



**PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA**



**SEGMENTAÇÃO SEMÂNTICA DE ÁREAS DESMATADAS UTILIZANDO REDES
NEURAIS CONVOLUCIONAIS NO SUL DA AMAZÔNIA LEGAL**

FERNANDA CAETANO COSTA

MANAUS

2020

FERNANDA CAETANO COSTA

SEGMENTAÇÃO SEMÂNTICA DE ÁREAS DESMATADAS UTILIZANDO REDES
NEURAIS CONVOLUCIONAIS NO SUL DA AMAZÔNIA LEGAL

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho
Co-Orientador(a): Prof^a. Dr^a. Marly Guimarães Fernandes Costa

MANAUS
2020

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

C837s Costa, Fernanda Caetano
Segmentação Semântica de Áreas Desmatadas utilizando Redes
Neurais Convolucionais no Sul da Amazônia Legal / Fernanda
Caetano Costa . 2020
89 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho
Coorientadora: Marly Guimarães Fernandes Costa
Dissertação (Mestrado em Engenharia Elétrica) - Universidade
Federal do Amazonas.

1. Áreas Desmatadas. 2. Redes Neurais Convolucionais. 3.
Imagem-Mosaico. 4. Landsat-8. 5. Segmentação. I. Costa Filho,
Cícero Ferreira Fernandes. II. Universidade Federal do Amazonas
III. Título

FERNANDA CAETANO COSTA

**SEGMENTAÇÃO SEMÂNTICA DE ÁREAS DESMATADAS
UTILIZANDO REDES NEURAIS CONVOLUCIONAIS NO SUL DA
AMAZÔNIA LEGAL**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 10 de novembro de 2020.

BANCA EXAMINADORA

Cícero F. F. Costa Filho

Prof. Dr. Cícero Ferreira Fernandes Costa Filho, Presidente

Universidade Federal do Amazonas

José Raimundo Gomes Pereira

Prof. Dr. José Raimundo Gomes Pereira, Membro

Universidade Federal do Amazonas

Clahildek Matos Xavier

Prof. Dr. Clahildek Matos Xavier, Membro

Universidade Federal do Amazonas

AGRADECIMENTOS

Agradeço primeiramente a Deus por abençoar e iluminar meus caminhos e por ter me dado forças para persistir e a Nossa Senhora, Maria Santíssima, por sua intercessão.

Agradeço aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e a Prof. Dr^a. Marly Guimarães Fernandes Costa, pelo comprometimento, dedicação, ensinamentos e orientações dada durante o andamento do mestrado.

Aos meus pais e aos meus irmãos e família, pelo apoio e incentivo constante, por estarem sempre ao meu lado em todos os momentos.

Aos meus amigos do mestrado Paula, Rodrigo, Mikaela, Patrese, Mateus, Bashi e Moisés pelo apoio e momentos de estudo.

Ao analista ambiental Jefferson Lobato pelas dicas e instruções sobre a área de sensoriamento remoto.

Aos meus amigos da Coordenação Técnica do Centro Regional de Manaus Renata, Rodrigo, Clahildek, Joel, Ailton, Ricardo, Ronaldo e a ao Coordenador Sebastião Lopes no qual todos torceram e acompanharam de perto a jornada do mestrado.

E por fim, a todos os meus amigos da caminhada espiritual do Shalom que sempre rezam por mim em especial nos momentos mais difíceis.

RESUMO

Neste trabalho propomos a utilização de redes neurais convolucionais para segmentação de áreas desmatadas na região do sul da Amazônia Legal em imagens de satélite Landsat-8. Para realização deste trabalho foi construído um banco de imagens mosaico, composto por amostras de áreas desmatadas e áreas de floresta extraídas das imagens de satélite Landsat-8, identificando as áreas desmatadas através da utilização dos dados de desmatamento do projeto PRODES do INPE. As arquiteturas de redes convolucionais utilizadas foram as propostas no trabalho de Serrão et al. (2020) e de Miyagawa et al. (2019). O treinamento das redes foi realizado com 32 épocas utilizando métodos de otimização SGDM, RMSProp e ADAM e os métodos de regularização L2 e *Dropout*. Combinando as três arquiteturas com esses métodos, totalizaram 36 simulações. Para avaliar o desempenho das arquiteturas para segmentação das áreas desmatadas, permitindo uma comparação entre os modelos, foi escolhida a métrica acurácia. Após a avaliação do desempenho dos modelos no conjunto de validação, seis deles foram selecionados para serem avaliados com o conjunto de teste. O modelo que apresentou o melhor resultado, uma acurácia de 99.97%, foi o que utilizou a seguinte combinação: CNN2 + RMSProp + *Dropout*. O resultado deste trabalho foi comparado com os resultados dos trabalhos de Ortega et al. (2019), Adarme et al. (2020) e De Bem et al. (2020) tendo obtido resultados superiores àqueles obtidos por esses autores. Os resultados mostraram que as redes neurais convolucionais são capazes de realizar com alto desempenho a tarefa de segmentação de áreas desmatadas.

Palavras-chave: áreas desmatadas, redes neurais convolucionais, imagem-mosaico, landsat-8, segmentação, algoritmos de otimização, técnicas de regularização.

ABSTRACT

In this work, we propose the evaluation of convolutional neural networks architectures to segment deforested areas in the southern region of the Brazilian Legal Amazon, using Landsat-8 satellite images. In order to carry out this work, a mosaic image data set was elaborated, consisting of samples of deforested areas and forest areas extracted from the Landsat-8 satellite images. The deforested areas were identified through the use of deforestation data from the PRODES project at INPE. The architectures of convolutional networks used in our research were those proposed in the work of Serrão et al. (2020) and Miyagawa et al. (2019). The training of the networks was carried out over 32 epochs using SGDM, RMSProp and ADAM optimization methods and the L2 and Dropout regularization methods. The combining of the three architectures with these methods, resulting in 36 simulations. To measure the performance of the architectures for segmented deforested region, allowing a comparison between the models, the accuracy metric was chosen. After evaluating the performance of the models in the validation set, six of them were selected to be evaluated with the test set. The model that presented the best result, with an accuracy of 99.97%, was the one that used the following combination: CNN2 + RMSProp + Dropout. The results of this work were compared with the results of the work of Ortega et al. (2019), Adarme et al. (2020) and De Bem et al. (2020). Our work obtained results best performance that the results obtained by these authors. The results showed that the convolutional neural networks are capable of performing the task of classifying deforested areas with high performance.

Keywords: deforestation, convolutional neural networks, mosaic image, landsat-8, segmentation, optimization algorithms, regularization techniques.

LISTA DE FIGURAS

Figura 2.1 – Área de desmatamento na Amazônia Legal Brasileira por km ² .	19
Figura 2.2 – Distribuição da área de desmatamento entre os estados da Amazônia Brasileira.	20
Figura 2.3 – Obtenção de imagens por sensoriamento remoto.	22
Figura 2.4 – Exemplo de um neurônio artificial.	25
Figura 2.5 – Exemplo de uma rede neural multicamada.	26
Figura 2.6 – Exemplo de arquitetura de rede convolucional.	28
Figura 2.7 – Ilustração da operação de convolução: a) <i>kernel 2x2</i> ; b) imagem de entrada e resultado da operação de convolução.	30
Figura 2.8 – Operação de convolução com <i>padding</i> e <i>kernel 2x2</i> . As partes sombreadas são a entrada e os elementos da matriz do <i>kernel</i> usados pelo primeiro elemento de saída: $0x0 + 0x1 + 0x2 + 0x3 = 0$.	30
Figura 2.9 – Exemplo de <i>Max-Pooling</i> .	32
Figura 2.10 – Exemplo de uma convolução transposta.	33
Figura 2.11 – Otimização de uma função simples: a) Trajetória de otimização do algoritmo SGD; b) Trajetória de otimização do algoritmo SGDM.	34
Figura 2.12 – Exemplo da aplicação da técnica dropout. a): Uma rede neural padrão com 2 camadas ocultas; b) exemplo de rede neural reduzida produzida pela aplicação de <i>dropout</i> na rede em a). As unidades marcadas com um “x” foram descartadas.	38
Figura 2.13 – Arquitetura CNN com segmentação semântica .	39
Figura 2.14 – Arquitetura da Rede Convolucional SegNet.	40
Figura 4.1 – Estado do Amazonas com as regiões de interesse destacadas.	58
Figura 4.2 – Exemplo de imagem-mosaico. a) à esquerda é ilustrada a imagem-mosaico com as amostras de floresta e desmatamento e b) à direita, a sua correspondente imagem binarizada.	64
Figura 4.3 – Arquiteturas CNN1, CNN2 e CNN3.	66

LISTA DE TABELAS

Tabela 2.1 – Bandas do sensor LANDSAT-8/OLI (USGS, 2020a).....	24
Tabela 3.1 – Resumo dos artigos selecionados.....	48
Tabela 4.1 – Cenas das imagens utilizadas.....	57
Tabela 4.2 – Quantidade de Amostras.....	62
Tabela 4.3 – Valores para os hiperparâmetros de treinamento.....	67
Tabela 4.4 – Matriz de Confusão.....	68
Tabela 5.1 – Resultados do desempenho da CNN1 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.....	71
Tabela 5.2 – Resultados do desempenho da CNN1 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de métodos para melhoria da generalização.....	72
Tabela 5.3 – Resultados do desempenho da CNN2 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.....	72
Tabela 5.4 – Resultados do desempenho da CNN2 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de métodos para melhoria da generalização.....	72
Tabela 5.5 – Resultados do desempenho da CNN3 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.....	73
Tabela 5.6 – Resultados do desempenho da CNN3 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de método de regularização.....	73
Tabela 5.7 – Modelos selecionados do conjunto de validação para aplicação do conjunto de teste.....	78
Tabela 5.8 – Resultados do desempenho das redes CNN1, CNN2 e CNN3 do conjunto de teste.....	78
Tabela 5.9 – Desempenho da CNN2 com algoritmo de otimização RMSProp e com método de regularização <i>dropout</i> no conjunto de teste.....	81
Tabela 5.10 – Tabela de comparação entre os resultados obtidos com a metodologia proposta e os resultados obtidos em trabalhos da literatura.....	83

LISTA DE ABREVIATURAS

AWIFS	<i>Advanced Wide Field Sensor</i>
ADAM	<i>Adaptive Moment Estimation</i>
CNN	<i>Convolutional Neural Network</i>
DETER	<i>Sistema de Detecção de Desmatamento em Tempo quase Real</i>
DETER-B	<i>Sistema de Detecção de Desmatamento em Tempo quase Real com base nos dados AWIFS e WFI</i>
EUA	<i>Estados Unidos da América</i>
GV	<i>Green Vegetation</i>
LDA	<i>Linear Discriminant Analysis</i>
LSMM	<i>Linear Spectral Mixing Model</i>
MSE	<i>Mean Square Error</i>
NDFI	<i>Normalized Degradation Fraction Index</i>
NPV	<i>Non-Photosynthetic Vegetation</i>
PCA	<i>Principal Component Analysis</i>
PRODES	<i>Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite</i>
RMSProp	<i>Root Mean Square Propagation</i>
ReLU	<i>Rectified Linear Unit</i>
SGDM	<i>Stochastic Gradient Descent with Momentum</i>
SEC	<i>Seleção Escalar de Características</i>
WFI	<i>Wide Field Imager</i>

SUMÁRIO

CAPÍTULO 1.....	13
1. INTRODUÇÃO.....	13
1.1 OBJETIVO GERAL	16
1.2 OBJETIVOS ESPECÍFICOS.....	16
1.3 DESENVOLVIMENTO DO TRABALHO.....	17
CAPÍTULO 2.....	18
2. FUNDAMENTOS TEÓRICOS.....	18
2.1 DESMATAMENTO NA AMAZÔNIA BRASILEIRA E O USO DO SENSORIAMENTO REMOTO	18
2.1.1 PRODES.....	21
2.1.2 Sensoriamento Remoto	22
2.1.3 Imagens LANDSAT-8.....	23
2.2 REDES NEURAIS CONVOLUCIONAIS.....	24
2.2.1 Redes Neurais Artificiais	24
2.2.2 Redes Neurais Convolucionais	27
2.2.2.1 Camada Convolucional	29
2.2.2.2 Unidade Retificadora Linear - ReLU.....	31
2.2.2.3 Camada de Subamostragem (<i>Pooling</i>)	31
2.2.2.4 Camada de Convolução Transposta	32
2.2.2.5 Camada de Totalmente Conectada (<i>Full Connected Layer</i>).....	33
2.3 ALGORITMO DE APRENDIZAGEM E HIPERPARÂMETROS.....	33
2.3.1 Algoritmo de aprendizagem	33
2.3.2 Hiperparâmetros	36
2.4 TÉCNICAS DE REGULARIZAÇÃO.....	37
2.5 SEGMENTAÇÃO SEMÂNTICA.....	39
2.6 CONSIDERAÇÕES.....	40
CAPÍTULO 3.....	41
3. REVISÃO BIBLIOGRÁFICA.....	41
3.1 TRABALHOS RELACIONADOS	41
3.2 CONSIDERAÇÕES E CONTRIBUIÇÃO DESTE TRABALHO.....	54
CAPÍTULO 4.....	56
4. METODOLOGIA	56
4.1 MATERIAIS.....	56
4.1.1 Imagens Utilizadas.....	56

4.1.2	Dados Geoespaciais Utilizados.....	58
4.2	MÉTODOS.....	60
4.2.1	Construção do Banco de Imagens Mosaico	60
4.2.2	Propostas das Arquiteturas de Redes Neurais Convolucionais	64
4.2.3	Treinamento, Validação e Teste das Arquiteturas	66
4.2.4	Seleção do Melhor Modelo de Arquitetura	67
4.2.5	Avaliação das Arquiteturas Propostas.....	68
4.2.6	Ambiente de Desenvolvimento.....	69
4.3	CONSIDERAÇÕES.....	69
CAPÍTULO 5.....		71
5.	RESULTADOS E DISCUSSÕES	71
5.1	Resultados das Simulações das Arquiteturas no Conjunto de Validação	71
5.2	Análise dos Resultados das Arquiteturas no Conjunto de Validação.....	73
5.3	Resultados da Seleção do Modelo com Melhor Desempenho	78
5.4	Resultado dos Testes dos Modelos Selecionados	78
5.5	Avaliação do Resultado com uso da Técnica Mosaico.....	82
5.6	Comparação com trabalhos da literatura.....	82
5.7	CONSIDERAÇÕES.....	83
CAPÍTULO 6.....		84
6.	CONCLUSÕES E TRABALHOS FUTUROS	84
REFERÊNCIAS		86

1. INTRODUÇÃO

Nos últimos cinco anos, as áreas desmatadas de floresta vêm aumentando de forma significativa na Amazônia Legal. Para se ter uma ideia, entre o período de agosto de 2018 a julho de 2019, a área desmatada correspondeu a 9.762 km², o que representa um aumento de 29,54% em relação ao ano anterior, quando o desmatamento foi de 7.536 km². No Estado do Amazonas, no ano de 2019, foi registrado um total de 1.421 km² de área desmatada, o que o coloca entre os quatro maiores estados da Amazônia Legal em termos de área desmatada (INPE, 2020b).

O monitoramento de desmatamento na Amazônia é feito por imagens de satélite, obtidas a partir de sistemas de sensoriamento remoto. O sensoriamento remoto é um conjunto de técnicas e métodos para coletar informações espaciais sobre a superfície da Terra à distância, sem que haja contato com o objeto monitorado (Coops e Tooke, 2017; Jensen, 2011). No Brasil, o principal responsável pelo monitoramento na Amazônia é o Instituto Nacional de Pesquisas Espaciais (INPE) que possui o Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES). Esse programa estima de anualmente a taxa de desmatamento por corte raso desde 1988. As estimativas desse programa são consideradas confiáveis pelos cientistas nacionais e internacionais (Kintisch, 2007). O PRODES considera desmatamento por corte raso a remoção completa da cobertura florestal em um curto intervalo de tempo, substituída por outras coberturas e usos (agrícola, pastagem, urbano, hidroelétrica, etc.), ou ainda pode ser abandonada e entrar em processo de regeneração (INPE, 2020a).

Neste sentido, diversos trabalhos sobre monitoramento florestal foram propostos na literatura com o objetivo de realizar mapeamento, detecção e classificação das áreas em estudo. A seguir elencamos alguns desses trabalhos.

Diniz et al. (2015) compararam as áreas classificadas de 188 municípios pelo sistema DETER-B com a classificação feita pelo PRODES. Nessa comparação, as áreas envolvidas somavam 373.529,8 ha de desmatamento com e sem vegetação e mineração, e 406.963,7 ha de desmatamento e classes residuais, obtendo um coeficiente de correlação (R^2) = 0,83, o que indica uma forte correlação linear entre ambos os sistemas. O DETER – Sistema de Detecção de Desmatamento em Tempo quase Real – é um sistema de alerta precoce projetado para apoiar a vigilância e o controle do desmatamento, tendo como objetivo mapear a ocorrência de áreas de corte e degradação florestal superiores a 25 ha. O DETER-B é o sistema DETER que utiliza dados de sensores AWiFS (56 m) e WFI (64 m).

Oliveira et al. (2017) propuseram avaliar métodos para melhoria da generalização das redes neurais para a detecção de áreas desmatadas na região sul do Estado do Amazonas, associadas aos métodos de extração de características Análise dos Componentes Principais (*Principal Component Analysis* – PCA), Análise de Discriminante Linear (*Linear Discriminant Analysis* – LDA) e Seleção Escalar de Características - SEC. Os autores utilizaram uma rede neural para realizar a classificação com uso de critérios de parada antecipada, regularização e MSE. Os autores obtiveram uma acurácia de 97% com o método de generalização de parada antecipada.

Shimabukuro et al. (2019) apresentaram um método semiautomatizado para monitoramento de desmatamento e degradação florestal na Amazônia brasileira usando imagens de fração multitemporal derivadas do Modelo Linear de Mistura Espectral (MLME). A abordagem empregada foi a de mapear áreas desmatadas e de degradação florestal causadas por incêndios através de técnicas de segmentação de imagens. A principal contribuição desses autores consiste em separar explicitamente a degradação florestal causada por atividades de corte seletivas de exploração e degradação florestal causada por incêndios não controlados. O método proposto pelos autores obteve acurácias de 67% e 68%, respectivamente, para a classificação de áreas degradadas por incêndios e corte seletivo.

Ortega et al. (2019) realizaram um estudo para avaliar métodos de detecção automática para desmatamento florestal com Rede Convolutacional de Fusão Precoce (EF) e Rede Neural Convolutacional Siamesa (S-CNN), fazendo um comparativo com Máquina de Vetores de Suporte (SVM). Esses métodos foram avaliados em uma região do Pará, localizado na Amazônia Brasileira, utilizando duas imagens LandSat-8 de 2016 e 2017. Os experimentos mostraram que o método S-CNN apresentou, para a métrica F1-score, um valor de 65% e uma acurácia de 98%. Com isso, os autores concluíram que mais pesquisas são necessárias para obter resultados viáveis, como por exemplo o ajuste dos hiperparâmetros dos métodos estudados.

Adarme et al. (2020) avaliaram técnicas baseadas em Aprendizagem Profunda aplicadas à detecção de desmatamento em duas regiões tropicais com diferentes padrões de desmatamento: os biomas Amazônia e Cerrado. Os autores obtiveram, para a região amazônica, uma acurácia de 97,4% para uma rede CNN de Fusão Antecipada.

De Bem et al. (2020) propuseram o uso de CNNs na detecção de desmatamento na Amazônia brasileira, comparando os resultados obtidos por técnicas de Aprendizagem Profunda com os de algoritmos clássicos de Aprendizagem de Máquina (ML). Os autores obtiveram acurácia global de 99,93%, com a rede ResUnet.

Nesse breve resumo, vemos que o trabalho de Shimabukuro et al. (2019), que utilizou segmentação de imagens para mapear áreas desmatadas, apresentou um baixo valor de acurácia, em torno de 70%. Melhores valores para a acurácias, superiores a 90%, foram

obtidos nos trabalhos de Oliveira et al. (2017), que utilizaram redes neurais inteiramente conectadas, e nos trabalhos de Ortega et al.(2019), Adarme et al. (2020) e De Bem et al. (2020) que usaram redes CNN.

O intuito deste trabalho é utilizar redes convolucionais para segmentação de áreas desmatadas com a técnica de imagem mosaico. A técnica da imagem mosaico foi utilizada para realizar um balanceamento entre as classes de desmatamento e floresta, que ocorre devido a predominância das áreas de floresta em relação as áreas desmatadas nas imagens de satélites. As redes convolucionais, diferentemente das redes neurais convencionais, são compostas por uma sequência de diferentes tipos de camadas, em que os neurônios de uma camada se conectam apenas a uma pequena região de neurônios da saída da camada anterior. Em comparação com as redes neurais inteiramente conectadas, essa característica implica numa redução na quantidade de parâmetros em cada camada. A principal vantagem das redes convolucionais frente as redes inteiramente conectadas é o fato de não ser necessário extração de características para servirem de entrada às mesmas. No trabalho de Oliveira et al. (2017), por exemplo, que utilizou redes inteiramente conectadas, os autores utilizaram as seguintes técnicas de extração de características: PCA, LDA e SEC. Por fim, por admitirem entradas multidimensionais, as redes convolucionais são muito apropriadas para o processamento de imagens de satélite, que disponibilizam várias bandas com informações diferenciadas.

A utilização de redes convolucionais para detecção de regiões desmatadas na região amazônica ainda é pouco explorada. Trabalhos mais recentes como os de Ortega et al. (2019), Adarme et al. (2020) e De Bem et al. (2020) mostram que o uso de CNN resultou em uma alta taxa de acurácia na detecção do desmatamento. No entanto, o desempenho significativo desses classificadores requer um grande número de amostras da área em estudo. Além disso, as características espectrais não são as mesmas para o desmatamento em toda a região.

Este trabalho pretende explorar a utilização de redes convolucionias para detecção de regiões desmatadas no sul do Estado do Amazonas, mais precisamente, em regiões selecionadas nos municípios de Apuí, Humaitá e Lábrea, como explorado no trabalho de Oliveira et al (2017). Além disso, estenderemos o estudo para regiões situadas nos municípios de Boca do Acre, Canutama, Manicoré e Novo Aripuanã e para uma porção das regiões vizinhas dos estados do Mato Grosso, Rondônia, Acre e Pará. Para essas regiões de estudo foram selecionadas 21 cenas das imagens LANDSAT-8 de forma a obter quantidade de amostras suficiente das áreas desmatadas.

Considerando que a região sul do Estado do Amazonas e regiões vizinhas estão sob constante ameaça de desmatamentos ilegais, acreditamos ser oportuno e necessário a aplicação de técnicas computacionais mais recentes e mais simples, como redes

convolucionais, que dispensam etapas de extração de características, para o mapeamento dessas áreas. Esperamos, dessa forma, contribuir para a pesquisa na área, tão necessária em órgãos federais que lidam com essa problemática.

1.1 OBJETIVO GERAL

O principal objetivo deste trabalho é avaliar arquiteturas de redes convolucionais para a segmentação semântica de áreas desmatadas por corte raso da floresta Amazônica no Sul da Amazônia Legal.

1.2 OBJETIVOS ESPECÍFICOS

Para alcançar o propósito principal deste trabalho, os seguintes objetivos específicos foram elencados:

- 1) Construir uma base de dados de imagens em mosaico utilizando os dados de satélite LANDSAT-8 para as áreas de estudo;
- 2) Propor e comparar o desempenho de três arquiteturas de redes convolutivas para a classificação e segmentação de áreas desmatadas;
- 3) Avaliar se a utilização da técnica de regularização *dropout* e *L2* aumenta a capacidade de generalização dos modelos treinados com redes neurais convolucionais;
- 4) Avaliar o desempenho das arquiteturas propostas e suas parametrizações utilizando métricas quantitativas;
- 5) Avaliar a utilização da técnica mosaico para construção de conjuntos de imagens de treinamento, teste e validação, no estudo de desmatamento por corte raso;
- 6) Comparar os resultados nesse trabalho com os registrados pelo estado da arte na literatura;

1.3 DESENVOLVIMENTO DO TRABALHO

Este trabalho está organizado nos seguintes capítulos:

- Capítulo I: Introdução;
- Capítulo II: Fundamentos Teóricos;
- Capítulo III: Revisão Bibliográfica;
- Capítulo IV: Metodologia;
- Capítulo VI: Resultados preliminares e discussão;

O Capítulo I apresenta um contexto para o trabalho, situando o mesmo frente a outras pesquisas da área e delineando as diferenças e semelhanças. São apresentados também os objetivos geral e específicos.

O Capítulo II apresenta os principais fundamentos teóricos utilizados no desenvolvimento deste trabalho: a) Sensoriamento remoto e desmatamento; b) Redes neurais artificiais; c) Redes neurais convolucionais e seu desenvolvimento na segmentação de imagens; d) Métodos de otimização, e; f) Segmentação semântica.

O Capítulo III apresenta os principais trabalhos que foram publicados na literatura sobre o tema abordado nessa dissertação, qual seja, a detecção de áreas desmatadas.

No Capítulo IV são apresentados os materiais e métodos utilizados nesta pesquisa. Nesse capítulo, é descrito como foi realizada a elaboração da base de dados, as arquiteturas propostas e suas parametrizações e quais métricas serão utilizadas para avaliação do desempenho dos métodos de segmentação semântica.

Por fim, no Capítulo VI são apresentados e analisados os resultados alcançados neste trabalho, comparando-os com os registrados na literatura.

CAPÍTULO 2

2. FUNDAMENTOS TEÓRICOS

Neste capítulo introduzimos os principais conceitos necessários ao entendimento deste trabalho. Estruturalmente, este capítulo consiste de sete seções. A Seção 2.1 apresenta os conceitos de desmatamento e o uso do sensoriamento remoto. A Seção 2.2 introduz o programa de monitoramento de desmatamento do INPE. A Seção 2.3 aborda o programa LANDSAT-8, do qual as imagens desse trabalho foram extraídas. Na Seção 2.4 são abordados os conceitos de redes neurais convolucionais. Na Seção 2.5 são apresentados os conceitos dos algoritmos de treinamentos e seus hiperparâmetros. A Seção 2.6 descreve conceitos sobre técnicas de regularização utilizadas neste trabalho. Por fim, na Seção 2.7 são introduzidos os conceitos de segmentação semântica.

2.1 DESMATAMENTO NA AMAZÔNIA BRASILEIRA E O USO DO SENSORIAMENTO REMOTO

A Amazônia Legal Brasileira é considerada a maior floresta tropical do mundo, ocupando grande parte na região norte do Brasil. Seu bioma vegetal ocupa cerca de 61% do território brasileiro com uma área total de 5.217.423 km² (IBGE, 2014). A região abrange nove Estados: Amazonas, Acre, Amapá, Mato Grosso, Maranhão, Pará, Roraima, Rondônia e Tocantins.

A Amazônia Legal Brasileira possui uma grande biodiversidade com relevância econômica, social e ambiental que chama a atenção de muitos exploradores de madeira e de minérios, o que, inevitavelmente, resulta no desmatamento da floresta.

Segundo Blanc (2016), o desmatamento é definido como destruição completa da cobertura florestal. Para o INPE, o desmatamento é a conversão por supressão de áreas de fisionomia florestal primária por ações antropogênicas (INPE, 2019). Assim, podemos entender de forma clara, que o desmatamento é a retirada completa da cobertura vegetal de uma área. A retirada dessa cobertura vegetal é devida a diversas atividades como a exploração madeireira, expansão da agricultura e pecuária, mineração e a expansão desordenada das áreas urbanas dos municípios.

É importante ressaltar que os principais impactos do desmatamento, de acordo com Fearnside (2006), são a perda do uso sustentável da floresta, como o manejo florestal, a perda da biodiversidade da floresta, a redução da ciclagem da água (e da precipitação) e a emissão

de gás carbônico (CO₂), o que contribui para o aquecimento global, com mudanças climáticas e biológicas relevantes.

De acordo com o MMA (2018), o desmatamento, conforme onde ocorra, é classificado nas seguintes categorias fundiárias: áreas privadas/outras, assentamento, gleba, terra indígena e unidades de conservação de proteção integral ou uso sustentável. Um dado interessante é que, no ano de 2015, as categorias áreas privadas/outras, assentamentos e glebas, foram as que apresentaram a maior proporção de desmatamento.

O desmatamento na Amazônia Brasileira aumentou significativamente nos últimos cinco anos. A Figura 2.1 mostra a evolução da área desmatada em km² no período de 2015 a 2019. Para se ter uma ideia, entre o período de agosto de 2018 a julho de 2019 o aumento foi de 29,54% em relação ao período de agosto de 2017 a julho de 2018. Registramos ainda que, no período de 2015 a 2019, na Amazônia Legal Brasileira, houve um aumento em torno 57% no desmatamento, tendo a área desmatada saltado de 6.207 km² para 9.762 km².

A distribuição do desmatamento não ocorre de forma uniforme entre as regiões da Amazônia. Cada Estado possui suas peculiaridades. Na Figura 2.2 podemos acompanhar a evolução da área de desmatamento em cada estado, observando que os Estados do Pará, Mato Grosso, Rondônia e Amazonas possuem as maiores áreas de desmatamento entre o período de 2015 a 2019, ficando o Estado do Amazonas em quarto lugar. Esses quatro estados concentraram 84,38% da área total desmatada em 2019, que foi em torno de 9.733 km², com o Estado do Amazonas representando 14,60% do total.

O INPE, através do projeto PRODES, realiza o cálculo das áreas desmatadas para cada ano, considerando o período entre agosto do ano anterior e julho do ano corrente. O PRODES possui uma precisão próxima de 95% na identificação das áreas desmatadas (INPE, 2020a).

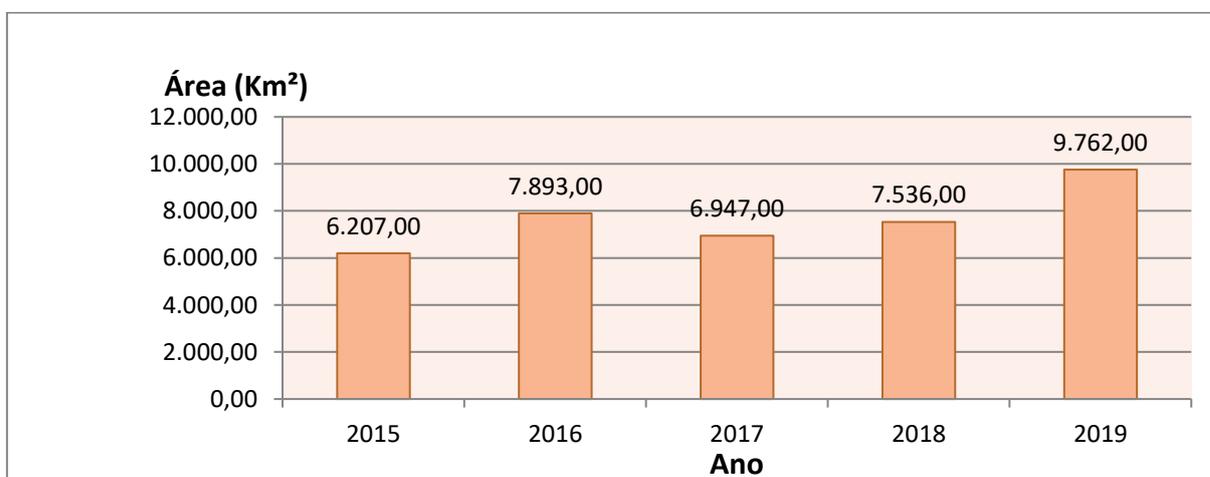


Figura 2.1 – Área de desmatamento na Amazônia Legal Brasileira por km².
Fonte: INPE (2020c).

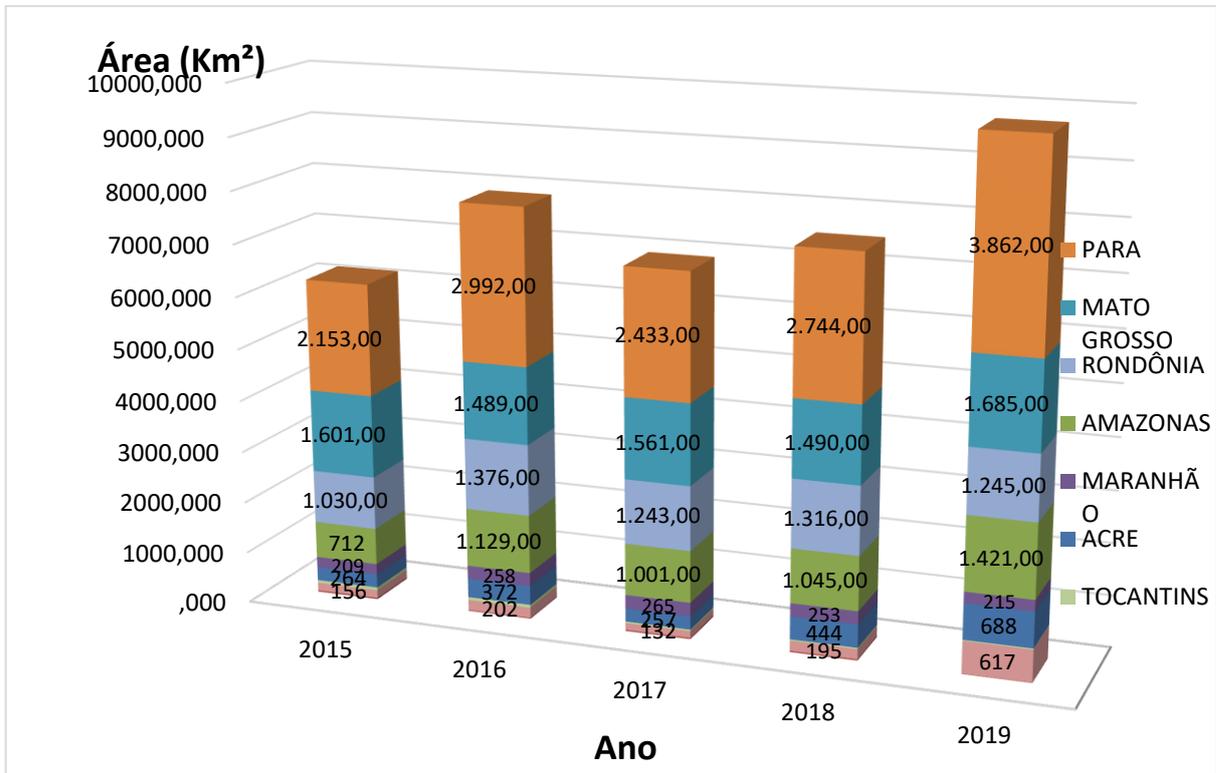


Figura 2.2 – Distribuição da área de desmatamento entre os estados da Amazônia Brasileira.
Fonte: INPE (2020c).

Devido à grande preocupação mundial com o desmatamento, diversos meios de sensoriamento remoto estão sendo utilizados para coletar informações à distância da superfície da Terra, sem a necessidade da presença humana no local monitorado. Esse monitoramento, quase em tempo real, é efetuado por meio de drones, sensores em satélites ou aeronaves, entre outros. Na Subseção 2.1.2 falaremos um pouco mais sobre o sistema de sensoriamento remoto.

No Brasil, o Instituto Nacional de Pesquisas Espaciais (INPE) é o principal responsável pelo monitoramento da Amazônia, possuindo três projetos que utilizam imagens satelitais para monitoramento da Amazônia: a) PRODES – Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite – estima a taxa de desmatamento anual por corte raso (remoção completa da cobertura florestal em um curto intervalo de tempo); b) DETER – Sistema de Detecção de Desmatamento em Tempo Real – sistema de alertas de alteração da cobertura florestal da Amazônia para apoio à fiscalização e controle de desmatamento e degradação florestal, e; c) TerraClass - sistema de mapeamento do uso e ocupação do solo.

É importante ressaltar que imagens satelitais obtidas por sensoriamento remoto são cada vez mais utilizadas para monitoramento de paisagens. De acordo com Blanc (2016), as florestas são facilmente reconhecíveis em imagens de satélite, diferenciando-se claramente

de outros tipos de uso do solo. Segundo o INPE (2019), o desmatamento por corte raso deixa cicatrizes evidentes que são facilmente reconhecidas em imagens de satélites.

As imagens utilizadas neste trabalho foram obtidas através do projeto PRODES, por serem consideradas confiáveis, conforme avaliação de pesquisadores nacionais e internacionais (Kintisch, 2007). São dados validados e de fácil disponibilidade na internet pelo sítio do INPE. O INPE disponibiliza tanto as imagens utilizadas na identificação de áreas desmatadas como seus dados geoespaciais de desmatamento. Assim sendo, falaremos um pouco mais sobre o PRODES na Subseção 2.1.1.

2.1.1 PRODES

Segundo o INPE, desde 1988 o PRODES realiza o monitoramento da região amazônica por meio de imagens satelitais obtidas por sensoriamento remoto com o intuito de identificar as áreas desmatadas por corte raso, com tamanho superior a 6,35 hectares. A taxa de desmatamento é calculada somando-se as áreas que tiveram no mínimo 90% de desmatamento no ano anterior. Essas taxas anuais são consideradas importantes para o governo federal definir suas políticas públicas de uso da terra.

As imagens utilizadas pelo PRODES são imagens de satélite da série Landsat da NASA/USGS (EUA) com 20 a 30 m de resolução espacial e período de revisitação do satélite de 16 dias. Essas resoluções são descritas na próxima Subseção. Todas essas imagens são ortorretificadas, ou seja, as imagens estão prontas para serem utilizadas junto com dados geoespaciais, sem nenhum processamento digital adicional na imagem. Na Subseção 2.1.3 falaremos um pouco sobre as imagens LANDSAT-8/OLI.

De acordo com o INPE (2019), o projeto PRODES realiza um mapeamento incremental através da fotointerpretação por especialistas. Para produção do mapeamento incremental é utilizada uma máscara de exclusão, que encobre as áreas desmatadas dos anos anteriores. A metodologia consiste das seguintes fases: 1) Seleção de imagens com a menor cobertura de nuvens que compreende o período de 01 de agosto de um ano até 31 de julho do ano subsequente; 2) Mapeamento dos polígonos de desmatamento por fotointerpretação de imagens realizada por especialistas. Esses especialistas delimitam as áreas desmatadas na tela do computador, avaliando o padrão de alteração de cobertura florestal pela tonalidade, textura e contexto da imagem, e; 3) Cálculo da taxa anual de desmatamento.

2.1.2 Sensoriamento Remoto

O sensoriamento remoto é uma das tecnologias mais eficiente para obter dados de forma automática para o monitoramento de paisagens florestais. Segundo Meneses e Almeida (2012) essa tecnologia é capaz de imagear a superfície terrestre do nosso planeta através de sensores a distância, não havendo contato físico de qualquer espécie entre o sensor e o objeto. Esses sensores, por meio de medições de radiação eletromagnética, conseguem registrar o objeto alvo.

A radiação eletromagnética necessita de dois elementos para funcionar: onda e energia. Desde o século passado, sabe-se que a radiação eletromagnética tem uma natureza dupla, ondulatória e corpuscular. Em alguns fenômenos físicos manifesta-se a natureza ondulatória e, em outros, a natureza corpuscular (fótons).

A Figura 2.3 ilustra como ocorre a obtenção de imagens através do sensoriamento remoto. O sol, que é a principal fonte de energia desse sistema, emite energia solar sobre a superfície terrestre. Uma parte dessa energia é absorvida pela terra e outra parte é refletida. Essa energia refletida é capturada pelos sensores, que são dispositivos de detecção, e enviada posteriormente para uma estação receptora, como por exemplo, uma antena receptora.

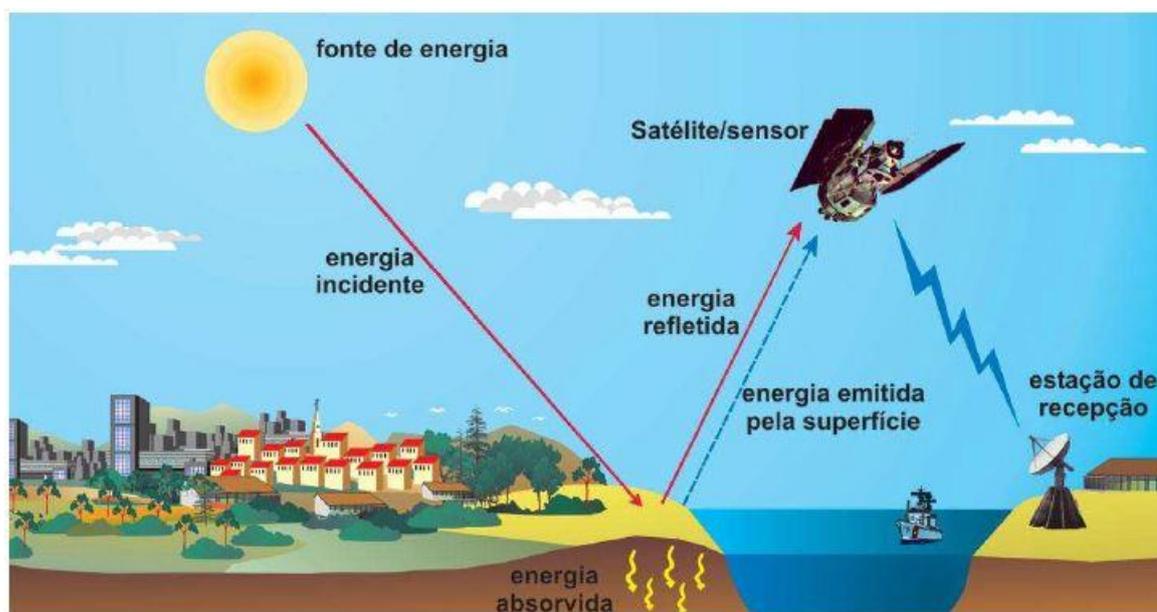


Figura 2.3 – Obtenção de imagens por sensoriamento remoto.

Fonte: FLORENZANO (2011).

Em sensoriamento remoto, os sensores utilizam a radiação eletromagnética para obter informações do objeto visado. Os sensores podem ser classificados como ativos e passivos. Os sensores ativos são os que produzem sua própria fonte de energia eletromagnética, como

por exemplo, os radares e as câmeras fotográficas. Os sensores passivos são os que dependem de uma fonte de energia externa, que geralmente é a energia solar, como por exemplo, os sensores satelitais do LANDSAT-8.

Os principais tipos de dados capturados por sensores são imagens digitais. As principais características que devem ser consideradas ao observar um objeto em uma imagem são os tipos de resoluções em que essas imagens foram obtidas: espacial, espectral, radiométrica e temporal.

Segundo Menezes e Almeida (2012), a **resolução espacial** é definida como o menor tamanho de um objeto a ser identificado na imagem. O objeto apenas pode ser reconhecido ou detectado na imagem quando seu tamanho for igual ou maior do que o tamanho do elemento de resolução do terreno. Por exemplo, se um objeto tem um tamanho de 10m x 10m, a resolução espacial mínima deveria ser de 10 metros para que este objeto seja identificado. A **resolução temporal** é o tempo de frequência que o sensor revisita a mesma área e obtém imagens periódicas. Por exemplo, os sensores TM-Landsat possuem resolução temporal de 16 dias (USGS, 2020). Esta resolução é fundamental para monitorar mudanças e alterações no terreno, como por exemplo, o desmatamento. A **resolução espectral** é a obtenção simultânea de imagens em múltiplas bandas espectrais. Possui três parâmetros de medida: número de bandas do sensor, largura das bandas e as posições das bandas na região do espectro eletromagnético. Finalmente, a **resolução radiométrica** está relacionada com a intensidade de radiação da área de cada pixel unitário. Quanto maior for a capacidade de medir as diferenças de intensidade dos níveis de radiação, maior será a resolução radiométrica. Essa resolução define o número de níveis de radiação que o detector pode discriminar.

Neste trabalho, os dados utilizados são imagens obtidas por sensoriamento remoto do sensor OLI (*Operational Land Imager* – Imageador Operacional do Terreno) do satélite LANDSAT-8. Essas imagens possuem resolução espacial de 30m, resolução temporal de 16 dias, resolução radiométrica de 16 bits e diferentes resoluções espectrais que podem ser visualizadas na Tabela 2.1 na Subseção 2.1.3 a seguir.

2.1.3 Imagens LANDSAT-8

LANDSAT é um programa de satélite de observação da Terra do Serviço Geológico dos Estados Unidos. O mais novo satélite da série LANDSAT é o LANDSAT-8, lançado em 2013. O mesmo oferece uma melhor resolução e maior sensibilidade ao brilho e cor do que os satélites anteriores. Segundo Acharya e Yang (2015), este novo satélite realiza o imageamento da Terra inteira a cada 16 dias, produzindo mais de 400 cenas por dia. Cada cena captura uma área de 170x183 km² (Norte-Sul, Leste-Oeste) com tamanho do arquivo da

imagem em média de 1 GB. As bandas multiespectral e pancromática possuem resolução espacial de 30 e 15 m, respectivamente. A Tabela 2.1 mostra a descrição das bandas do sensor OLI. Este trabalho fez uso das bandas de 2 a 7 por serem consideradas as mais adequadas para identificar regiões de desmatamento (USGS, 2020).

Tabela 2.1 – Bandas do sensor LANDSAT-8/OLI (USGS, 2020).

Sensor	Banda	Nome da Banda	Resolução Espectral (μm)	Resolução Espacial	Resolução Temporal	Resolução Radiométrica	Área Imageada
Operational Land Imager (OLI)	1	Aerosol	0,435 - 0,451	30 m			170 x 183 km ²
	2	Azul	0,452 - 0,512	30 m			
	3	Verde	0,533 - 0,590	30 m			
	4	Vermelho	0,636 - 0,673	30 m			
	5	Infravermelho próximo	0,851 - 0,879	30 m			
	6	Infravermelho de ondas curtas 1	1,5166 - 1,651	30 m	16 dias	16 bits	
	7	Infravermelho de ondas curtas 2	2,107 - 2,294	30 m			
	8	Pancromática	0,503 - 0,676	15 m			
	9	Cirrus	1,363 - 1,384	30 m			

Cada imagem obtida do LandSat é identificada pelo número de órbita / número da cena. Esse par de informações recebe a denominação de ponto. Por exemplo, a imagem 1/66 refere-se ao município de Lábrea no estado do Amazonas.

2.2 REDES NEURAIIS CONVOLUCIONAIS

Neste trabalho, a técnica de aprendizado de máquina utilizada foi Redes Neurais Convolucionais – (*Convolutional Neural Network* - CNN). Antes de abordar as CNNs, abordaremos alguns conceitos relacionados à Redes Neurais Artificiais - (*Artificial Neural Network* - ANN) úteis para sua compreensão. As CNN's serão abordadas na Subseção 2.2.2.

2.2.1 Redes Neurais Artificiais

Segundo Silva et al. (2010) as redes neurais artificiais são métodos computacionais inspirados no funcionamento dos sistemas nervosos biológicos. São constituídas por um conjunto de neurônios interligados entre si por meio de um grande número de interconexões (sinapses artificiais), representadas por vetores ou matrizes de pesos sinápticos. Esse conjunto de neurônios, através de um processo de aprendizagem, é capaz de aprender e armazenar informações para desempenhar uma determinada tarefa.

O processo de aprendizagem consiste na implementação de um algoritmo de treinamento em que os pesos dos neurônios devem ser ajustados de forma a obter a saída desejada. Os principais exemplos de algoritmo de treinamento são o Gradiente Descendente Estocástico com Momento (SGDM) e o Levenberg-Marquardt (LM), (Hagan et al., 2014).

As principais características das redes neurais artificiais são as seguintes: tolerância a falhas, alta capacidade de armazenamento e aprendizagem e alta habilidade de generalizar o conhecimento para exemplos que não fazem parte do conjunto de treinamento (Silva et al., 2010). As redes neurais são utilizadas em diversas aplicações: na área de tecnologia, para reconhecimento de face e voz; na área da medicina, para exames de ressonância magnética; no setor automotivo e aeroespacial; na indústria de alimentos; na área ambiental, para o monitoramento de paisagens, etc.

A estrutura de uma rede neural é baseada em um neurônio artificial semelhante ao neurônio biológico. A Figura 2.4 ilustra o neurônio artificial da rede constituído por sete elementos básicos:

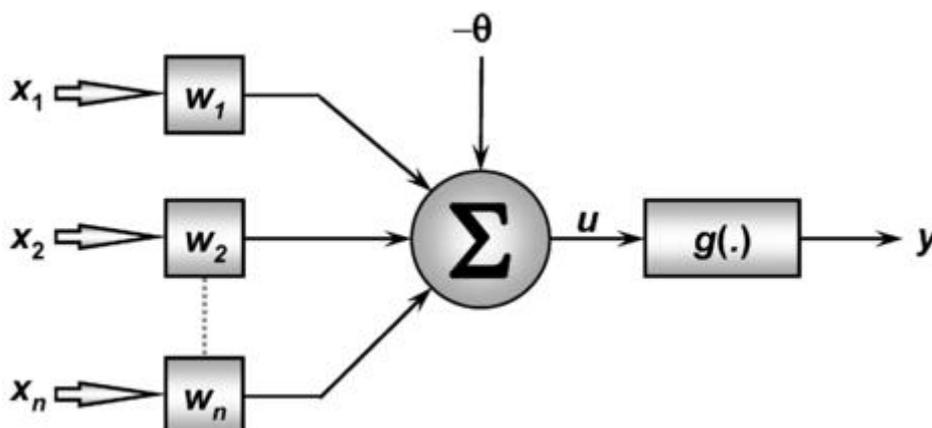


Figura 2.4 – Exemplo de um neurônio artificial.
Fonte: Silva et al. (2010).

- 1) Sinais de entrada $\{x_1, x_2, \dots, x_n\}$ – são valores obtidos do meio externo;
- 2) Pesos sinápticos $\{w_1, w_2, \dots, w_n\}$ – são valores para ponderar cada valor de entrada na rede;
- 3) Somador $\{\Sigma\}$ – agrega os sinais de entradas que foram ponderados pelos seus respectivos pesos sinápticos e a polarização ($u = \sum_{i=1}^n w_i x_i - \theta$).
- 4) Limiar de Ativação $\{\theta\}$ – polarização do neurônio;
- 5) Entrada líquida da função $g\{u\}$ – resultado da operação do somador;
- 6) Função de ativação $\{g\}$ - função linear ou não linear (identidade, sigmoide, tangente hiperbólica, softmax, etc.);
- 7) Sinal de saída $\{y\}$ – também denominada de ativação do neurônio (variável binária ou real).

De forma resumida, o funcionamento do neurônio artificial consiste inicialmente em apresentar um vetor de características na sua entrada. Essas características são multiplicadas por seus respectivos pesos sinápticos e obtém-se, através do somador, o valor da entrada líquida da função de ativação. Aplica-se a função g nessa entrada líquida e obtém-se o valor de ativação do neurônio.

A arquitetura de uma rede neural define como os neurônios estão estruturados e conectados. A arquitetura de uma rede neural pode ser dividida em três tipos de camadas: entrada, escondidas ou intermediárias e saídas. A camada de entrada é responsável por receber informações ou dados obtidas do meio externo, vetor $x = (x_1, x_2, x_3, \dots, x_n)$. As camadas escondidas ou intermediárias (podem ser constituídas por uma ou mais camadas de neurônios) são responsáveis por gerar novas características a partir dos dados de entrada. Os neurônios da camada de saída produzem a ativação da rede que, no exemplo da Figura 2.5, constitui-se do vetor y .

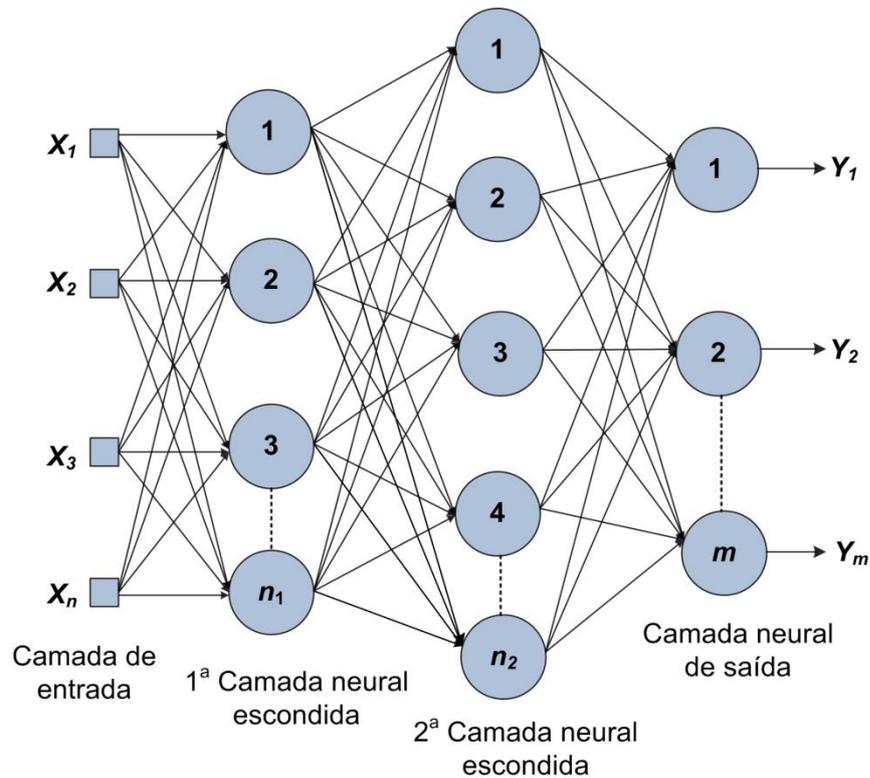


Figura 2.5 – Exemplo de uma rede neural multicamada.
Fonte: Adaptado de Silva et al. (2010).

O treinamento de uma ANN consiste em ajustar os pesos sinápticos e limiares dos seus neurônios, de forma a obter a ativação desejada na saída. Esse ajuste é efetuado através de um algoritmo de aprendizagem. Ao longo desse processo de treinamento, a rede é capaz de gerar novas características a partir dos vetores apresentados de entradas. O algoritmo de aprendizagem mais utilizado no treinamento de uma ANN para atualização de pesos e polarizações é o algoritmo de *backpropagation* (Awad e Khanna, 2015).

De forma resumida, o algoritmo *backpropagation* constitui-se de três etapas: 1) Inicialização: nesta etapa todos os pesos sinápticos da rede são inicializados aleatoriamente assumindo valores próximos a 0 (zero); 2) Propagação para frente (*forward propagation*): as saídas das camadas são calculadas de forma sucessiva, da camada de entrada, para a camada de saída. O vetor obtido na saída da rede é comparado com um valor desejado, através de uma função de erro; 3) Propagação para trás (*backward propagation*): os pesos e polarizações são atualizados por algum algoritmo de aprendizagem, normalmente o gradiente descendente. A ordem de atualização dos pesos e polarizações é inversa àquela do *forward propagation*, ou seja, primeiramente são atualizados os pesos da última camada, depois da penúltima camada, até chegar nos pesos da camada de entrada.

2.2.2 Redes Neurais Convolucionais

Rede Neural Convolucional (CNN) é um tipo de rede neural artificial complexa, considerada uma das melhores ferramentas disponíveis atualmente para reconhecimento, classificação, segmentação e processamento de imagens. Segundo Khan (2019), a estrutura da CNN é baseada em uma rede de múltiplos estágios de aprendizagem de propagação direta. Os estágios de aprendizagem consistem na combinação de camadas convolucionais e camadas de subamostragem. Nas camadas convolucionais ocorre a operação de convolução, que consiste na aplicação de um filtro de pesos para extrair características dos pixels de uma sub-região da imagem de entrada, gerando um mapa de características da camada. Esse mapa de característica é apresentado a uma unidade de processamento não linear, denominada Unidade Retificadora Linear (ReLU – *Rectifier Linear Unity*). Nas camadas de subamostragem ocorrem a redução de forma gradual do mapa de características, através da redução de parâmetros utilizando cálculos computacionais complexos. A camada de subamostragem pode ser seguida de combinação de outras camadas que serão descritas nas seções seguintes.

Diferentemente das redes neurais artificiais convencionais, vimos que as redes convolucionais são compostas por uma sequência de combinações de diferentes tipos de camadas, em que os neurônios de uma camada se conectam apenas a uma pequena região de neurônios da saída da camada anterior. Em comparação com as redes neurais inteiramente conectadas, essa característica implica numa redução na quantidade de parâmetros em cada camada. A principal vantagem das redes convolucionais é o fato de não ser necessário extração de características da imagem de entrada para servirem de entrada às mesmas. As próprias camadas iniciais da CNN são responsáveis por extraírem essas características.

Outro fato a ser mencionado da estrutura das CNNs, é que cada camada na rede é uma estrutura de grade tridimensional, com altura, largura e profundidade. Além disso, em cada camada há filtros (*kernels*) que são sobrepostos em uma região da imagem, levando em consideração a informação da vizinhança do pixel, gerando uma nova matriz reduzida (convolução) ou maior (deconvolução), que se constitui no mapa de características. Todas essas características obtidas da imagem se propagam da camada de entrada para a camada de saída da rede.

A Figura 2.6 ilustra um exemplo de arquitetura CNN simplificada. Primeiramente, a camada de entrada recebe imagens padronizadas. Em seguida, um *kernel* convolucional é movido sobre essa imagem de entrada, até percorrer a imagem toda, resultando em um mapa de características da camada convolucional. Esse mapa de características consiste de cada um dos planos na camada convolucional da Figura 2.6. Em seguida, temos a camada de *pooling* que faz uma subamostragem desse mapa de características. Após uma sequência de estágios, constituídos por uma camada convolucional e uma camada de *pooling*, segue-se uma camada completamente conectada, responsável pela etapa de classificação. Essa camada completamente conectada é seguida pela camada de saída. O número de saídas nessa camada corresponde ao número de classes em que as amostras de um dado problema são classificadas.

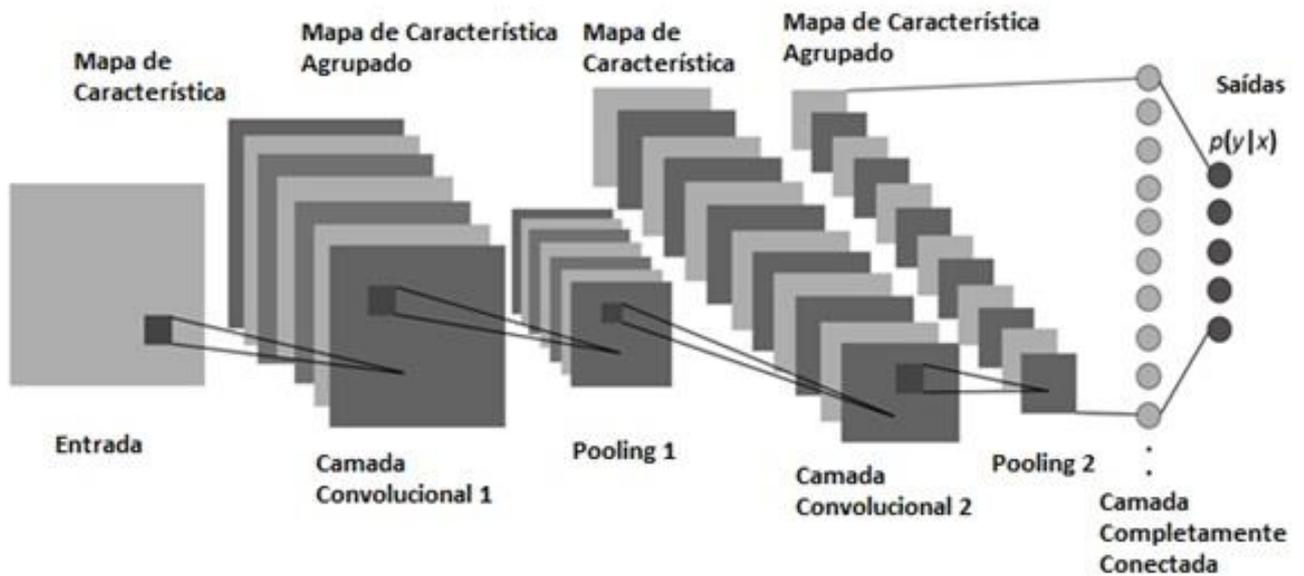


Figura 2.6 – Exemplo de arquitetura de rede convolucional.
Fonte: Adaptado de Mukherjee (2020).

2.2.2.1 Camada Convolutiva

Uma camada convolutiva compreende um conjunto de filtros (*kernels*) convolucionais que realiza uma operação de convolução na imagem. A operação de convolução consiste em sobrepor o *kernel* sobre uma região da imagem e calcular o produto entre os pixels do *kernel* e os pixels da região de abrangência do filtro na imagem, obtendo-se um único número representando a soma desses produtos. Essa operação de convolução ocorre sobre toda a imagem, movendo o *kernel* da esquerda para a direita e de cima para baixo, de acordo com o passo (*stride*) definido, até o filtro chegar no canto inferior direito da imagem. Após essa operação de convolução, esta camada obterá como resultado o mapa de características.

A operação de convolução pode ser expressa matematicamente como:

$$f_l^k(p, q) = \sum_c \sum_{x, y} i_c(x, y) \cdot e_l^k(u, v) \quad (2.1)$$

Em que:

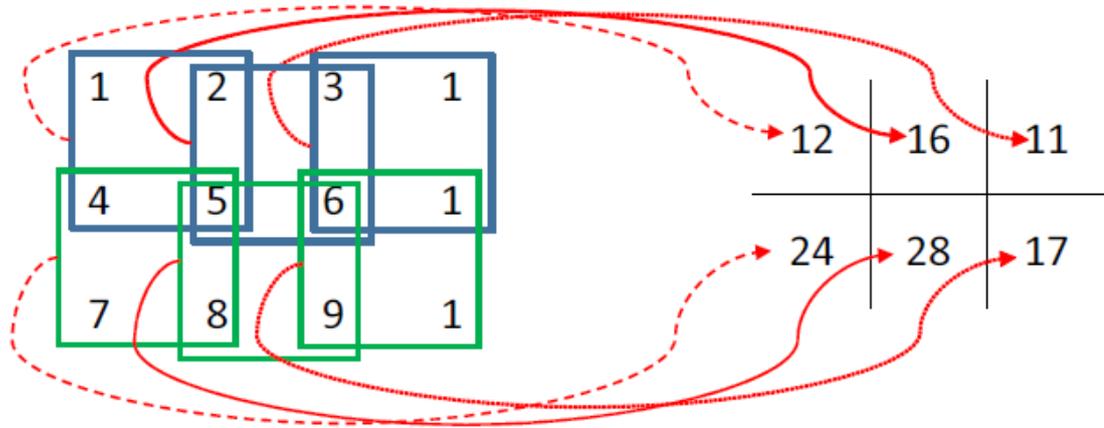
- $i_c(x, y)$ é um elemento da imagem de entrada I_c ;
- $e_l^k(u, v)$ é o elemento (u, v) do k -ésimo *kernel* convolutiva da l -ésima camada;
- u e v representam a u -ésima linha e v -ésima coluna do *kernel*, respectivamente;
- x e y são as coordenadas da imagem;
- c é o índice do canal da imagem;

O mapa de característica da k -ésima operação convolutiva pode ser expresso como $\mathbf{F}_l^k = [f_l^k(1,1), \dots, f_l^k(p, q), \dots, f_l^k(P, Q)]$. Em que p é a p -ésima linha e q é a q -ésima coluna de um total de P linhas e Q colunas.

A Figura 2.7 ilustra a operação de convolução: a) representa um exemplo de *kernel*, e; b) imagem de entrada e resultado da operação de convolução. A imagem de entrada mostra o *kernel* sobreposto. O resultado da operação de convolução é denominado de mapa de características.

$$\begin{array}{c|c} 1 & 1 \\ \hline 1 & 1 \end{array}$$

(a)



(b)

Figura 2.7 – Ilustração da operação de convolução: a) *kernel* 2x2; b) imagem de entrada e resultado da operação de convolução.

Fonte: Fonte (Wu, 2020).

Em CNNs, para realizar a operação de convolução, deve-se considerar os seguintes parâmetros: o tamanho do *Kernel*, o passo (*Stride*) e o espaçamento (*Padding*). O tamanho do *Kernel* é definido pela sua dimensão $k_h \times k_w$. O passo (*Stride*) consiste no número de linhas ou colunas em que o kernel convolucional é deslocado ao mover-se sobre a imagem, na direção horizontal ou vertical. O espaçamento (*Padding*) é a adição de pixels extras de preenchimento ao redor da borda da imagem, aumentando o tamanho da imagem. O *Padding* evita a perda de pixels na borda da imagem ao realizar uma operação de convolução.

A Figura 2.8 ilustra uma operação de convolução aplicando *padding* de uma linha e uma coluna de zeros nas quatro bordas da imagem de entrada de 3x3. Nessa operação foram utilizados um *kernel* de 2x2 e *stride* de 1. O resultado é uma imagem de saída de 4x4.

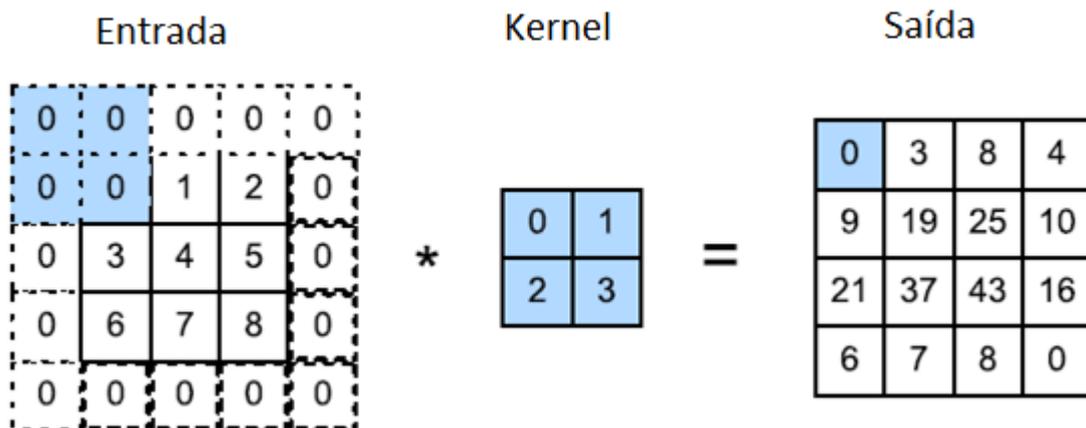


Figura 2.8 – Operação de convolução com *padding* e *kernel* 2x2. As partes sombreadas são a entrada e os elementos da matriz do *kernel* usados pelo primeiro elemento de saída: $0 \times 0 + 0 \times 1 + 0 \times 2 + 0 \times 3 = 0$.

Fonte: Zhang (2020).

Em geral, a camada de convolução tem como resultado final a geração de uma imagem de saída denominada de mapa de característica. Utilizando uma imagem com dimensões $n_h \times n_w$, e um kernel com dimensões $k_h \times k_w$, usando padding de p_h e p_w , nas direções horizontal e vertical, respectivamente, as dimensões dessa imagem de saída serão dadas por:

$$(n_h - k_h + p_h + s_h) \times (n_w - k_w + p_w + s_w) \quad (2.2)$$

2.2.2.2 Unidade Retificadora Linear - ReLU

Em uma CNN, as operações de convolução são normalmente intercaladas com a aplicação de uma função de ativação não linear ReLU - Unidade Retificadora Linear e por uma camada de subamostragem. A aplicação da função ReLU não altera as dimensões de uma camada, pois a mesma é aplicada na saída de cada neurônio da camada de convolução. O uso da função ReLU possui vantagens sobre outras funções de ativação, como a tangente hiperbólica e a sigmoide, em termos de velocidade e pelo fato da mesma não apresentar saturação dos valores positivos.

2.2.2.3 Camada de Subamostragem (*Pooling*)

A camada de subamostragem tem como finalidade reduzir de forma gradual a dimensionalidade da representação do mapa de característica, obtendo desta forma a redução do número de parâmetros e da complexidade do cálculo computacional da rede. A equação (2.3) mostra a operação de subamostragem.

$$\mathbf{Z}_l^k = g_p(\mathbf{F}_l^k) \quad (2.3)$$

Em que:

- \mathbf{Z}_l^k representa o mapa de características agrupado da l -ésima camada para a k -ésima entrada do mapa de características \mathbf{F}_l^k ;
- $g_p(\cdot)$ define o tipo de operação de subamostragem.

As operações de subamostragem mais comuns são o valor máximo (*max pooling*), o valor médio (*average pooling*) e o valor de soma (*sum pooling*), porém o mais utilizado é o valor máximo (*max pooling*). Na Figura 2.9 podemos visualizar um exemplo da operação $g_p(\cdot)$.

de subamostragem de valor máximo (*max pooling*), em que seleciona o valor máximo de uma sub-região de dimensionalidade 2x2 aplicado a um passo de 2 na imagem. Esses parâmetros reduzem a dimensão do mapa de características em 25%. Neste exemplo, o filtro da operação é deslizado sobre a imagem, como uma convolução, obtendo o maior valor do filtro como resultado.

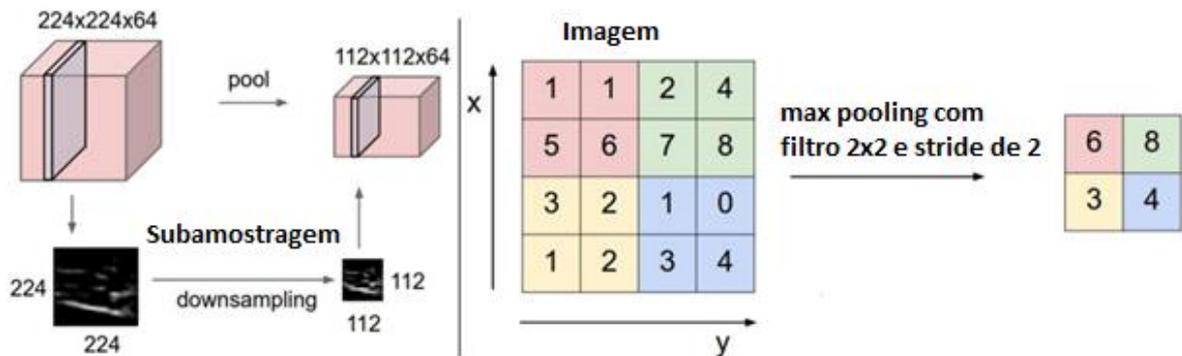


Figura 2.9 – Exemplo de *Max-Pooling*.
Fonte: Santos (2020).

2.2.2.4 Camada de Convolução Transposta

A camada de convolução transposta, também denominada de camada de deconvolução, funciona de forma contrária à camada de convolução, utilizando sobreamostragem da rede. Enquanto que a camada de convolução tem a função de comprimir a imagem de entrada, a convolução transposta descomprime. Nessa camada, o filtro da camada de convolução é usado de forma invertida horizontalmente e verticalmente. (Aggarwal, 2018).

Segundo Aggarwal (2018), a operação de deconvolução ocorre de forma semelhante à operação de convolução, porém o filtro é utilizado de forma invertida. O filtro invertido é movido sobre uma sub-região da imagem de entrada para executar o produto escalar, fazendo o movimento do filtro invertido semelhante ao movimento que ocorre na camada de convolução.

Na Figura 2.10 ilustramos um exemplo da convolução transposta. Pode-se notar a aplicação de um filtro 3x3 na imagem de entrada, com *padding* de 2. Aplicando a operação de deconvolução sobre essa imagem, utilizando-se um filtro invertido de 3x3 obtém-se como resultado uma imagem de 5x5. Desta forma, ampliamos a imagem de entrada de 3x3 para uma imagem resultante 5x5.

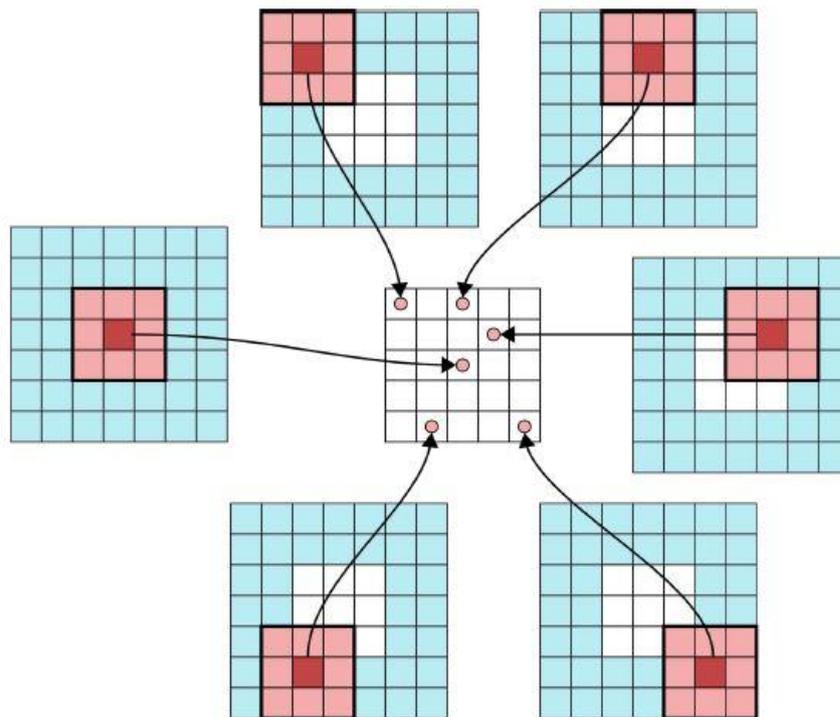


Figura 2.10 – Exemplo de uma convolução transposta.
Fonte: Glassner (2018).

2.2.2.5 Camada de Totalmente Conectada (*Full Connected Layer*)

A camada completamente conectada é geralmente usada no final da rede CNN com o intuito de realizar uma classificação. Diferentemente da camada de convolução e da camada de subamostragem, os neurônios dessa camada se conectam a todos os neurônios da camada anterior. Podemos afirmar que na CNN todas as camadas anteriores à camada inteiramente conectada são responsáveis por extração de características, enquanto a camada inteiramente conectada, associada à camada *softmax* são responsáveis pela classificação. Então, nessa camada é operacionalizada uma combinação não linear das características extraídas pelas camadas anteriores, responsável pela classificação dos dados.

2.3 ALGORITMO DE APRENDIZAGEM E HIPERPARÂMETROS

2.3.1 Algoritmo de aprendizagem

O Algoritmo de aprendizagem consiste no método utilizado para melhorar o desempenho de uma rede neural, através da otimização dos parâmetros de uma função objetivo. Esses parâmetros consistem basicamente dos pesos e polarizações das camadas. Esses algoritmos

são geralmente utilizados no treinamento da rede. Neste trabalho, os algoritmos de aprendizagem utilizados nos experimentos são SGDM, RMSProp e Adam. O Gradiente Descendente Estocástico com Momento (SGDM) é uma adaptação do algoritmo gradiente descendente estocástico (SGD) com a introdução de momento. O SGD é um dos métodos de otimização mais utilizado em redes neurais. No mesmo, para cada exemplo aplicado à entrada da rede, ocorre uma atualização dos parâmetros. Tal estratégia de otimização gera um processo muito demorado de convergência para uma solução. Para acelerar o desempenho desse algoritmo foi introduzido o termo momento, inspirado na primeira lei do movimento de Newton, com objetivo de propiciar uma convergência mais rápida, com menos oscilações. As trajetórias de otimização de um algoritmo SGD e de um algoritmo SGDM, para uma função simples, são mostradas na Figura 2.11. Observa-se nessa figura que o algoritmo SGDM atinge o ponto ótimo em menos passos e com menos oscilação que o algoritmo SGD.

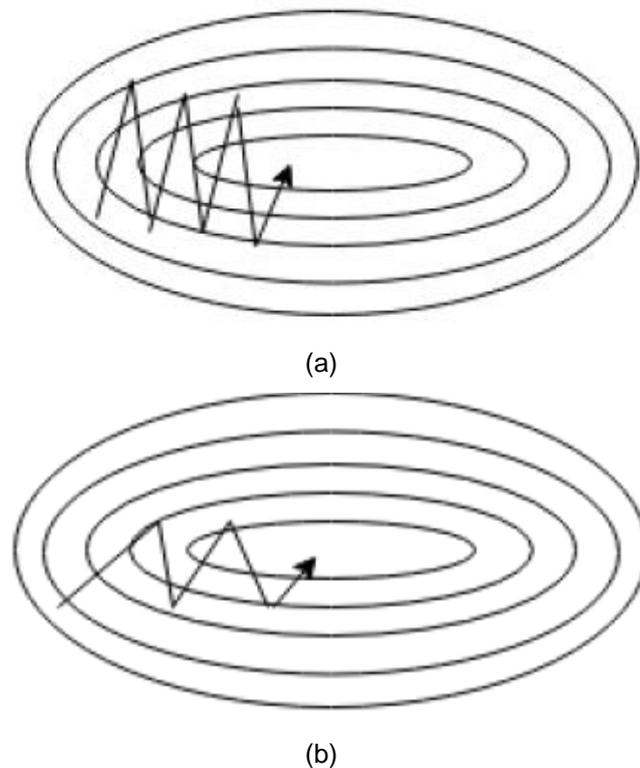


Figura 2.11 – Otimização de uma função simples: a) Trajetória de otimização do algoritmo SGD; b) Trajetória de otimização do algoritmo SGDM.
Fonte: Setiawan et al. (2018).

As equações (2.4) e (2.5) mostram como o momento é utilizado para acelerar o SGD na direção relevante, amortecendo as oscilações, como pode ser visto na Figura 2.11(b) (Ruder, 2017).

$$V_t \leftarrow \gamma V_{t-1} + \eta \left(\frac{\partial L}{\partial w_i} \right); \quad \forall i \quad (2.4)$$

$$w_i \leftarrow w_i - V_t; \quad \forall i \quad (2.5)$$

Em que:

- $\gamma \in (0,1)$ é o termo do momento;
- V_t é o valor do gradiente de primeira ordem da iteração presente suavizado exponencialmente;
- V_{t-1} é o valor do gradiente de primeira ordem da iteração anterior suavizado exponencialmente;
- $\eta \in (0,1)$ é a taxa de aprendizagem;
- w_i é o i -ésimo parâmetro da rede;
- L é a função de custo. Especificamente, a função de erro.

O *Root Mean Square Propagation* – RMSProp é um método de otimização baseado no algoritmo AdaGrad. O AdaGrad atualiza as taxas de aprendizagem individualmente, a cada iteração, para cada parâmetro do modelo, baseando-se na soma de todos os valores do gradiente ao quadrado, resultando, desta forma, na diminuição da taxa de aprendizagem. Isso faz com que o progresso do AdaGrad se torne muito lento. Além disso, os fatores de escalonamento agregados ficam estagnados, pois dependem dos valores históricos antigos, resultando no aumento da imprecisão. Já o RMSProp, para reduzir essa taxa de aprendizagem, utiliza uma média exponencial ponderada. A equação (2.6) mostra a expressão para calcular o método RMSProp (Aggarwal, 2018):

$$A_i \leftarrow \rho A_i + (1 - \rho) \left(\frac{\partial L}{\partial w_i} \right)^2; \quad \forall i \quad (2.6)$$

$$w_i \leftarrow w_i - \frac{\alpha}{\sqrt{A_i}} \left(\frac{\partial L}{\partial w_i} \right); \quad \forall i \quad (2.7)$$

Em que:

- $\rho \in (0,1)$ é uma constante que define o fator de decaimento;
- A_i é o valor médio exponencial do i -ésimo parâmetro w_i .
- L é a função de custo (ou erro);
- w_i é o i -ésimo peso da rede;

O *Adaptive Moment Estimation* – ADAM é outro algoritmo de otimização que calcula as taxas de aprendizagem para cada parâmetro, através de uma média exponencial

decrecente do primeiro e segundo momento do gradiente. A equação (2.8) mostra a suavização do gradiente de segunda ordem. A equação (2.9) mostra a suavização do gradiente de primeira ordem. Os dois gradientes são usados, então, para atualização dos pesos da rede na equação (2.10), (Aggarwal, 2018).

$$A_i \leftarrow \rho A_i + (1 - \rho) \left(\frac{\partial L}{\partial w_i} \right)^2; \quad \forall i \quad (2.8)$$

$$F_i \leftarrow \rho_f F_i + (1 - \rho_f) \left(\frac{\partial L}{\partial w_i} \right); \quad \forall i \quad (2.9)$$

$$w_i \leftarrow w_i - \frac{\alpha_t}{\sqrt{A_i}} F_i; \quad \forall i \quad (2.10)$$

Em que:

- F_i é o i -ésimo componente exponencialmente suavizado do gradiente de primeira ordem;
- $\rho_f \in (0,1)$ é o fator de decaimento;
- α_t é a taxa de aprendizagem na t -ésima iteração;

2.3.2 Hiperparâmetros

Os hiperparâmetros são as variáveis que configuram a arquitetura da rede e que afetam o processo de otimização na fase de treinamento. Os hiperparâmetros relacionados com a arquitetura da rede são: número de camadas ocultas, números de neurônios, funções de ativação, tamanho dos filtros, passo (*stride*), preenchimento com zeros (*padding*), taxa de *dropout* e inicialização dos pesos, entre outros. Por sua vez, os hiperparâmetros relacionados ao algoritmo de treinamento são: taxa de aprendizado, momento, número de épocas e tamanho do lote.

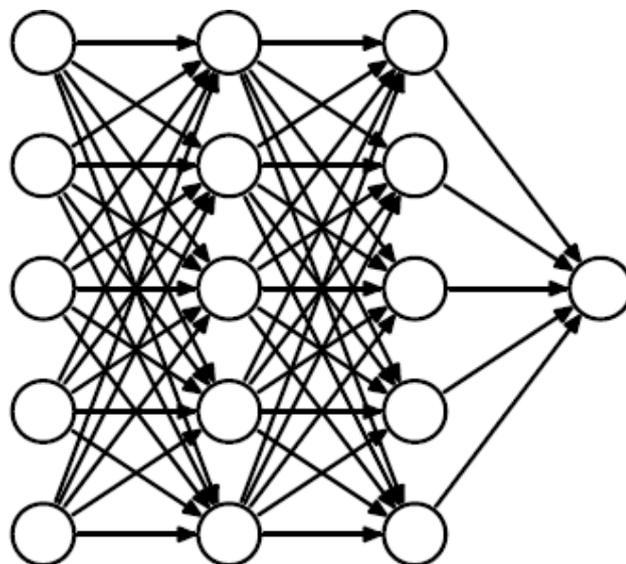
Segundo Loussaief e Abdelkrim (2018), um dos aspectos mais desafiadores da CNN é como encontrar os valores para os seus hiperparâmetros. Algumas abordagens utilizadas para encontrar os hiperparâmetros ótimos para as redes convolucionais são a busca exaustiva de todas as combinações possíveis ou o uso de heurísticas tais como o algoritmo genético, que encontram uma solução aproximada.

2.4 TÉCNICAS DE REGULARIZAÇÃO

Segundo Kim et al. (2016), regularização é qualquer modificação realizada em um algoritmo de aprendizagem com o objetivo de reduzir seu erro de generalização, mas não seu erro de treinamento. Por sua vez, generalização é capacidade da rede realizar uma classificação com ótimo desempenho para um conjunto de dados que não foram treinados, como por exemplo um conjunto de teste.

Entre os métodos de regularização mais utilizados destacamos o *Dropout* e o método de regularização L_2 . O primeiro desabilita neurônios em uma rede. O segundo adiciona uma penalidade aos pesos.

Para uma época do treinamento, a técnica de *dropout* remove, temporariamente da rede, neurônios e todas as suas ligações de entrada e saída. Essa remoção é feita de forma aleatória, com uma probabilidade predefinida. A Figura 2.12 ilustra como essa técnica funciona. Dessa forma, em cada época, é treinada uma rede inteiramente diferente da outra. O resultado é que os parâmetros da rede serão ajustados como se diversas arquiteturas de rede diferentes tivessem sido treinadas, aumentando o poder de generalização e simplificando um treinamento que, de outra forma, exigiria que diversas arquiteturas fossem treinadas separadamente.



(a)

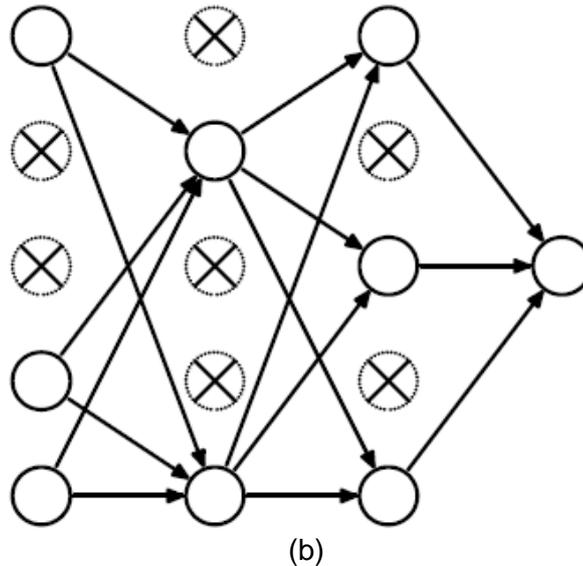


Figura 2.12 – Exemplo da aplicação da técnica dropout: a) Uma rede neural padrão com 2 camadas ocultas; b) exemplo de rede neural reduzida produzida pela aplicação de *dropout* na rede em (a). As unidades marcadas com um “x” foram descartadas.

Fonte: (Srivastava et al., 2014).

A regularização L_2 ou regularização de Tikhonov é uma técnica de generalização que utiliza uma penalidade adicionada aos pesos no treinamento da rede. Essa penalidade consiste em somar, na expressão a ser otimizada, os pesos elevados à segunda potência. A consequência verificada da introdução desses termos é obter valores menores para os pesos, evitando o sobre ajuste do mesmo aos dados do conjunto de treinamento, melhorando, assim, a generalização da rede. A equação (2.4) mostra uma expressão a ser otimizada, constituída do erro médio quadrático, à qual foram somados novos termos dos pesos elevados à segunda potência.

$$L = \sum_{(x,y) \in D} (y - \hat{y})^2 + \lambda \sum_{i=0}^d w_i^2 \quad (2.11)$$

Em que:

- (x, y) são instâncias do conjunto de treinamento D ;
- y é a saída desejada da rede;
- \hat{y} é a saída obtida da rede;
- λ é um parâmetro de regularização maior do que 0;
- w_i^2 é o quadrado do i -ésimo peso da rede;
- d é número de pesos da rede;

As atualizações dos pesos w_i ocorrem através do algoritmo SGD: cada valor do peso w_i é decrescido pelo produto entre a taxa de aprendizado e a derivada parcial da expressão (2.12) em relação ao peso w_i , conforme a equação a seguir:

$$w_i \leftarrow w_i - \alpha \frac{\partial L}{\partial w_i} \quad (2.12)$$

Em que:

- w_i é o peso a ser atualizado;
- α é a taxa de aprendizado;
- $\frac{\partial L}{\partial w_i}$ é a derivada parcial da função de otimização L em relação ao peso w_i .

2.5 SEGMENTAÇÃO SEMÂNTICA

A segmentação semântica é uma técnica de classificação da imagem em nível de pixel. Assim, cada pixel é classificado como pertencente a uma das classes definidas no problema de segmentação, que normalmente referem-se aos diferentes objetos presentes em uma imagem. Ao classificar os pixels de um objeto como pertencentes a uma classe distinta, a segmentação semântica propicia, na verdade, a segmentação dos diversos objetos presentes em uma imagem. A primeira rede de segmentação semântica foi proposta por Long et al. (2015). A Figura 2.13 mostra um exemplo de uma CNN com segmentação semântica. A primeira parte da rede, em que ocorre uma diminuição do tamanho dos mapas de ativação, é denominada de codificador (*encoder*). A segunda parte da rede, em que ocorre um aumento do tamanho dos mapas de ativação, é denominada de decodificador (*decoder*).

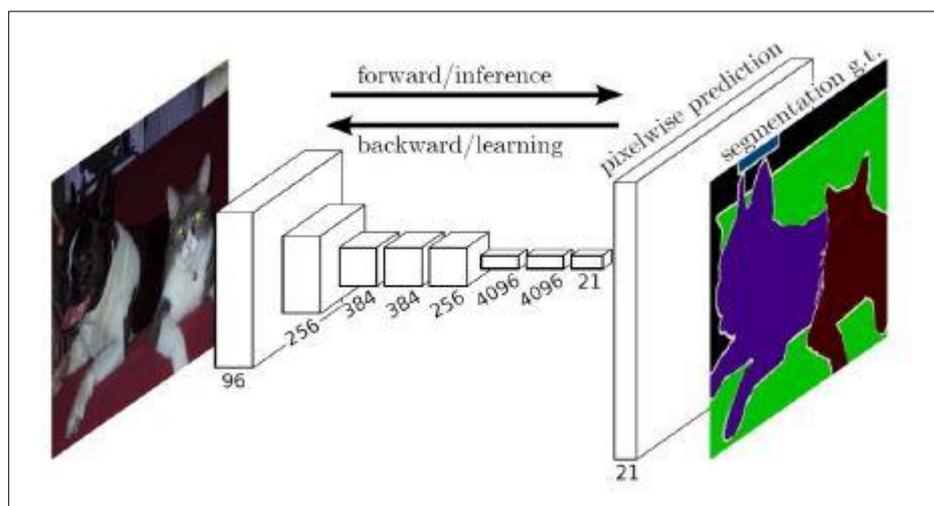


Figura 2.13 – Arquitetura CNN com segmentação semântica.
Fonte: Long et al (2015).

Algumas arquiteturas mais modernas como a SegNet (Badrinarayanan et al., 2015) utilizam uma conexão alternativa entre camadas de subamostragem e as camadas de

sobreamostragem, conforme mostrado na Figura 2.14. Segundo os autores, essa conexão alternativa propicia uma melhor precisão da segmentação semântica, preservando os detalhes de cada objeto. Na rede SegNet, conforme pode ser observado, não há camada convolucional totalmente conectada. A classificação dos pixels nas suas correspondentes classes é feita através da função *softmax* para múltiplas classes.

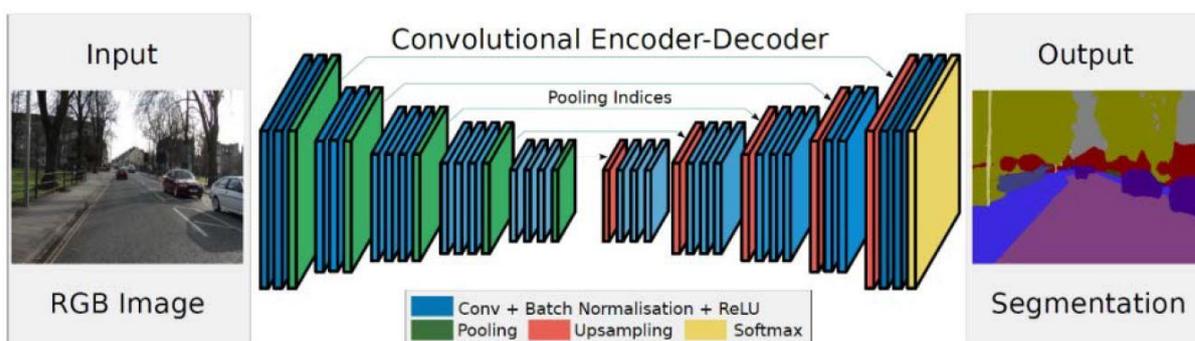


Figura 2.14 – Arquitetura da Rede Convolucional SegNet.
Fonte: Badrinarayanan et al. (2015).

2.6 CONSIDERAÇÕES

Neste Capítulo foram mostrados os fundamentos teóricos das técnicas utilizadas neste trabalho, com alguns comentários sobre como determinado conceito, técnica ou método se encaixa no escopo deste trabalho.

Foram mostrados dados recentes para contextualização do problema de desmatamento na Amazônia Legal Brasileira, identificando os estados com maiores taxas de desmatamento, bem como as dimensões territoriais do desmatamento. Situamos também as regiões que serão avaliadas nesse trabalho, a saber, sul do Estado do Amazonas e regiões vizinhas dos Estados do Mato Grosso, Rondônia, Acre e Pará. Diante desse contexto, ficou evidente a necessidade de monitoramento das atividades ilegais e predatórias que causam desmatamento, através de técnicas de Sensoriamento Remoto com imagens de satélite.

Apresentamos também a ferramenta de aprendizado de máquina a ser utilizada nesse trabalho para reconhecimento de áreas desmatadas, as redes convolucionais, com suas camadas, algoritmos de treinamento e principais parâmetros. Tais ferramentas, após treinadas, com exemplos de regiões positivas (áreas desmatadas) e de regiões negativas (áreas não desmatadas), podem identificar numa imagem ambos os tipos de regiões. Uma visão mais detalhada sobre redes convolucionais pode ser obtida através de uma consulta às referências indicadas neste capítulo.

No próximo capítulo, faremos uma revisão dos principais trabalhos da literatura referentes aos temas relacionados com o desmatamento florestal.

CAPÍTULO 3

3. REVISÃO BIBLIOGRÁFICA

Neste capítulo serão apresentados as principais técnicas, classificadores utilizados e os resultados mais significativos dos trabalhos na literatura para detecção de desmatamento em áreas de florestas. O intuito deste capítulo é mostrar ao leitor abordagens utilizadas pelos trabalhos publicados para que possamos tecer um comparativo no que o trabalho ora proposto se diferencia.

3.1 TRABALHOS RELACIONADOS

Nesta seção apresenta-se alguns trabalhos da literatura, publicados no período de 2015 a 2020, que abordam temas relacionados ao desmatamento florestal. No final dessa seção é apresentada a Tabela 3.1, mostrando um resumo das principais características desses artigos, tais como: a) objetivos – o que cada trabalho se propõe a fazer; b) materiais – constituídos pelo banco de imagens utilizados e resolução das mesmas; c) técnica de filtragem – compreende as máscaras utilizadas para identificação floresta/não floresta; d) classificador utilizado; e) medidas de desempenho utilizadas: acurácia, sensibilidade, especificidade, taxa de erro e taxa de acerto, entre outras.

A seguir são descritos os resumos dos trabalhos selecionados na Tabela 3.1.

Diniz et al. (2015) apresentaram a metodologia e os resultados do DETER-B (Sistema de Detecção de Desmatamento em Tempo quase Real) comparando com os dados do PRODES. Para construção do sistema foram utilizadas 291 imagens AWiFS do ano de 2013, dados ortorretificados do Landsat, os limites de bordas oficiais da Amazônia Legal Brasileira, correspondentes a área de estudo, e os dados do PRODES até 2013. A metodologia do sistema DETER-B baseia-se na interpretação visual e compreende as seguintes etapas: aquisição e composição da imagem, correção geométrica, Modelo Linear de Mistura Espectral (MLME), recuperação das áreas não observadas, geração da máscara PRODES, interpretação visual e auditoria. Para verificar a validação estatística do DETER-B, os autores compararam as classes de desmatamento e resíduos do PRODES com as seguintes classes do DETER-B: desmatamento, desmatamento com vegetação e mineração, considerando as diferenças metodológicas, espaciais e temporais de cada sistema. Para todos os estados da Amazônia Legal foram obtidas as seguintes áreas classificadas: 433.289,13 ha de

desmatamento e classes residuais e 404.898,87 ha de desmatamento com e sem vegetação e mineração, para o PRODES e DETER-B, respectivamente. Os resultados mostraram que os dados do DETER-B e PRODES 2013 são estatisticamente semelhantes e apresentaram um coeficiente de correlação (R^2) = 0,83, considerando 188 municípios, indicando uma forte correlação linear. Para cada um dos 188 municípios são calculadas as áreas desmatadas em ha obtidas pelo PRODES e pelo DETER-B. Os pontos formados por esses dois valores são plotados em um gráfico em que o eixo x indica os valores do PRODES e o eixo y, os valores obtidos pelo DETER. A partir disso, traça-se uma reta para a obtenção do coeficiente de correlação entre os pontos plotados. Os autores analisaram que ambos os sistemas são eficientes para detecção de desmatamento em áreas menores que 25 ha, apresentando aproximadamente 80% de suas detecções totais nessa faixa, representando quase 45% da área mapeada por cada sistema.

Joshi et al. (2015) apresentaram um método com objetivo de detectar perturbações florestais como desmatamento, degradação e uso da terra, utilizando trajetória de retro espalhamento do radar, e identificar áreas com recuperação lenta e rápida. Os autores definiram distúrbios florestais como a perda difusa da cobertura florestal por degradação e efeitos do corte raso por desmatamento. Esses distúrbios são estudados analisando as diferenças em retro espalhamento por radar. O local de estudo foi na região das províncias de Tahuamanu e Tambopata, localizadas na região de Madre de Dios – Amazônia Peruana. Os dados utilizados foram duas cenas da estação seca, da fase de banda L, adquiridas do sensor do Radar de Abertura Sintética a bordo do *Advanced Land Observing Satellite* (ALOS PALSAR), no período compreendido entre julho e agosto, de 2007 a 2010. Sobre esses dados utilizaram o filtro de Lee de janela 3x3 pixels para reter texturas das imagens com ruído. O algoritmo proposto no estudo consistiu nos seguintes passos: 1) utilização de dados multitemporais para verificar locais de perturbações; 2) identificação de áreas com retro espalhamento de recuperação lenta e rápida. Retro espalhamento foi definido relacionando a biomassa para as áreas e observando a diminuição da perturbação. Esse algoritmo utilizou um conjunto de limiares (T1 a T5) para identificar o pixel perturbado. Os autores obtiveram uma taxa de acerto de 63% na detecção de perturbações. A principal contribuição do estudo foi apresentar a utilidade do radar para detectar a dinâmica de desmatamento e degradação das florestas.

Barreto et al. (2016) propuseram o uso do método de segmentação de superpixel pelo Agrupamento Iterativo Linear Simples (SLIC – *Simple Linear Iterative Clustering*) combinado com a técnica de detecção de imagens por Correlação de Objeto e com um Perceptron multicamada, aplicadas a imagens SAR multitemporal de banda X. Os autores utilizaram 14 subconjuntos de imagens SAR banda X com polarização HH e resolução de 1m, da Floresta Atlântica no Estado de São Paulo – Brasil. No total, 1877 amostras foram coletadas e

rotuladas por interpretação visual nas seguintes classes: *Sem mudança* – azul (877); Desmatamento – vermelho (527), e; Outras Alterações – verde (473). Os autores utilizam o índice de coeficiente Kappa, a taxa positiva verdadeira, a taxa falsa positiva e área de curva ROC para avaliar seus resultados, no qual obtiveram 66%, 79%, 13% e 90%, respectivamente. Nisto, concluíram que resultados obtidos com o modelo MLP mostram uma melhoria de 60,84% para 78,69% em termos de amostras classificadas corretamente (CCS), apresentando um percentual CCS de 95% para a classe *Sem mudança*.

Rosa et al. (2017) apresentaram um método de detecção em imagens SAR multitemporais, banda X e P simultaneamente, com objetivo de criar uma ferramenta eficiente para identificar regiões onde ocorre algum tipo de alteração na superfície e produzir máscaras binárias para criação de vetores cartográficos. Os autores utilizaram coerência cruzada, para obter o indicador de detecção de mudanças das imagens SAR, e o algoritmo de Agrupamento Iterativo Linear Simples (SLIC) para criar superpixels das imagens da banda X. Depois de criados os superpixels, os atributos são calculados e utilizados como entrada de uma Rede Neural Perceptron Multicamada (MLP), que detecta mudanças nas imagens. Foi utilizada validação cruzada com 10 pastas para evitar *overfitting*. Os autores avaliaram a rede através da matriz de confusão. A entrada da rede foi obtida através da aplicação de Análise de Componentes Principais (PCA) a três atributos calculados de superpixels R_t^a , R_{t+1}^b e R_{t+1}^c , que representam a coerência cruzada entre imagens de diferentes bandas. Os autores concluíram que o método do super pixel é muito bom para reduzir os falsos alarmes criado por ruídos salpicados. Os autores obtiveram 80% de acurácia para detecção de mudança.

Oliveira et al. (2017) avaliaram no seu estudo o desempenho de diversos métodos aplicados para a melhoria da generalização de redes neurais tipo Perceptron Multicamadas, no reconhecimento de áreas desmatadas na região sul do estado do Amazonas. Para entrada da rede foram utilizados três métodos de extração de características: Análise de Componentes Principais (PCA – *Principal Component Analysis*), Análise Discriminante Linear (LDA – *Linear Discriminant Analysis*) e Seleção Escalar de Características (SEC). Os dados utilizados foram imagens Landsat 8 e máscara de áreas desmatadas e não desmatadas do projeto PRODES, disponibilizado pelo INPE. As seguintes características foram extraídas a partir das imagens: métricas de textura como média, variância e desvio padrão; homogeneidade; energia; correlação; contraste e entropia. Como métodos para melhoria da generalização da rede foram utilizados parada antecipada, regularização e MSE para o treinamento. Os autores utilizaram o método de Levenberg-Marquardt para acelerar o treinamento da rede. Os parâmetros utilizados para avaliar o método de classificação foram área sob a curva ROC e a acurácia. Os resultados mostraram que a técnica de seleção LDA apresentou um melhor desempenho comparada à SEC e à PCA. Na técnica LDA os autores observaram que o número de características utilizadas não tem influência sobre o

desempenho do classificador. Em relação às técnicas para melhoria da generalização, os melhores resultados foram obtidos com o método de parada antecipada. O melhor resultado de classificação foi: área sob a curva ROC de 0,997 e acurácia de 0,97.

Bullock et al. (2018) propuseram uma metodologia que fosse capaz de mapear e estimar a degradação florestal e desmatamento de grandes áreas, utilizando uma série temporal de dados do Landsat para monitoramento contínuo. As imagens Landsat foram transformadas em frações espectrais (GV – *Green Vegetation*, NPV – *Non-Photosynthetic Vegetation*, Solo, Sombra e Nuvem), utilizando a função *unmix* do Google Earth Engine, para calcular o índice de fração de degradação normalizado (NDFI). A área de estudo foi o estado de Rondônia. A metodologia consistiu em ajustar o modelo de regressão para as frações espectrais e o NDFI. Os coeficientes obtidos do modelo de regressão anterior foram utilizados como entrada do classificador de Floresta Randômico (*Random Forest*), para obter as máscaras de Floresta e Não-Floresta. A partir dos pixels classificados como Floresta é calculado o NDFI futuro para o período monitorado. Se o valor de cinco observações NDFI é menor do que o limiar de mudança (-5), temos uma indicação positiva para Detecção de distúrbios. Em seguida, caracteriza-se a pós-perturbação da cobertura da terra com o modelo de regressão para o NDFI e as frações espectrais, resultando em duas classificações: Floresta e Não-Floresta. Com esta metodologia, os autores obtiveram acurácia do produtor (que é a proporção da classe de referência classificada como classe mapeada) de 68,1% para degradação e 85,3% para desmatamento. Os autores concluíram que os dados históricos do Landsat podem ser efetivamente usados para monitorar e estimar a degradação florestal.

Costa et al. (2019) realizaram um estudo para avaliar a precisão de técnicas de detecção de corte seletivo e a extensão espaço-temporal nas áreas de florestas impactadas por essas atividades, utilizando dados de sensoriamento remoto. As áreas de estudo compreenderam três estados da Amazônia Legal Brasileira: Mato Grosso, Pará e Rondônia. Essas áreas estão localizadas espacialmente dentro de 5 cenas de imagens Landsat. A metodologia proposta pelos autores foi baseada na aplicação de uma técnica semiautomática que usou em conjunto o algoritmo de textura para detectar pátios de aterrissagem de toras de madeira e a inspeção visual para mapear florestas registradas seletivamente, mostrando claramente as florestas em degradação através das imagens Landsat. Os autores criaram uma máscara de floresta e não floresta a partir do banco de dados de desmatamento do PRODES para mapear as terras de desmatamento, pois o estudo se concentra apenas em áreas florestais. A classificação das imagens Landsat foi inspecionada por interpretação visual em 12 imagens RapidEye de 5m de resolução. Os autores utilizaram matriz de confusão para estimar a acurácia do produtor, erros de comissão e omissão e precisão da técnica de mapeamento de corte seletivo. A acurácia do produtor da classe j é a proporção da área de classe de referência j que é mapeada como classe j . Neste caso, a classe em questão é a de

desmatamento. Eles obtiveram o valor de acurácia do produtor de 88,1%, 92,6% e 98% para áreas localizadas no estado do Mato Grosso, Pará e Rondônia, respectivamente. Os autores também utilizaram o coeficiente Kappa, como medida de resultado, apresentando um valor acima de 80%. Esse valor foi considerado uma boa concordância entre a imagem validada em campo e a classificação de corte seletivo.

Shimabukuro et al. (2019) propuseram um método semiautomatizado para monitoramento de desmatamento e degradação florestal na Amazônia brasileira usando imagens de fração multitemporal procedentes do modelo de mistura espectral linear (LSMM - *Linear Spectral Mixing Model*). O modelo LSMM é usado para estimar a fração dos componentes de solo, vegetação e sombra em cada pixel da imagem. A principal contribuição do trabalho desses autores consiste em separar explicitamente a degradação florestal causada por atividades de corte seletivo de exploração e degradação florestal causada por incêndios não controlados. A abordagem empregada foi a de mapear áreas desmatadas e de degradação florestal causadas por incêndios através de técnicas de segmentação de imagens por serem áreas contínuas e tamanho maior que a resolução do pixel. Porém, para mapear as áreas de degradação por atividade de corte seletivo foi usada a abordagem baseada em pixel, identificando desembarque de toras e estradas de toras. Os autores utilizaram 21 imagens Landsat TM do período compreendido entre 2005 e 2011, 15 imagens Landsat OLI do período compreendido entre 2013 e 2017 e duas imagens RapidEye (com datas próximas das imagens OLI de 2013) com resoluções de 30m, 30m e 5m, respectivamente. A área de estudo corresponde à parte da cena (path/row 226/068) localizada no estado do Mato Grosso. Com o objetivo de comparar com os resultados das imagens RapidEye como uma referência, o resultado da classificação do OLI foi reamostrado para 5m de resolução espacial usando a função de interpolação bilinear do pacote raster do R. Os autores aplicaram edição manual por interpretação visual à classificação de referência, assumindo que essa classificação pode ser tratada como padrão ouro de campo. O método proposto pelos autores obteve acurácias de 67% e 68%, respectivamente, para a classificação de áreas degradadas por incêndios e corte seletivo.

Ortega et al. (2019) realizaram um estudo comparativo para avaliar métodos de detecção automática para desmatamento com Rede Convolutacional de Fusão Precoce (EF), Rede Neural Convolutacional Siamesa (S-CNN) e Máquina de Vetores de Suporte (SVM). Os autores avaliaram esses métodos em uma região do Estado do Pará, localizado na Amazônia Legal Brasileira. Esses autores utilizaram em seus experimentos duas imagens Landsat 8 dos anos 2016 e 2017, com tamanho de 1.100 x 2.600 pixels e sete bandas espectrais: Costeira/Aerossol, Azul, Verde, Vermelho, Infravermelho próximo (NIR - *Near-Infrared*), Infravermelho de ondas curtas (SWIR - *Shortwave Infrared*)-1 e SWIR-2. Utilizaram a métrica F1 para avaliação e desempenho dos métodos. Em um dos experimentos obtiveram valores

de F1-Score em torno de 52%, 62% e 65% para SVM, EF e S-CNN, respectivamente, e acurácias em torno de 96%, 97% e 98% para SVM, EF e S-CNN, respectivamente. Os métodos melhoram o desempenho à medida que o número de amostras de treinamento aumenta. A S-CNN apresentou o melhor resultado, identificando melhor as áreas desmatadas. Nisso, os autores concluíram que mais pesquisas são necessárias para obter resultados viáveis, como, por exemplo, o ajuste dos hiper parâmetros dos métodos estudados.

Adarme et al. (2020) avaliaram técnicas baseadas em *Deep-Learning* aplicadas à detecção de desmatamento em duas regiões tropicais com diferentes padrões de desmatamento: os biomas Amazônia e Cerrado. As estratégias baseadas em Aprendizagem Profunda (*Deep Learning* – DL) foram a Fusão Antecipada (*Early Fusion* – EF), Rede Siamesa (SN – *Siamese Network*) e Máquina de Vetores de Suporte Convolutiva (CSVM – *Convolutional Support Vector Machine*). Os autores usam Máquina de Vetores de Suporte (SVM – *Support Vector Machine*) como linha de base. O principal objetivo é reduzir a intervenção humana na detecção de áreas desmatadas, em programas de monitoramento de desmatamento florestal. Os autores utilizaram, em seus experimentos, duas imagens Landsat 8-OLI adquiridas em datas diferentes, de tamanho 1.100 x 2.600 pixels, fazendo uso das sete bandas espectrais: Costeira/Aerossol, Azul, Verde, Vermelho, Infravermelho próximo (NIR – *Near-Infrared*), Infravermelho de ondas curtas (SWIR – *Shortwave Infrared*)-1 e SWIR-2. Além disso, foi calculado o Índice de Vegetação das Diferenças Normalizadas (NDVI) e adicionado a cada imagem, como um oitavo componente. Os dados de desmatamento utilizados foram as máscaras do projeto PRODES. As imagens utilizadas foram divididas em 15 fragmentos de tamanho 15 por 15 pixels. Para cada imagem dividida, foram separados 4 fragmentos para treinamento, 2 para validação e 9 para teste. Ao final, os autores obtiveram para o bioma Amazônico os seguintes valores de acurácia: 96,2%, 97,8%, 98%, 96,2%, para SVM, EF, SN e CSVM, respectivamente. Também foi obtido os seguintes valores F1-Score: 52,4%, 63,2%, 63% e 52,3% para SVM, EF, SN e CSVM, respectivamente.

De Bem et al. (2020) propuseram o uso de CNNs na detecção de desmatamento na Amazônia brasileira, comparando os algoritmos de DL com os algoritmos clássicos de aprendizado de máquina aplicados ao mapeamento de desmatamento. As três CNNs usadas foram U-Net, SharpMask e ResUnet. Os métodos de aprendizado de máquina usados foram Floresta Aleatória (*Random Forest* – RF) e Perceptron Multicamada (*Multi Layer Perceptron* – MLP). Os autores utilizaram duas cenas para treinamento e uma cena para teste de imagens LandSat 8/OLI da Amazônia Legal brasileira, do mês de julho de 2017, 2018 e 2019. Os autores extraíram 844 amostras de fragmentos das três cenas, cada amostra tendo um tamanho de 200x200 pixels e uma sobreposição de 10 pixels em cada lado por local, por sequência de tempo, com um total de 3.376 amostras de treinamento e 1.688 amostras de teste. Os autores utilizaram como padrão-ouro as máscaras de desmatamento PRODES para

os anos de 2018 e 2019. O modelo ResUnet obteve os melhores resultados na maioria das métricas de desempenho, exceto para a métrica precisão. Os valores obtidos por esse classificador foram os seguintes: F1-Score de 0,9432, precisão de 0,9252, sensibilidade de 0,9619 e acurácia global de 0,9993, para o período de 2017-2018, e F1-Score de 0,9465, precisão de 0,9358, sensibilidade de 0,9574 e acurácia global de 0,9990, para o período de 2018-2019.

Tabela 3.1 – Resumo dos artigos selecionados.

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
2015/ Diniz et al.	<i>DETER-B: The New Amazon Near Real-Time Deforestation Detection System</i>	Apresentar a metodologia e os resultados do DETER com base nos dados do AWIFS, chamados DETER-B para detecção de desmatamento e outras classes.	Dados: 291 imagens AWIFS do LandSat das bandas 5, 4 e 3; dados do PRODES até 2013;	Amazônia Legal Brasileira	Fração de solo, vegetação e sombra. Máscara Prodes (Floresta/Não Floresta).	Modelo Linear de Mistura Espectral (MLME).	Detecção de 433.289,13 ha e 404.898,87 ha para o PRODES e DETER. Com a remoção das áreas discrepantes, AC, e TO e MA, ficaram: 406.963,69 ha e 373.529,83 ha para o PRODES e DETER. Obtiveram um coeficiente de correlação entre o PRODES E DETER de: $(R^2) = 0,83$
2015/ Joshi et al.	<i>Mapping Dynamics of Deforestation and Forest Degradation in Tropical Forest Using Radar Satellite Data</i>	Apresentar um método para detectar a dinâmica na mudança na cobertura de floresta, incluindo degradação, desmatamento e uso da terra.	Dados: Duas cenas da estação seca da banda L do Radar ALOS PALSAR, de julho-agosto dos anos 2007 a 2010. Resolução de 90 m.	Províncias de Tahuamanu e Tambopata, região de Madre Dios, na Amazônia peruana.	Filtro aprimorado de Lee. Janela de 3x3 pixels	Algoritmo de detecção utilizando Limiares para detecção de perturbação em floresta: T1 = -1,5 dB T2 = 56% T3 = 2 pixels T4 = 67.5% T5 = 3 pixels	Foram detectadas perturbações em 1721 de 2.740 pixels. Taxa de Acerto: 63% Taxa de Erro: 27%

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
2016/ Barreto et al.	<i>Deforestation Change Detection Using High-Resolution Multi-Temporal XBAND SAR Images and Supervised Learning Classification</i>	Propõe uma nova abordagem de classificação de mudanças capaz de distinguir as mudanças ocorridas em diferentes classes, além da classificação usual binária com ou sem alteração.	Os autores utilizaram 14 subconjuntos de imagens SAR banda X com polarização HH e resolução de 1m, 1877 amostras foram coletadas e rotuladas por interpretação visual nas seguintes classes: <i>Sem mudança</i> - azul (877); Desmatamento - vermelho (527), e; Outras Alterações - verde (473).	Floresta Atlântica no Estado de São Paulo – Brasil	Máscara de segmentação a partir do coeficiente das imagens de correlação de objeto.	Perceptron Multicamada	Índice de Coeficiente Kappa = 66%; Taxa positiva verdadeira = 79%; Taxa falsa positiva = 13%; Área de curva ROC = 90%.
2017/ Rosa et al.	<i>Deforestation Detection in Amazon Rainforest with Multitemporal X-Band and P-Band SAR Images using cross-coherences and super pixels.</i>	Desenvolver um algoritmo para detectar alterações na superfície em um intervalo de tempo, usando imagens SAR espaçadas temporariamente	Dados: Imagens SAR Multitemporal das bandas X e P	Floresta Amazônica (Porto Velho, Brasil)	Máscara Binária	Algoritmo Iterativo Linear Simples – SLIC Superpixels SLIC Rede Neural com Perceptron Multicamada (MLP)	Acurácia: 0,8

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
2017/ Oliveira et al.	<i>Detecção de áreas desmatadas utilizando redes neurais artificiais associadas a métodos de generalização.</i>	Reconhecer áreas desmatadas na região sul do Amazonas.	Dados: Imagens ópticas Landsat 8 (Bandas 2,3,4,5,6,7) <i>Shapes</i> do PRODES do ano de 2015 do INPE.	Região Sul do Amazonas: Apuí, Humaitá e Lábrea	Máscara de áreas desmatadas e não desmatadas do PRODES	Técnicas de extração de características: SEC, PCA e LDA Técnica de regularização: MSE, Parada Antecipada e Regularização. Classificador: Redes Neurais Artificiais.	Acurácia de 0,97; Área sob Curva ROC de 0,997.
2018/ Bullock et al.	<i>Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis</i>	Desenvolver uma nova metodologia que seja capaz de mapear e estimar uma degradação e o desmatamento em grandes áreas.	Série Temporal de Imagens Landsat (1990 – 2013)	Estado de Rondônia.	Função <i>unmix</i> do GEE para gerar as frações GV, NPV, Solo, Sombra e Nuvem	Classificador de Floresta Aleatória	Acurácia do produtor: 68,1% para degradação florestal e de 85,3% para desmatamento
2019/ Costa et al.	<i>Selective Logging Detection in the Brazilian Amazon</i>	Avaliar a precisão de técnicas de detecção de corte seletivo usadas na Amazônia.	Dados: Cinco Imagens Landsat-5 (2003,2006 e 2010) e Landsat-8 (2014) 12 Imagens RapidEye (2014) com resolução de 5m;	Estados do Mato Grosso (MT), Pará (PA) e Rondônia (RO)	Máscara de Floresta/Não-Floresta (PRODES)	Técnica semiautomática que utiliza algoritmo de textura e Interpretação Visual	Acurácia do produtor 88,1%, 92,6% e 98%, e coeficiente de Kappa de 82%,82% e 85,3%, para os Estado do MT, PA e RO, respectivamente.

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
2019/ Shimabukuro et al.	<i>Monitoring deforestation and forest degradation using Multi-temporal fraction images derived from Landsat Sensor data in the Brazilian Amazon</i>	Apresenta um procedimento semiautomático para o monitoramento de desmatamento e degradação florestal na Amazônia. Principal contribuição é separar explicitamente degradação florestal causada por corte seletivo e incêndio florestal.	Dados: 21 Imagens sem nuvens Landsat TM entre 2005 – 2011; 15 Imagens OLI entre 2013 a 2017	Imagem Landsat Thematic Mapper™ - path/row 226/068 do Estado do Mato Grosso.	Segmentação de Imagem Multitemporal. Utilizou os valores de pixels de fração de solo, sombra e vegetação. Valor de Pixel TM (R5, G4e B3)	Técnica de Limiarização	Acurácia: 67% para área degradada por queimada, e; 68% para área degradada por corte seletivo.
2019/ Ortega et al.	<i>Evaluation of Deep Learning Techniques for Deforestation Detection in the Amazon Forest.</i>	Avaliar e comparar o desempenho de 3 métodos de detecção de desmatamentos.	Dados: duas imagens Lansat 8 de 2016 e 2017.	Amazônia Legal Brasileira Coordenadas: (03 17 '23 "S e 050 55' 08" W)	Não Informado.	Rede convolutiva de Fusão Antecipada (EF) Rede Convolutiva Siamesa (S-CNN) Máquina de Vetores de Suporte (SVM)	F1 = 52%, 62% e 65% para SVM, EF e S-CNN, respectivamente. Acurácia = 96%, 97% e 98% para SVM, EF e S-CNN, respectivamente

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
2020/ Adarme et al.	<i>Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery</i>	Avalia técnicas baseadas em <i>Deep-Learning</i> aplicadas à detecção de desmatamento, visando reduzir a intervenção humana envolvido em programas de monitoramento do desmatamento na Amazônia e no Cerrado.	Duas cenas Landsat 8-OLI. Imagens de 1.100x2.600 pixels e sete bandas de 1 a 7. Imagens divididas em 15 blocos: a) quatro para treinamento (8.118 amostras para cada classe); b) um para validação (com 2.889 amostras por classe), e; c) o restante para teste (com 40.392 amostras de desmatamento e 1.675.608 amostras para não desmatamento).	Na Amazônia Legal, a área de estudo está localizada no estado do Pará, Brasil, centralizada nas coordenadas 03 03 '23 "sul e 050 55' 08" oeste.	Foram usados os dados do PRODES como máscara do padrão-ouro.	Três abordagens baseadas em DL: Fusão Antecipada (EF), Rede Siamesa (SN), e SVM Convolutacional (CSVM). Uma Técnica de ML: Máquina de Vetores de Suporte (SVM – <i>Support Vector Machine</i>), usada como linha de base.	Acurácias de 96,2%, 97,8%, 98%, 96,2%, para SVM, EF, SN, CSVM, respectivamente. F1-Score de 52,4%, 63,2%, 63% e 52,3% para SVM, EF, SN, CSVM, respectivamente
2020/ De Bem et al.	<i>Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks</i>	Investigar o uso de CNNs na detecção de desmatamento na Amazônia brasileira para verificar a hipótese de que os algoritmos DL (<i>Deep Learning</i> – Aprendizagem Profunda) são viáveis e	Três cenas Landsat 8/OLI para cada local (A, B e C) para os anos de 2017, 2018 e 2019, totalizando 9 cenas. Path/row de 227/63 (A), 227/65 (B) e 230/65 (C). Bandas de 1 a 7. As imagens foram empilhadas em cubos bi-temporal da imagem base e a imagem do próximo ano, constituindo 14 bandas.	Três regiões na Amazônia Legal ao longo da rodovia Transamazônica (BR-230) e rodovia “Cuiabá – Santarém” (BR-163). Duas delas	Foram usados os dados do PRODES como máscara do padrão-ouro.	Três classificadores DL, arquiteturas CNN: 1) SharpMask; 2) U-Net, e; 3) ResUnet Dois classificadores ML: 1) Floresta Aleatória (RF); 2) Perceptron Multicamada (MLP);	Melhores resultados para a ResUnet, exceto precisão: 1) Para o período de 2017-2018: F1-Score = 0,9432, Precisão = 0,9252; Sensibilidade = 0,9619, e; Acurácia Global = 0,9993; Para o período de 2018-2019: F1-Score = 0,9465,

Ano/Autor	Título	Objetivos	Materiais	Área de Estudo	Máscara ou Filtragem	Classificador	Medida de Desempenho e Resultado
		<p>possivelmente melhor alternativa em comparação com os algoritmos clássicos de ML (<i>Machine Learning</i> – Aprendizagem de Máquina) para mapear o desmatamento</p>	<p>Treinamento: 3.376 amostras de duas cenas (loais A e B); Teste: 1.688 amostras (local C), uma cena. Amostra de 200x200 pixels, com 10 pixels de sobreposição;</p>	<p>(loais A e B) no Estado do Pará, e uma (local C), no Estado do Amazonas.</p>		<p>Parâmetros: Para RF: 500 árvores de 3 variáveis; Para o MLP com três camadas: 256 neurônios na camada escondida; Para o MLP e métodos DL: Algoritmo de otimização: NADAM; Taxa de aprendizagem de 2×10^{-3}; Taxa de decaimento exponencial para o primeiro β_1 e β_2 momentos, de 0,9 e 0,999, respectivamente. Número de épocas: 250; Tamanho do lote: 16;</p>	<p>Precisão de 0,9358; Sensibilidade = 0,9574, e; Acurácia Global = 0,9990;</p>

3.2 CONSIDERAÇÕES E CONTRIBUIÇÃO DESTE TRABALHO

A partir da dos trabalhos revisados na seção anterior, podemos concluir que:

1. Os dados de desmatamento do PRODES e as imagens LANDSAT-8 são dados confiáveis e diversas pesquisas foram feitas usando esses dados com o objetivo de detecção de diversos usos do solo, conforme podemos notar no trabalho de Diniz et al. (2015);
2. Alguns trabalhos de detecção de áreas desmatadas fizeram uso combinado de técnicas de processamento digital de imagens com uso de redes neurais artificiais, como são os casos de Barreto et al. (2016), Rosa et al. (2017) e Oliveira et al. (2017), mostrando que os estudos com redes neurais está cada vez mais avançado. Em especial, Oliveira et al. (2017) obteve acurácia acima de 90% utilizando uma rede neural com técnicas de extração de características das imagens;
3. Os trabalhos que fizeram somente o uso de métodos de processamento digital de imagens, como segmentação por limiar, obtiveram o valor de acurácia em torno de 70%, como é o caso do Shimabukuro et al. (2019);
4. Trabalhos como os de Costa et al. (2019) e Shimabukuro et al. (2019) validam a classificação em imagens LandSat com a inspeção visual de imagens de alta resolução como a RapidEye de 5m de resolução;
5. Os trabalhos mais recentes que fizeram uso de técnicas de *Deep-Learning* para classificação obtiveram as melhores acurácias acima de 97%, como os casos de Adarme et al. (2020) e De Bem et al. (2020), usaram sete ou mais bandas. Esses poucos trabalhos aplicados na área de sensoriamento remoto motivam a exploração de mais estudos aplicado a esse nicho.

Dessa forma, nossa proposta de classificação de desmatamento com novas arquiteturas de CNN constitui-se em uma importante contribuição para a área, na medida em que o desempenho de redes convolutivas em outras áreas que utilizam processamento de imagens tem se mostrado bem superior ao desempenho de outras técnicas de reconhecimento de padrões tradicionalmente empregadas, mormente quando se dispõe de uma grande quantidade de dados para treinamento.

No trabalho ora proposto, não se fará uso de técnicas de extração de características como foi utilizado por Oliveira et al. (2017), pois as redes convolucionais possuem diferentes tipos de camadas, que podem ser combinadas, de forma que, ao mesmo tempo que extrai as características, as mesmas são armazenadas em mapas de características na própria rede, dispensando, assim, as técnicas de extração. Além disso, no nosso trabalho usaremos

apenas seis bandas, enquanto nos trabalhos de Adarme et al. (2020) e De Bem et al. (2020) foram utilizadas mais de seis bandas.

Enfatizamos novamente que outra novidade deste trabalho é a aplicação da técnica mosaico para construção de uma base de dados. A partir de imagens satelitais e dos dados geoespaciais de desmatamento são gerados fragmentos de amostras de áreas desmatadas e áreas de florestas, que serão as classes consideradas neste trabalho, para a geração das imagens-mosaico que servirão de entrada para as redes convolucionais utilizadas neste trabalho, sendo este um dos trabalhos precursores com aplicação desta abordagem para a área de sensoriamento remoto.

4. METODOLOGIA

A proposta metodológica deste trabalho tem por objetivo caracterizar a aplicação de redes CNNs na segmentação de áreas desmatadas por corte raso. Serão propostas três arquiteturas de CNNs. O treinamento, a validação e o teste dessas arquiteturas serão realizados utilizando-se imagens mosaico. O conceito de imagens mosaico foi introduzido no trabalho de Serrão et al. (2020), com o objetivo de treinar redes CNNs para o reconhecimento de bacilos da tuberculose. Neste trabalho, o conceito de imagens mosaico se refere a agregação em um retângulo maior, de um conjunto de fragmentos de amostras com tamanho de 40x40, extraídas de imagens satelitais, correspondendo tanto a áreas desmatadas como a áreas de floresta. Ressaltamos ser esse o primeiro trabalho em que imagens mosaico são propostas para treinamento de redes CNNs visando a identificação de áreas específicas em imagens de satélite. Nesse capítulo, será descrito com mais detalhes o processo de construção dessas imagens mosaico. Serão apresentados também os materiais, as arquiteturas das CNNs e as medidas utilizadas para avaliação do desempenho das mesmas.

4.1 MATERIAIS

4.1.1 Imagens Utilizadas

As imagens utilizadas foram obtidas pelo sensor OLI do satélite LANDSAT-8 conforme as cenas mostradas na Tabela 4.1. Essas cenas cobrem as regiões do sul do Estado do Amazonas e regiões vizinhas dos Estados do Mato Grosso, Rondônia, Acre e Pará. A Figura 4.1 ilustra as cenas no mapa cobrindo as regiões de estudo.

Tabela 4.1 – Cenas das imagens utilizadas.

Órbita/Ponto	Data da Imagem
001/65	22/07/2018
001/66	23/08/2018
001/67	23/08/2018
002/65	30/08/2018
002/66	14/08/2018
002/67	29/07/2018
229/65	19/07/2018
229/66	19/07/2018
229/67	19/07/2018
230/65	26/07/2018
230/66	26/07/2018
230/67	26/07/2018
231/65	17/07/2018
231/66	17/07/2018
231/67	17/07/2018
232/65	24/07/2018
232/66	24/07/2018
232/67	24/07/2018
233/65	15/07/2018
233/66	15/07/2018
233/67	15/07/2018

As imagens foram obtidas a partir do sítio do Serviço Geológico mantido pelo Governo dos Estados Unidos (USGS, 2019). As bandas utilizadas foram: Azul (Banda 2); Verde (Banda 3); Vermelho (Banda 4); Infravermelho próximo (Banda 5); Infravermelho de ondas curtas 1 (Banda 6), e; Infravermelho de ondas curtas 2 (Banda 7). Segundo (USGS, 2020) essas bandas são as mais adequadas para a detecção de vegetação. Além disso, Oliveira et al. (2017) utilizaram essas 6 bandas citadas em seu trabalho em que obteve um resultado em torno de 97% de acerto. A resolução espacial dessas imagens é de 30m (7.591 x 7.741 pixels). A Figura 4.1 ilustra o estado do Amazonas, destacando as 21 cenas das imagens que serão analisadas.

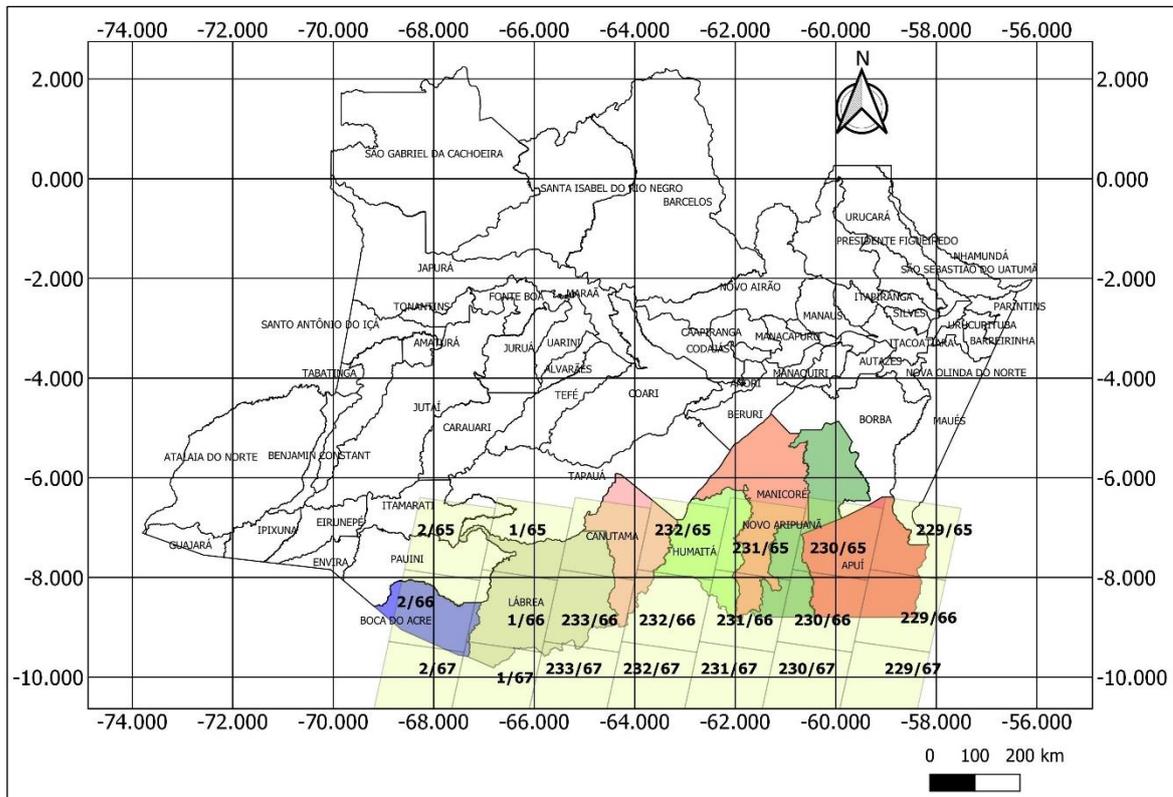


Figura 4.1 – Estado do Amazonas com as regiões de interesse destacadas.
 Fonte: *Shapefiles* – arquivos com dados geoespaciais – obtidos do IBGE (2019).

4.1.2 Dados Geoespaciais Utilizados

Os dados geoespaciais utilizados foram os dados do PRODES referente as classes Desmatamento e Floresta, do ano de 2018, correspondente a cada cena de imagem selecionada na Tabela 4.1. Esses dados vetoriais foram adquiridos do INPE (2020d).

Os dados geoespaciais são dados vetoriais compostos por polígonos e metadados. Os polígonos representam as áreas mapeadas. No caso deste trabalho, as áreas utilizadas são de desmatamento e de floresta, e os metadados são informações dessas áreas mapeadas. As Figuras 4.2 e 4.3 ilustram os polígonos das áreas desmatadas e áreas de florestas, respectivamente. Os polígonos em vermelho representam as áreas desmatadas e os em verde representam as áreas de floresta. É comum referenciar esses polígonos como máscaras. Neste trabalho, utilizamos a denominação de máscaras de desmatamentos e máscaras de florestas.

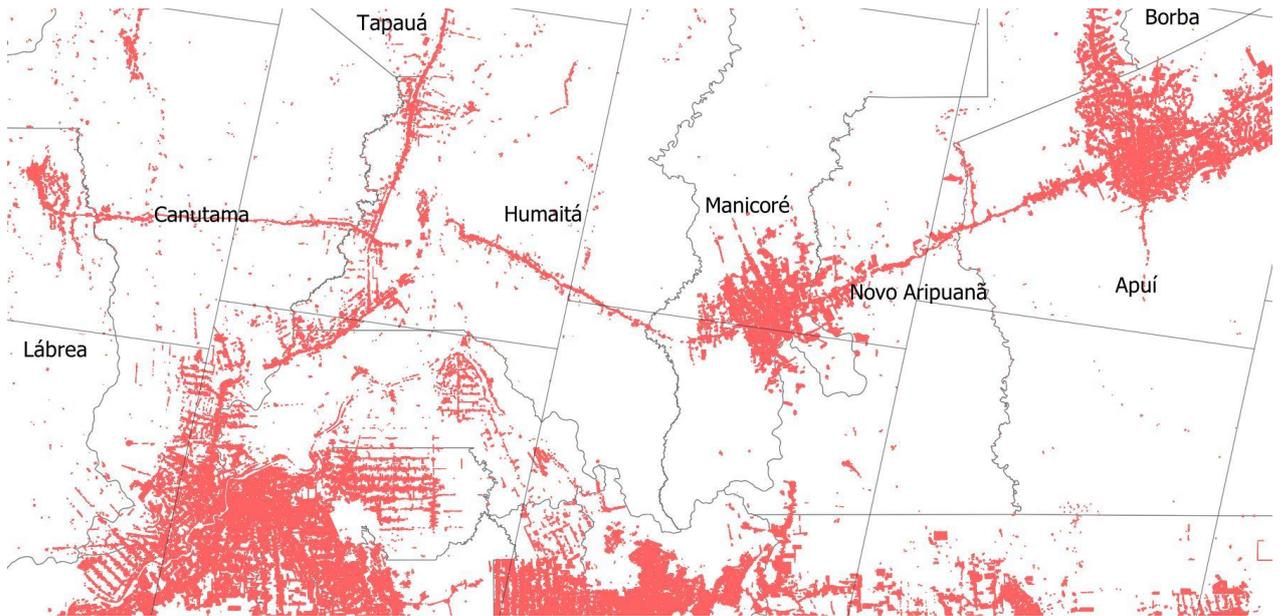


Figura 4.2 – Máscaras de áreas desmatadas (Polígonos em vermelho). Fonte: INPE (2020d).

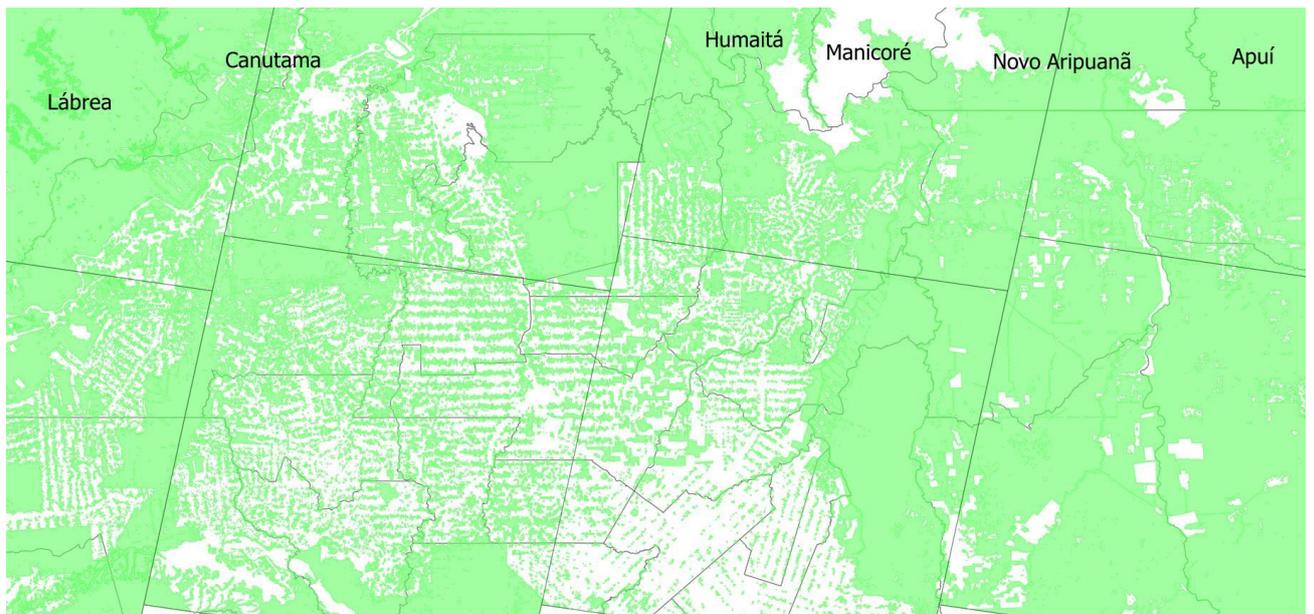


Figura 4.3 – Máscaras de áreas de floresta (Polígonos em verde). Fonte: INPE (2020d).

4.2 MÉTODOS

O desenvolvimento desta pesquisa foi efetuado através das etapas descritas no diagrama em bloco da Figura 4.4.

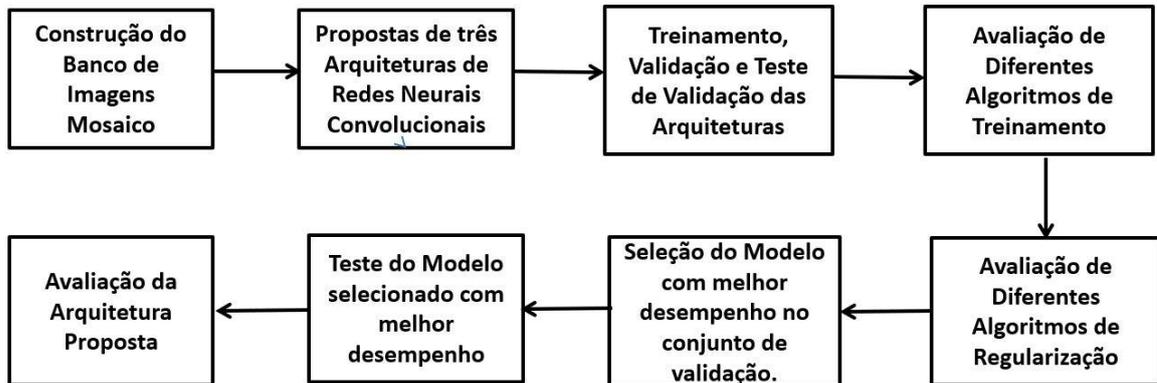


Figura 4.4 – Fluxograma do Processo Metodológico.

A seguir descreveremos cada uma das etapas mostradas nesse diagrama em blocos.

4.2.1 Construção do Banco de Imagens Mosaico

A Figura 4.5 apresenta as etapas para a elaboração do banco de imagens em mosaico.

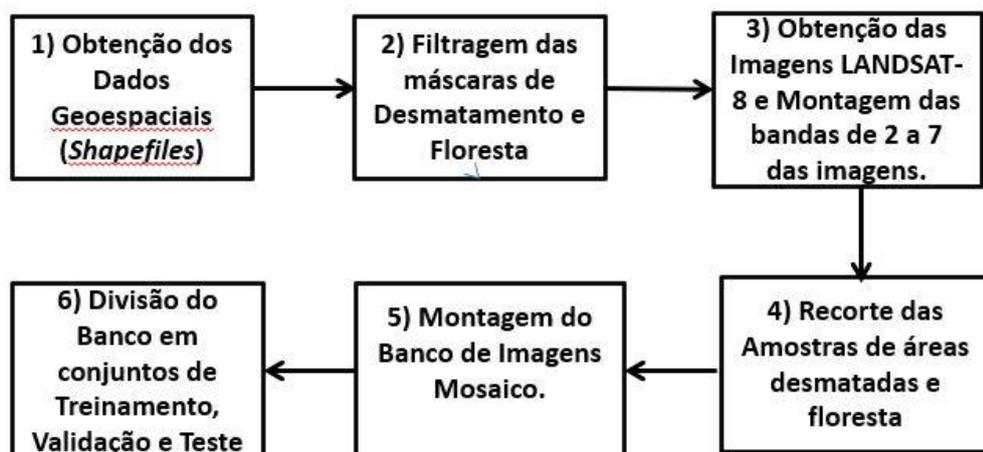


Figura 4.5 – Fluxograma da Elaboração do Banco de Imagens Mosaico.

A proposta de construir uma imagem mosaico para treinamento de uma rede CNN advém da seguinte constatação. Normalmente, em uma imagem da região amazônica, existe um desbalanceamento entre áreas de floresta e áreas desmatadas. Via de regra, a maioria das imagens contém uma área muito grande de floresta em detrimento de áreas desmatadas. Esse desbalanceamento provoca uma grande dificuldade no treinamento da CNN, pois a mesma se especializará no reconhecimento da classe que ocorre com mais frequência, ou seja, a área de floresta. Essa especialização da rede CNN faz com que a mesma cometa muitos erros na detecção de áreas desmatadas. Assim sendo, a ideia da imagem mosaico é de, a partir de pequenos fragmentos de amostras de áreas de floresta e de desmatamento, construir uma imagem em formato retangular maior, em que, em média, teremos 50% de área de floresta e 50% de área desmatada. O processo da elaboração desse banco de imagens mosaico é descrito nos passos a seguir:

- 1) Obtemos os arquivos com dados geoespaciais (*shapefiles*) para a região Sul do Amazonas;
- 2) Realizamos a filtragem das máscaras de desmatamento e floresta, utilizando a ferramenta QuantumGIS (QGIS) para cada arquivo com dados geoespaciais (*shapefiles*). Nessa etapa, verifica-se a data da imagem usada na marcação dos polígonos;
- 3) Acessando o sítio *earthexplorer* (USGS, 2019) obtemos as 6 bandas do sensor LANDSAT-8/OLI descritas na Tabela 2.1, para cada cena das imagens apresentadas na Tabela 4.1. Através da utilização da ferramenta de Ambiente para Processamento de Imagens (*Environment for Visualizing Images*), as bandas foram unidas, gerando uma única cena de imagem com as 6 bandas;
- 4) Nesta etapa são segmentadas (ou recortadas) as regiões de interesse para as classes desmatamento e floresta de cada cena em estudo, com base nos polígonos correspondentes à cada classe. A partir dessas regiões de interesse, para cada cena foram geradas duas imagens segmentadas somente com a região de interesse. A primeira imagem foi gerada para a classe desmatamento com fundo preto e a segunda, para a classe floresta também com fundo preto.
- 5) Com as imagens segmentadas descrita no item anterior, foi implementado um script em Matlab para gerar amostras de imagens de 40x40 pixels para cada classe, que foram divididas em três conjuntos: treinamento, validação e teste. A Tabela 4.2 mostra a quantidade dessas amostras separadas por conjunto e classe. As Figuras 4.6 e 4.7 ilustram os fragmentos das amostras de desmatamento e floresta, respectivamente. Essas amostras de imagens foram então utilizadas para gerar as imagens mosaico de 400x400 pixels para os conjuntos de treinamento,

validação e teste. A Tabela 4.2 mostra as quantidades de amostras por classe para cada conjunto. A escolha de uma amostra para compor um retângulo da imagem mosaico é feita de forma aleatória. Assim, não podemos garantir que 50% dos recortes de uma imagem mosaico corresponderão a amostras de desmatamento e outros 50% corresponderão às amostras de floresta. No entanto, como a seleção é aleatória, os percentuais de distribuição de área em cada imagem mosaico gira em torno de 50%.

Tabela 4.2 – Quantidade de Amostras.

	Desmatamento	Floresta
Treinamento	6.616	10.334
Validação	2.206	3.445
Testes	2.205	3.442



Figura 4.6 - Fragmentos de amostras de desmatamento 40x40.



Figura 4.7 - Fragmentos de amostras de floresta 40x40.

- 6) Segundo Aggarwall (2018), um erro comum na divisão de um banco de dados é utilizar o mesmo conjunto de dados para ajuste de parâmetro, que é o conjunto de treinamento, e avaliação final, que é o conjunto de teste. Essa abordagem resulta em uma acurácia excessivamente otimista. Para evitar esse problema, o autor sugere que o conjunto de dados seja dividido em três conjuntos: treinamento, validação e teste. No conjunto de treinamento obtém-se os parâmetros de pesos e polarizações da rede através de um algoritmo de treinamento. No conjunto de validação, os dados desse conjunto são utilizados para ajustar os parâmetros e selecionar o melhor modelo. Escolhido o melhor modelo, avalia-se sua precisão uma única vez utilizando os dados do conjunto de teste. Ratificamos a importância de garantir que os dados de teste não sejam utilizados na fase de treinamento e validação, para evitar um processo viciado de avaliação do desempenho da

máquina treinada. A Figura 4.8 ilustra a divisão da base de dados de imagens mosaicos desta pesquisa.



Figura 4.8 - Divisão do conjunto em treinamento, validação e teste.
Fonte: Adaptado de (Aggarwal, 2018).

O banco de dados deste projeto é composto por 4.000 imagens mosaicos, sendo 2.400 imagens do conjunto de treinamento, 800 do conjunto de validação e 800 do conjunto de teste para testes. Essas imagens mosaico têm dimensão de 400x400 pixels e são compostas por 100 fragmentos de amostras de floresta e desmatamento para cada imagem. Associada a cada imagem mosaico foi gerada uma imagem binarizada. A imagem binarizada é construída de tal forma à que, as pequenas áreas da imagem mosaico onde se verifica a presença de floresta correspondem a retângulos brancos na imagem binarizada, enquanto que as pequenas áreas da imagem mosaico onde se verifica a presença de desmatamento correspondem a retângulos pretos. Um total de 4.000 imagens mosaicos e suas imagens binárias correspondentes foram geradas, separando-se 60% para o conjunto de treinamento, 20% para o conjunto de validação e 20% para o conjunto de teste. A Figura 4.9 ilustra o exemplo de imagem mosaico e sua imagem binarizada correspondente.

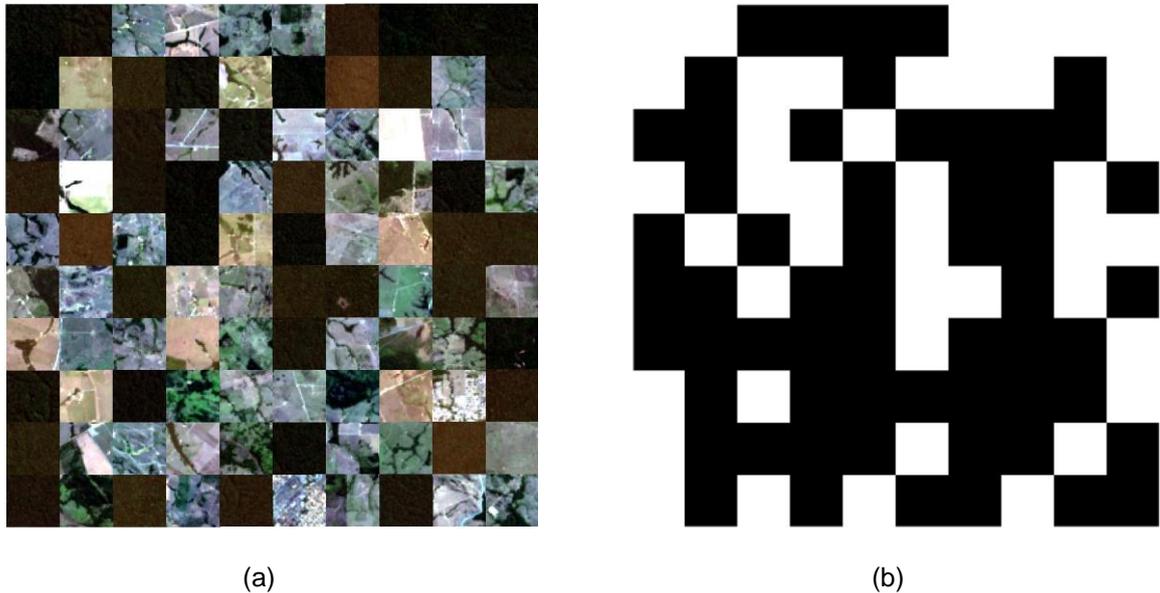


Figura 4.9 – Exemplo de imagem-mosaico: a) à esquerda, é ilustrada a imagem-mosaico com as amostras de floresta e desmatamento, e; b) à direita, a sua correspondente imagem binarizada.

4.2.2 Propostas das Arquiteturas de Redes Neurais Convolucionais

As arquiteturas propostas para este trabalho foram baseadas na primeira arquitetura proposta por Miyagawa et al. (2019). Esses autores propuseram automatizar a segmentação de lúmen e a classificação da região de bifurcação, utilizando imagens de tomografia de coerência óptica intravascular (IVOCT). Para a segmentação do lúmen, a arquitetura denominada CNN-S1 foi definida com 26 camadas ligadas diretamente, composta por quatro etapas de subamostragem, seguida da camada de *dropout* e quatro etapas de sobreamostragem. Cada etapa de subamostragem e da sobreamostragem foi dividida em várias camadas, de forma a extrair e preservar as características da imagem, tais como: camada de convolução, com kernel 3x3; camada *Bath Normalization*, e; a camada ReLU. No trabalho de Serrão et al. (2020), para realizar o reconhecimento de bacilos de tuberculose, os autores propuseram três arquiteturas CNN1, CNN2 e CNN3, baseadas na arquitetura CNN-S1 do trabalho de Miyagawa et al. (2019). De forma geral, as estruturas dessas arquiteturas são compostas por sequências de blocos da camada convolutiva com filtro 3x3, com camada *Bath Normalization* e ReLU, repetidas de 2 a 3 vezes, seguida da camada *Max Pooling* com filtro 2x2. Essa última etapa é definida como uma etapa de subamostragem, reduzindo de forma gradual a dimensionalidade da imagem. Na etapa seguinte, a arquitetura é composta

por mais uma sequência de blocos da camada de convolução transposta com filtro 4x4, com camada *Bath Normalization* e ReLU, repetidas de 2 a 4 vezes, seguida da camada convolutiva de kernel 1x1. Essa etapa é definida como a etapa de sobreamostragem, aumentando de forma gradual a dimensionalidade da imagem. Em seguida, as últimas camadas *softmax* e *pixel classification* são utilizadas para a classificação de cada pixel da imagem.

O presente trabalho foi desenvolvido para realizar a segmentação semântica da imagem descrita no item 4.2.1, obtida a partir das 6 bandas das imagens Landsat 8 para áreas de florestas desmatadas ou não. Para tanto, as arquiteturas utilizadas neste estudo foram as arquiteturas CNN1, CNN2 e CNN3 propostas por Serrão et al. (2020) baseadas na arquitetura CNN-S1 de Miyagawa et al. (2019). A Figura 4.6 ilustra a estrutura das arquiteturas CNN1, CNN2 e CNN3.

A arquitetura CNN1 possui quatro estágios de subamostragem e quatro estágios de sobreamostragem. A arquitetura CNN2 possui apenas dois estágios de subamostragem e dois estágios de sobreamostragem e a arquitetura CNN3 possui também dois estágios de subamostragem e dois de sobreamostragem, diferenciando-se no estágio de subamostragem que possui apenas uma camada convolucional seguida da camada *bath normalization* e ReLU. Entre os estágios de subamostragem e sobreamostragem das três arquiteturas é introduzida a camada de *dropout* para fins da aplicação da técnica de regularização. Na Figura 4.6, os valores numéricos 16, 32, 64, 128 e 2 representam a quantidade de camadas de neurônios em cada camada convolutiva, equivalendo também ao número de *kernel*s utilizados na camada convolutiva.

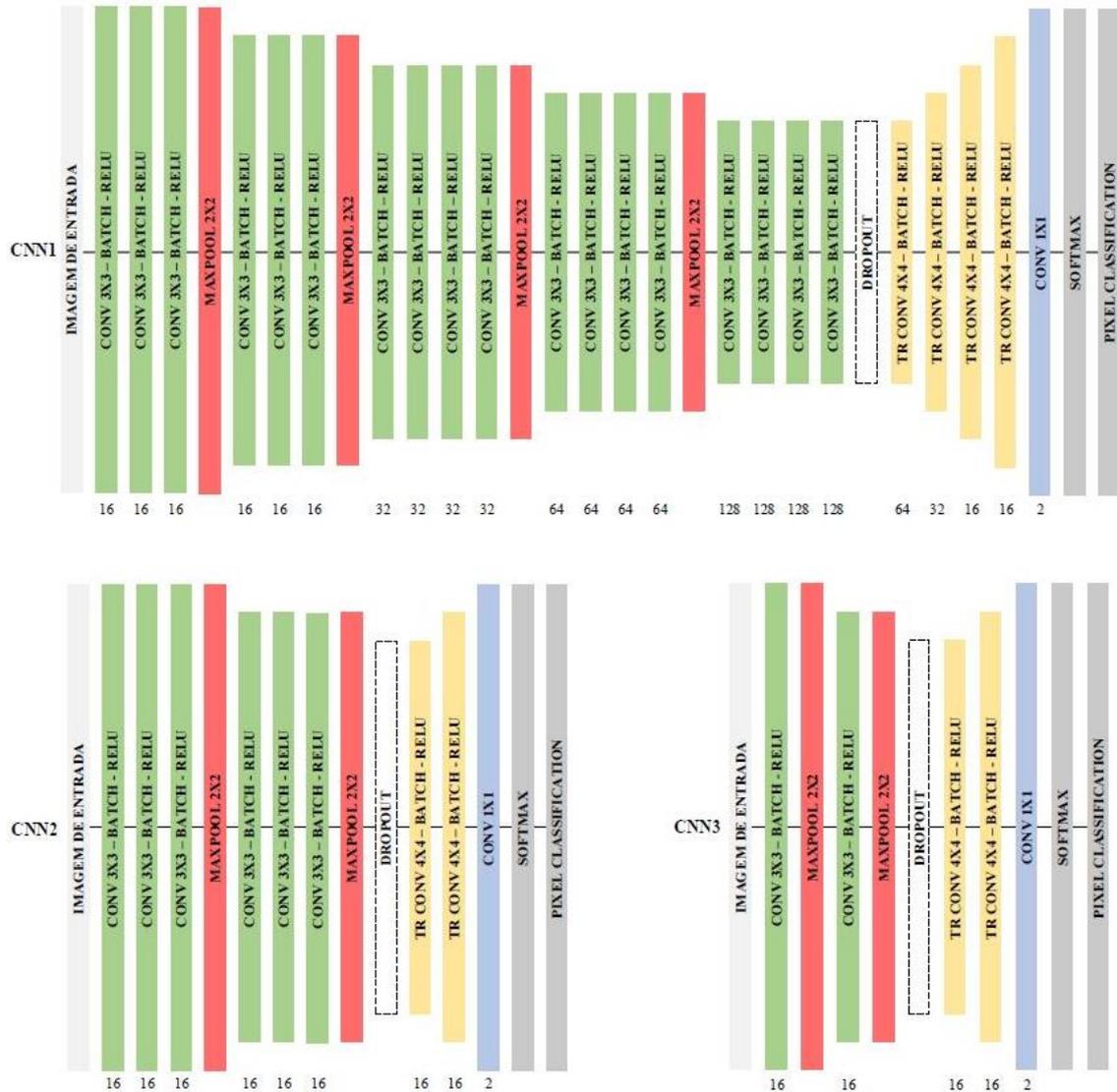


Figura 4.2 – Arquiteturas CNN1, CNN2 e CNN3.
Fonte: Serrão et al. (2020).

4.2.3 Treinamento, Validação e Teste das Arquiteturas

O treinamento das arquiteturas CNN é supervisionado, o que significa, no treinamento, apresentar na entrada da rede um conjunto de imagem-mosaico e na saída, a imagem-binária correspondente.

A Tabela 4.2 mostra os hiperparâmetros utilizados no treinamento e seus valores. O critério de parada para o treinamento foi estabelecido em 32 épocas. Os parâmetros numéricos mostrados na Tabela 4.3 foram obtidos experimentalmente, através das simulações efetuadas com as redes convolutivas, procurando um compromisso entre desempenho e agilização do treinamento.

Tabela 4.3 – Valores para os hiperparâmetros de treinamento.

Hiperparâmetro	Valor
Taxa de aprendizagem	0,001
Tamanho do mini-lote	8
Decaimento da taxa de aprendizagem	0,5
Número de épocas	32
Fator de Regularização L2	0,001
Técnica de Regularização	<i>Dropout</i> , regularização L2 ou <i>Dropout</i> + regularização L2
Algoritmo de Treinamento	SGDM, RMSProp ou ADAM

Com a rede treinada, foi realizada a validação utilizando-se o conjunto de validação. O objetivo da validação é selecionar, dentre as três arquiteturas, aquela que apresenta os melhores resultados. Em seguida, o desempenho da melhor arquitetura é avaliado através do conjunto de teste.

As três arquiteturas propostas serão treinadas utilizando os seguintes algoritmos de treinamento: SGDM, RMSProp e ADAM.

Para cada algoritmo de treinamento foram avaliados os seguintes métodos de regularização: nenhum método de regularização, regularização L_2 , camada de *dropout* e regularização L_2 + camada de *dropout*.

No total, foram realizadas 36 simulações (3 arquiteturas \times 3 métodos de otimização \times 4 métodos de regularização). Os conceitos sobre os métodos de regularização Dropout e L_2 foram descritos no capítulo da fundamentação teórica.

4.2.4 Seleção do Melhor Modelo de Arquitetura

A seleção do modelo de arquitetura com melhor desempenho será feita no conjunto de validação utilizando as seguintes métricas de avaliação como: Sensibilidade, Especificidade, Acurácia, Precisão e F1-Score. Essas métricas são descritas na Subseção 4.2.5 a seguir.

4.2.5 Avaliação das Arquiteturas Propostas

Para avaliação das arquiteturas propostas foram utilizadas as seguintes métricas: sensibilidade, especificidade, acurácia, precisão e F1-Score. Antes de iniciarmos a definição dessas métricas, apresentamos na Tabela 4.4 alguns conceitos que serão utilizados posteriormente. A Tabela 4.4 é mais conhecida como matriz de confusão.

Tabela 4.4 – Matriz de Confusão.

		Previsto	
		Desmatamento	Floresta
Real	Desmatameto	VP	FN
	Floresta	FP	VN

Em que:

- VP (Verdadeiro Positivo) – Desmatamento classificado como Desmatamento;
- FP (Falso Positivo) – Floresta classificada como Desmatamento;
- VN (Verdadeiro Negativo) – Floresta classificada como Floresta;
- FN (Falso Negativo) – Desmatamento classificado como Floresta;

A partir das variáveis mostradas na matriz de confusão podemos obter as métricas sensibilidade, especificidade e acurácia. A Sensibilidade (S) identifica o quanto o classificador é eficaz em reconhecer corretamente as áreas desmatadas e é dada pela equação (4.1).

$$S = \frac{VP}{VP + FN} \quad (4.1)$$

A sensibilidade é a proporção de VP, ou seja, avalia a capacidade do modelo de classificar as áreas desmatadas como desmatamento. A Especificidade (E) identifica o quanto o classificador é eficaz em reconhecer corretamente áreas que não apresentam desmatamento. É dada pela equação (4.2).

$$E = \frac{VN}{VN + FP} \quad (4.2)$$

A Acurácia é o acerto total do classificador e é dada pela equação (4.3).

$$Acc = \frac{(VP + VN)}{(VP + FP + VN + FN)} \quad (4.3)$$

A precisão representa a razão entre a área de desmatamento classificada corretamente (VP), e a área total classificada como desmatamento, incluindo áreas

corretamente classificadas como desmatamento e incorretamente classificadas como desmatamento (VP + FP). É dada pela equação (4.4):

$$Prec = \frac{VP}{VP + FP} \quad (4.4)$$

Alguns classificadores podem apresentar melhor desempenho para a sensibilidade, outros podem apresentar melhor desempenho para a precisão. Neste caso, para comparar o desempenho de diferentes classificadores, utiliza-se, em geral, a métrica F1 – Score. Essa métrica avalia o desempenho global do classificador e é dada pela equação (4.5) que representa uma média ponderada entre precisão e sensibilidade.

$$F1 - Score = 2 * \frac{Prec * S}{Prec + S} \quad (4.5)$$

Das métricas citadas acima, a acurácia foi escolhida como critério para selecionar a melhor arquitetura. Nas simulações do conjunto de validação foram selecionados os dois melhores modelos de cada arquitetura, um com aplicação de métodos de regularização e outro sem.

4.2.6 Ambiente de Desenvolvimento

Para a realização dos experimentos do projeto, treinamento e teste e adaptação/implementação das arquiteturas, foi utilizado o *software* Matlab® versão 2019b. Foram utilizadas as ferramentas QGIS e Envi para recorte das áreas de interesse e visualização das imagens, respectivamente.

As configurações do computador em que foram realizadas as simulações são: processador Intel i7-8700 com 3,2 GHz; memória RAM 32 GB; e NVIDIA Quadro GV100 32GB com 5.120 núcleos CUDA.

4.3 CONSIDERAÇÕES

Neste capítulo foi mostrado como foram adquiridas as imagens e os dados geoespaciais utilizados nesta pesquisa. Foi apresentado também o fluxograma com as etapas realizadas para obtenção do banco de dados constituído por imagens mosaico, uma

contribuição desse trabalho para a área de reconhecimento de padrões em sensoriamento remoto.

As arquiteturas de redes neurais propostas foram apresentadas e foram definidos os parâmetros para treinamento, validação e teste. Os métodos de regularização e algoritmos de treinamento também foram definidos. Por último, foram descritas as medidas para avaliação de desempenho dos resultados a serem obtidos para cada arquitetura.

No próximo capítulo mostramos os principais resultados alcançados para as arquiteturas propostas e analisamos o desempenho das mesmas com as métricas propostas nesse capítulo.

CAPÍTULO 5

5. RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados, analisados e discutidos os resultados obtidos com as simulações realizadas com as três arquiteturas CNN1, CNN2, CNN3 e com os três algoritmos de otimização SGDM, RMSProp e Adam, com a aplicação dos métodos de regularização.

5.1 Resultados das Simulações das Arquiteturas no Conjunto de Validação

Nesta seção, apresentamos a análise do desempenho das três arquiteturas CNN1, CNN2 e CNN3 para os algoritmos de otimização SGDM, RMSProp e Adam, sem a aplicação dos métodos para a melhoria da generalização e com a aplicação dos métodos regularização L_2 , *dropout* e $L_2+dropout$ no conjunto de validação. Esta seção compreende as etapas da metodologia de avaliação dos diferentes algoritmos de otimização e avaliação dos diferentes métodos de regularização.

Os resultados das simulações foram avaliados através das seguintes métricas de desempenho: Sensibilidade; Especificidade; Acurácia; Precisão e F1-Score. A seguir, as Tabelas 5.1 e 5.2 apresentam o desempenho da arquitetura CNN1 sem e com a aplicação dos métodos para melhoria da generalização, respectivamente, no conjunto de validação.

Tabela 5.1 – Resultados do desempenho da CNN1 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.

Algoritmo de Treinamento	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	0,99956	0,99958	0,99957	0,99958	0,99957
RMSProp	1,00000	0,99993	0,99996	0,99993	0,99996
ADAM	0,99999	0,99977	0,99988	0,99977	0,99988

Tabela 5.2 – Resultados do desempenho da CNN1 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de métodos para melhoria da generalização.

Algoritmo de Treinamento	Método de Regularização	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	L2	0,99928	0,99978	0,99953	0,99978	0,99953
	Dropout	0,99897	0,99991	0,99944	0,99991	0,99944
	L2+Dropout	0,99905	0,99992	0,99949	0,99992	0,99949
RMSProp	L2	0,99996	0,99996	0,99996	0,99996	0,99996
	Dropout	1,00000	0,99987	0,99993	0,99987	0,99993
	L2+Dropout	0,99992	0,99994	0,99993	0,99994	0,99993
ADAM	L2	0,99929	1,00000	0,99964	1,00000	0,99964
	Dropout	1,00000	0,99981	0,99990	0,99981	0,99990
	L2+Dropout	0,99939	0,99948	0,99944	0,99948	0,99944

As Tabelas 5.3 e 5.4 apresentam os valores das medidas de desempenho obtidos pela arquitetura CNN2 sem e com a aplicação dos métodos para melhoria da generalização, respectivamente, no conjunto de validação.

Tabela 5.3 – Resultados do desempenho da CNN2 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.

Algoritmo de Treinamento	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	0,99793	0,99874	0,99834	0,99874	0,99833
RMSProp	0,99969	0,99988	0,99979	0,99988	0,99978
ADAM	0,99895	0,99997	0,99946	0,99997	0,99946

Tabela 5.4 – Resultados do desempenho da CNN2 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de métodos para melhoria da generalização.

Algoritmo de Treinamento	Método de Regularização	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	L2	0,99786	0,99880	0,99833	0,99880	0,99833
	Dropout	0,99828	0,99871	0,99850	0,99870	0,99849
	L2+Dropout	0,99775	0,99907	0,99841	0,99907	0,99841
RMSProp	L2	0,99952	0,99994	0,99973	0,99994	0,99973
	Dropout	0,99971	0,99987	0,99979	0,99987	0,99979
	L2+Dropout	0,99928	0,99987	0,99957	0,99986	0,99957
ADAM	L2	0,99973	0,99975	0,99974	0,99975	0,99974
	Dropout	0,99886	0,99987	0,99937	0,99873	0,99936
	L2+Dropout	0,99896	0,99992	0,99944	0,99992	0,99944

As Tabelas 5.5 e 5.6 apresentam o desempenho da arquitetura CNN3 sem e com a aplicação dos métodos para melhoria da generalização, respectivamente, no conjunto de validação.

Tabela 5.5 – Resultados do desempenho da CNN3 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação sem aplicação de métodos para melhoria da generalização.

Algoritmo de Treinamento	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	0,99520	0,99759	0,99639	0,99758	0,99639
RMSProp	0,99866	0,99977	0,99922	0,99977	0,99921
ADAM	0,99880	0,99964	0,99922	0,99964	0,99922

Tabela 5.6 – Resultados do desempenho da CNN3 com métodos de otimização SGDM, RMSProp e ADAM no conjunto de validação com aplicação de método de regularização.

Algoritmo de Treinamento	Método de Regularização	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
SGDM	<i>L2</i>	0,99608	0,99769	0,99688	0,99768	0,99688
	<i>Dropout</i>	0,99367	0,99813	0,99590	0,99812	0,99589
	<i>L2+Dropout</i>	0,99411	0,99860	0,99635	0,99859	0,99634
RMSProp	<i>L2</i>	0,99888	0,99957	0,99923	0,99957	0,99922
	<i>Dropout</i>	0,99736	0,99991	0,99863	0,99991	0,99863
	<i>L2+Dropout</i>	0,99891	0,99915	0,99903	0,99915	0,99903
ADAM	<i>L2</i>	0,99896	0,99943	0,99919	0,99943	0,99919
	<i>Dropout</i>	0,99902	0,99968	0,99935	0,99968	0,99935
	<i>L2+Dropout</i>	0,99806	0,99963	0,99884	0,99963	0,99884

Após a apresentação dos resultados dos desempenhos das arquiteturas, os valores marcados em negrito, nas tabelas anteriormente apresentadas, representam os melhores valores de acurácia e F1-Score em cada uma das tabelas de resultados.

5.2 Análise dos Resultados das Arquiteturas no Conjunto de Validação

Para a arquitetura CNN1, a simulação com melhor desempenho, sem e com regularização, ocorreu com aplicação do algoritmo de otimização RMSProp. O método para melhoria da generalização com melhor desempenho foi o método de regularização L_2 . Essas duas simulações obtiveram o mesmo valor de **99,996%** para as duas métricas acurácia e F1-Score, conforme mostrado nas Tabelas 5.1 e 5.2. Assim, podemos inferir que a aplicação de métodos de regularização não influenciou significativamente nos resultados para a CNN1. Para essa arquitetura, esses dois modelos foram os que obtiveram os melhores resultados no conjunto de validação.

Na arquitetura CNN2, a simulação com melhor desempenho, sem e com regularização, ocorreu com aplicação do algoritmo de otimização RMSProp. O método para melhoria da generalização com melhor desempenho foi o método *dropout*. Para esses dois métodos, os valores de todas as métricas deram muito próximo, obtendo o mesmo valor de acurácia de **99,979%**, conforme mostrado nas Tabelas 5.3 e 5.4. Sendo assim, podemos observar que a aplicação de métodos de regularização não teve uma grande influência nos resultados significativamente na CNN2. Para essa arquitetura, esses dois modelos foram os melhores resultados obtidos no conjunto de validação.

Na arquitetura CNN3, a simulação com melhor desempenho, sem e com regularização, ocorreu com aplicação do algoritmo de otimização ADAM. O método para melhoria da generalização com melhor desempenho foi o método *dropout*. Esses dois métodos obtiverem os valores de acurácia de **99,922%** e **99,935%**, respectivamente, conforme mostrado nas Tabelas 5.5 e 5.6. Com a aplicação do método *dropout* houve um pequeno aumento no valor de acurácia. Para essa arquitetura, esses dois modelos produziram os melhores resultados no conjunto de validação.

De acordo com os melhores resultados obtidos das três redes, podemos concluir que a aplicação dos métodos de regularização *dropout* e L_2 não influenciam de forma significativa a capacidade de generalização das redes.

A Figura 5.1 apresenta um exemplo da imagem mosaico original do conjunto de validação. As Figuras 5.2, 5.3, 5.4, 5.5, 5.6 e 5.7 mostram exemplos dos melhores resultados da segmentação das redes CNN1, CNN2 e CNN3 utilizando essa imagem mosaico.

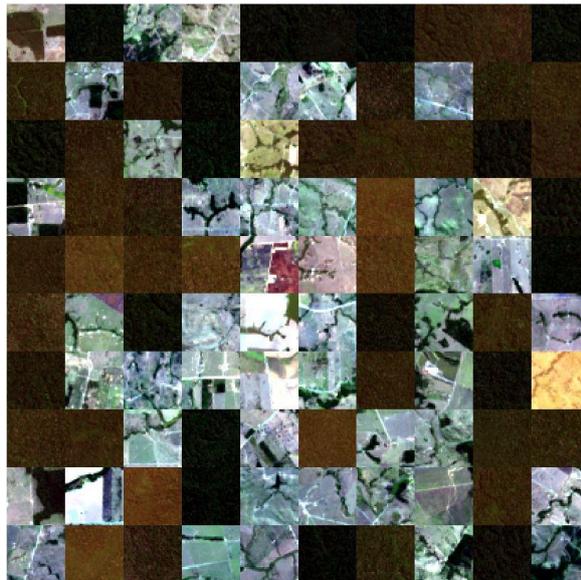


Figura 5.1 –Imagem-mosaico original do conjunto validação.

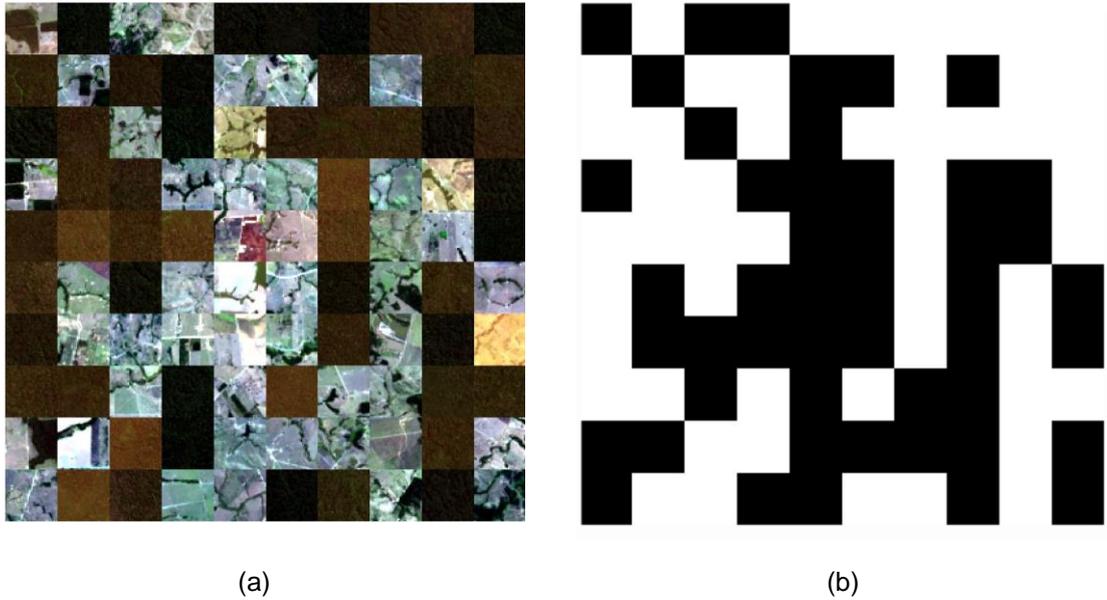


Figura 5.2 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN1+ RMSProp + sem regularização: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

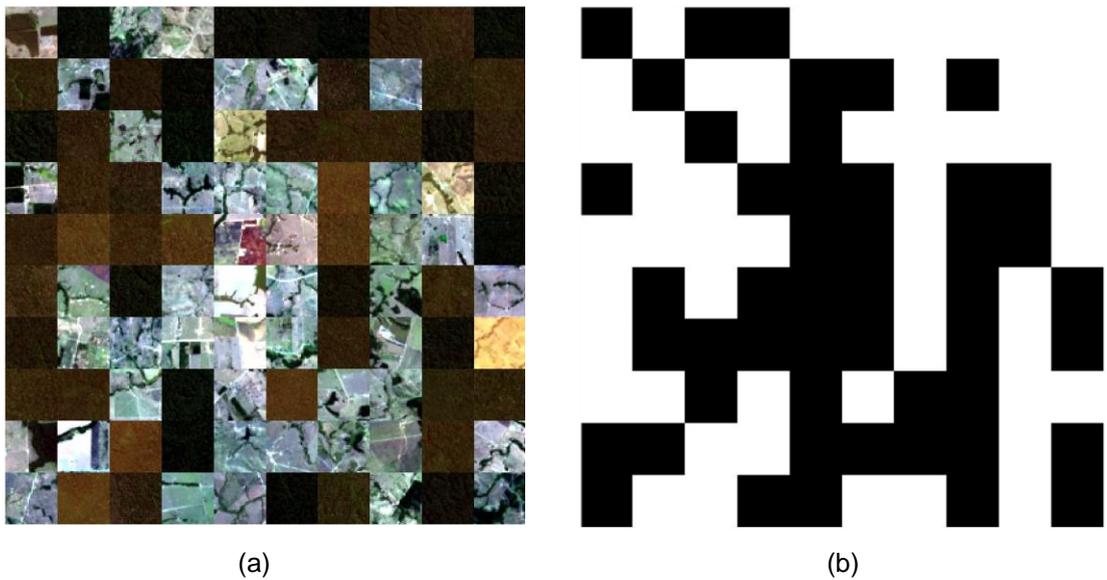


Figura 5.3 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN1+ RMSProp + L_2 : a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

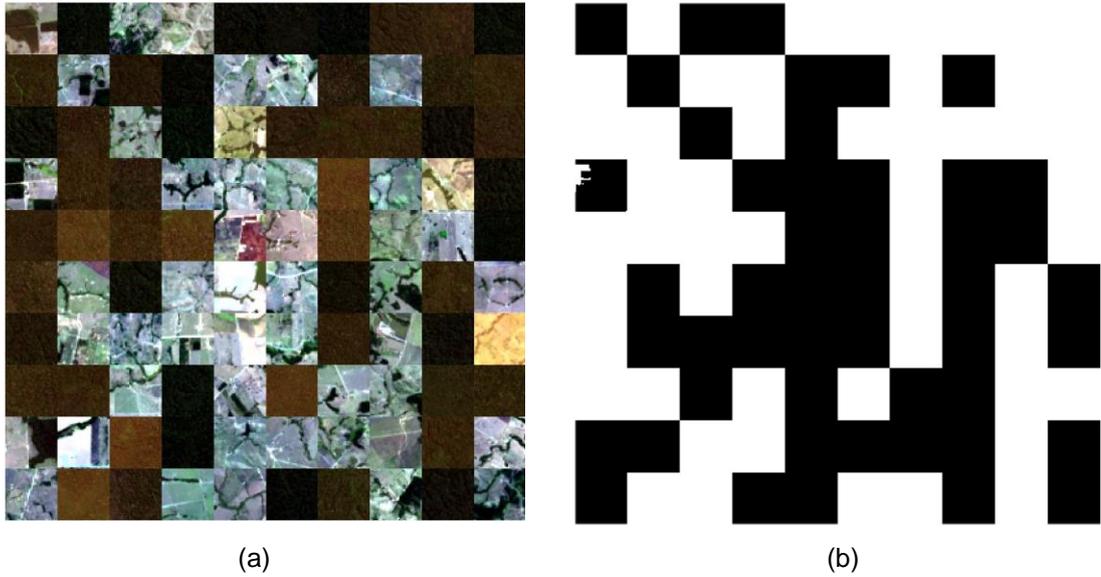


Figura 5.4 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN2 + RMSProp + sem regularização: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

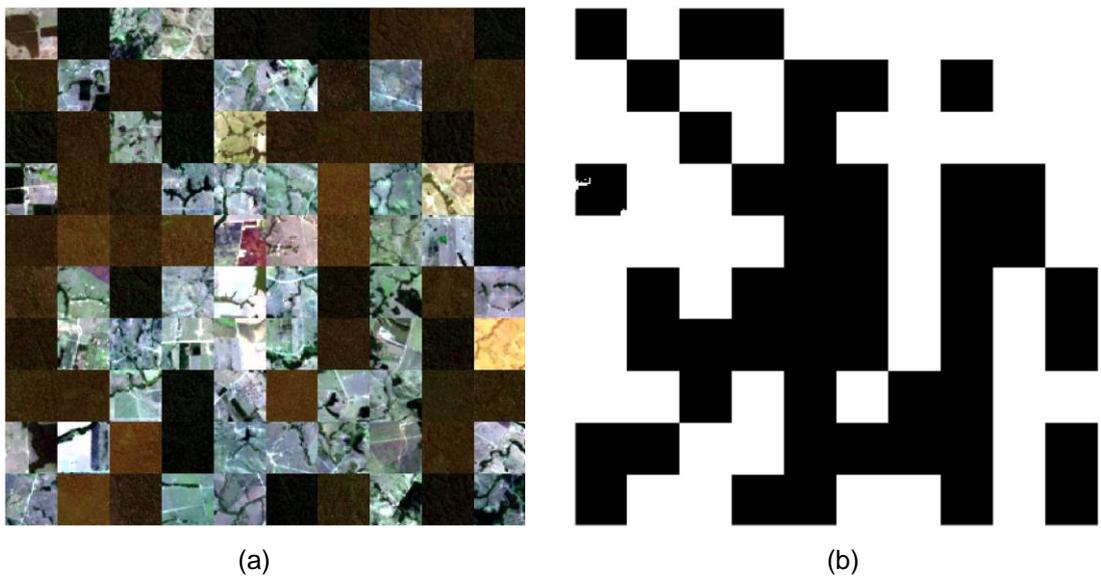


Figura 5.5 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN2 + RMSProp + *dropout*: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

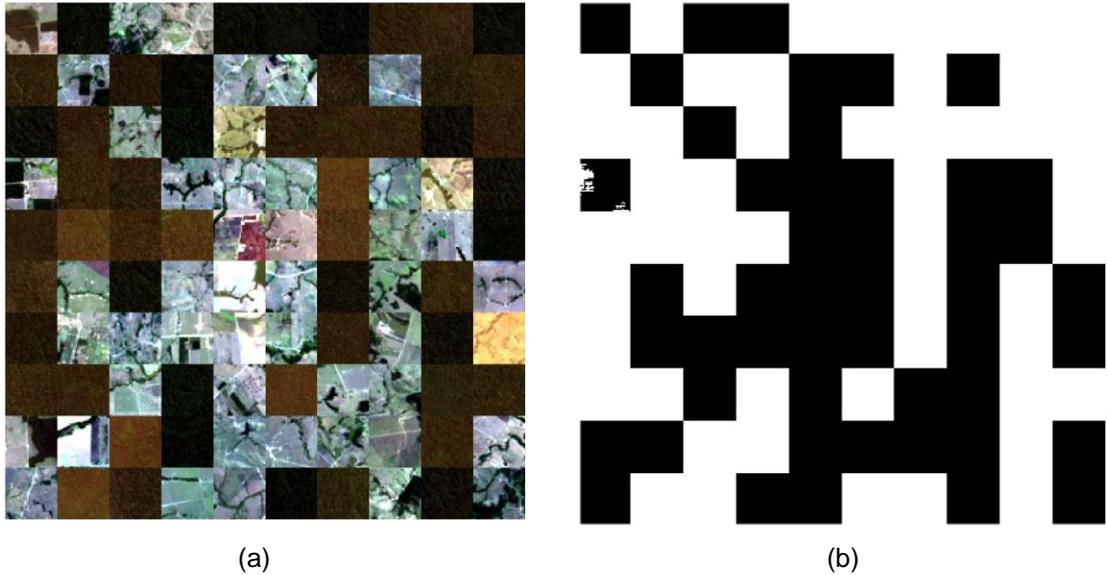


Figura 5.6 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN3 + ADAM + sem regularização: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

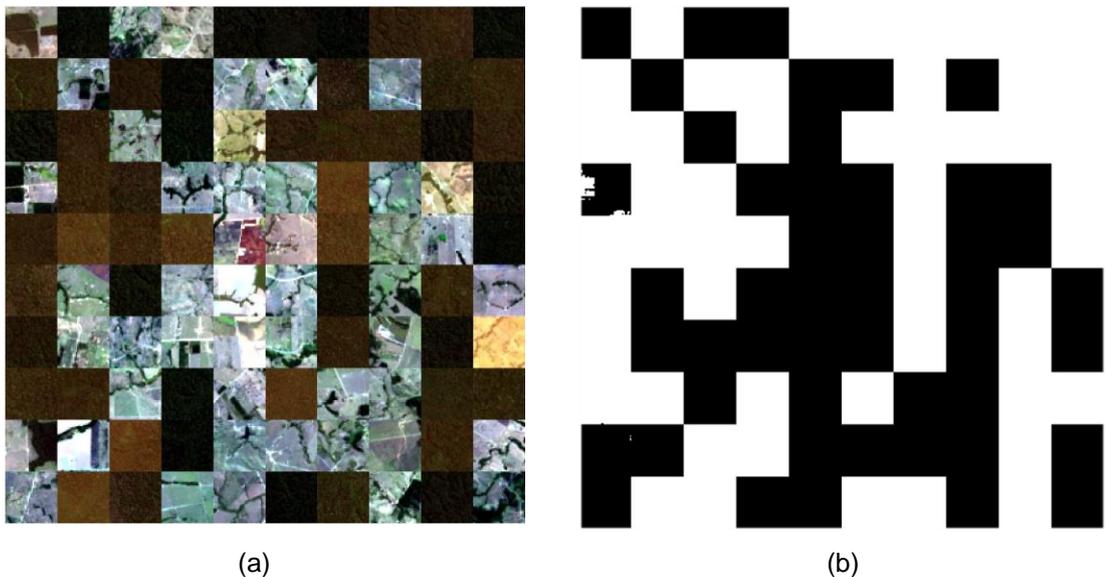


Figura 5.7 – Resultado da segmentação de uma imagem-mosaico do conjunto validação da rede CNN3 + ADAM + *dropout*: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

Analisando os resultados da segmentação anteriormente apresentados do conjunto de validação, de forma visível percebemos, nas imagens binarizadas de saída, que ocorrem poucos ruídos. Podemos notar, de forma visível, um pequeno ruído encontrado em um único fragmento na imagem de saída binarizada para as redes CNN2, Figuras 5.4(b) e 5.5(b), e CNN3, Figuras 5.6(b) e 5.7(b). Apesar disso, todos os outros fragmentos da imagem foram perfeitamente segmentados, mostrando que a CNN2 e a CNN3 não deixaram de segmentar de forma aceitável, lembrando que essas redes tiveram ótimos valores de acurácia. Já a rede

CNN1 segmentou perfeitamente as imagens, o que pode ser observado nas imagens binarizadas de saída produzidas pela mesma, Figuras 5.2(b) e 5.3(b).

5.3 Resultados da Seleção do Modelo com Melhor Desempenho

De acordo com as discussões dos resultados apresentados anteriormente para o conjunto de validação, notamos que as três redes tiveram ótimos resultados, e que alguns modelos obtiveram o mesmo valor de acurácia. Desta forma, foram selecionados dois modelos com os melhores resultados de cada arquitetura para realizar a aplicação no conjunto de teste. A Tabela 5.7 apresenta os modelos selecionados para aplicação no conjunto de teste.

Tabela 5.7 – Modelos selecionados do conjunto de validação para aplicação do conjunto de teste.

Rede	Métodos	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
CNN1	<i>RMSProp + sem regularização</i>	1,00000	0,99993	0,99996	0,99993	0,99996
	<i>RMSProp + L2</i>	0,99996	0,99996	0,99996	0,99996	0,99996
CNN2	<i>RMSProp + sem regularização</i>	0,99969	0,99988	0,99979	0,99988	0,99978
	<i>RMSProp + Dropout</i>	0,99971	0,99987	0,99979	0,99987	0,99979
CNN3	<i>ADAM + sem regularização</i>	0,99880	0,99964	0,99922	0,99964	0,99922
	<i>ADAM + Dropout</i>	0,99902	0,99968	0,99935	0,99968	0,99935

5.4 Resultado dos Testes dos Modelos Selecionados

A Tabela 5.8 apresenta os resultados de desempenho das redes CNN1, CNN2 e CNN3 aplicadas no conjunto de teste.

Tabela 5.8 – Resultados do desempenho das redes CNN1, CNN2 e CNN3 do conjunto de teste.

Rede	Métodos	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
CNN1	<i>RMSProp + sem regularização</i>	0,99892	0,99997	0,99945	0,99997	0,99944
	<i>RMSProp + L2</i>	0,99936	0,99997	0,99967	0,99997	0,99966

CNN2	<i>RMSProp + sem regularização</i>	0,99914	0,99984	0,99949	0,99983	0,99949
	<i>RMSProp + Dropout</i>	0,99955	0,99991	0,99973	0,99991	0,99973
CNN3	<i>ADAM + sem regularização</i>	0,99866	0,99967	0,99917	0,99966	0,99916
	<i>ADAM + Dropout</i>	0,99872	0,99971	0,99921	0,99971	0,99921

A partir dos resultados da Tabela 5.8, observamos que a combinação da rede CNN2 + RMSProp + Dropout apresentou o maior valor de acurácia de 99,973%, sendo esse valor de acurácia próximo àquele do conjunto de validação, de 99,979%, mostrado da Tabela 5.4. No conjunto de validação os melhores modelos foram CNN1 + RMSProp + L_2 e CNN1 + RMSProp + sem regularização, ambos com valor de acurácia de 99,996%, Tabelas 5.1 e 5.2. No conjunto de teste, seus valores foram de 99,967% e 99,945%, respectivamente. Concluímos que nem sempre o modelo que apresenta o melhor resultado no conjunto de validação obtém o melhor resultado no conjunto de teste.

Os modelos da rede CNN2 + RMSProp + Dropout e da rede CNN1 + RMSProp + L_2 foram os que obtiveram os maiores valores de acurácia de 99,973% e 99,967%, respectivamente, no conjunto de teste, havendo uma diferença não significativa de 0,006% entre esses dois modelos. Concluímos que a remoção dos blocos das camadas convolutivas, *bath normalization* e RELU, e a remoção das camadas de *max pooling* da CNN1, para gerar a rede CNN2, apresentou uma pequena influência nos resultados de segmentação das imagens obtidos pela rede CNN2.

De uma forma geral, a rede aprende melhor as características das áreas desmatadas e áreas de florestas à medida que são utilizadas mais camadas convolutivas e etapas de subamostragem na geração de mapas de características. A propósito, as redes CNN2 e CNN3 possuem um menor número de camadas convolutivas e etapas de subamostragem. Apesar disso, essas duas redes também não deixaram de realizar a classificação com um desempenho considerado muito bom.

A Figura 5.8 apresenta um exemplo da imagem mosaico original do conjunto de teste. As Figuras 5.9 e 5.10 mostram o exemplo do resultado de segmentação do modelo CNN2 + RMSProp + dropout e do modelo CNN1 + RMSProp + L_2 , respectivamente.

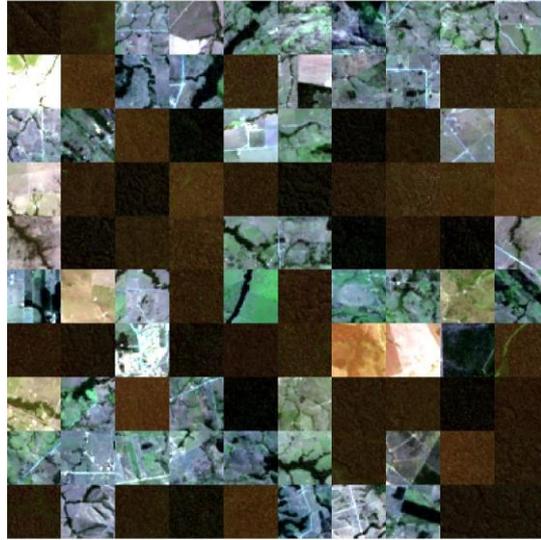


Figura 5.8 –Imagem-mosaico original do conjunto teste.

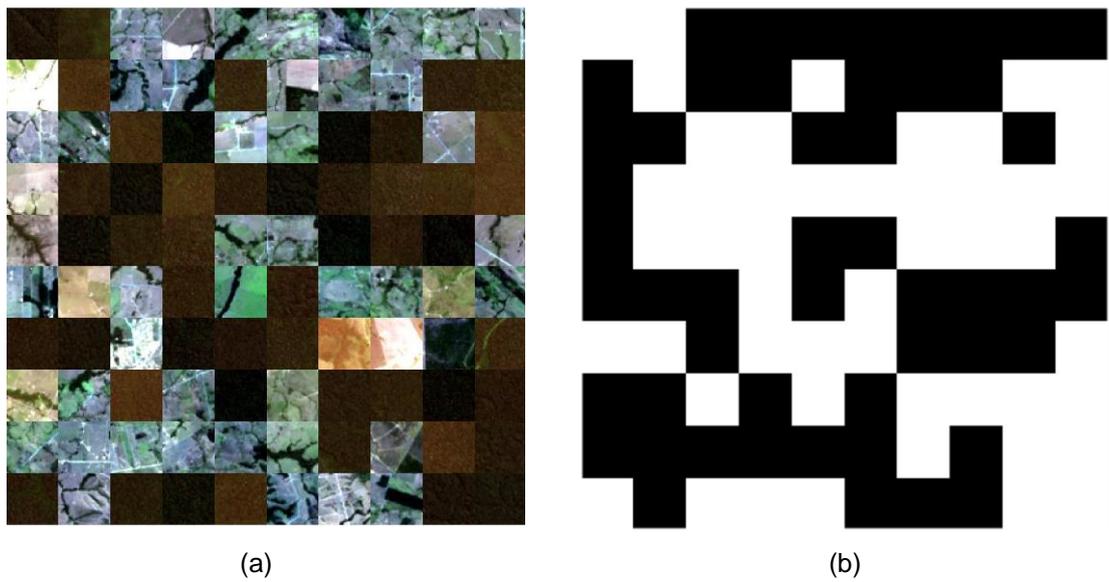


Figura 5.9 – Resultado da segmentação de uma imagem-mosaico do conjunto teste do modelo CNN2 + RMSProp + dropout: a) Imagem-mosaico (entrada), e; b) Imagem binarizada (saída).

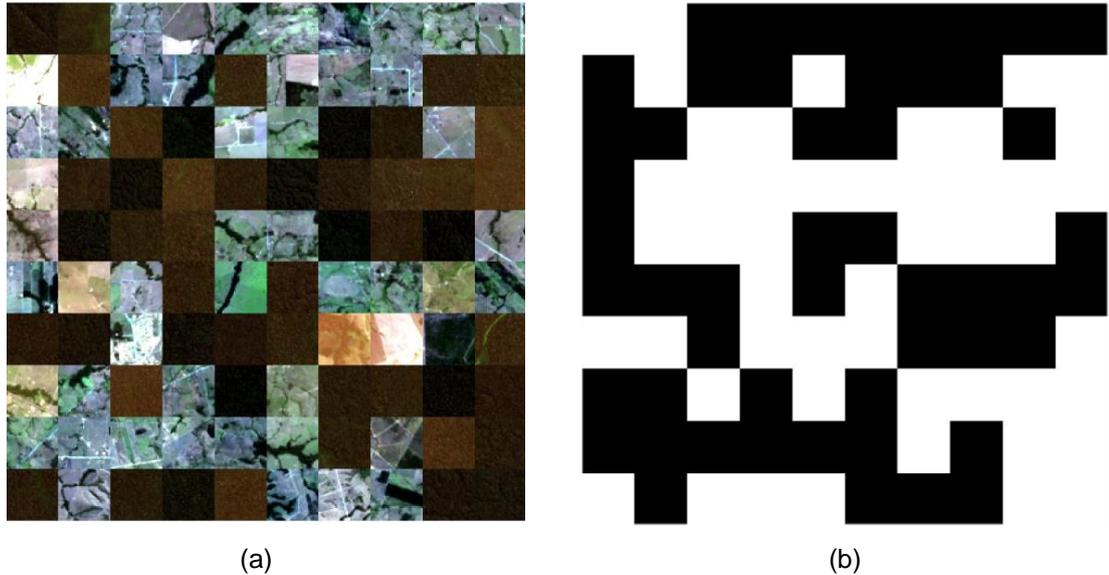


Figura 5.10 – Resultado da segmentação de uma imagem-mosaico do conjunto teste do modelo CNN1 + RMSProp + L_2 : a) Imagem-mosaico (entrada), e; b) Imagem binária (saída).

De acordo com os resultados obtidos, os dois modelos CNN2 + RMSProp + dropout e CNN1 + RMSProp + L_2 segmentaram perfeitamente as imagens. Observamos que as imagens binarizadas de saída não apresentaram ruídos. Analisando o valor da acurácia notamos que a diferença de valores foi muito pequena entre os modelos. Embora a CNN2 possua um número menor de blocos de aprendizagem, a mesma apresentou um excelente desempenho. Considerando os maiores valores das métricas de sensibilidade, especificidade, acurácia, precisão e F1-Score, podemos concluir que o melhor modelo escolhido para a segmentação de áreas desmatadas foi o modelo CNN2 com algoritmo de otimização RMSProp e método de regularização *dropout*. A Tabela 5.9 apresenta o desempenho desse modelo.

Tabela 5.9 – Desempenho da CNN2 com algoritmo de otimização RMSProp e com método de regularização *dropout* no conjunto de teste.

Experimento	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
CNN2 + RMSProp + Dropout	99,955%	99,991%	99,973%	99,991%	99,973%

A Figura 5.11 apresenta um comparativo da imagem mosaico original com a imagem binarizada de saída do melhor modelo. Nessa figura, observamos que o resultado da segmentação é livre de ruídos.

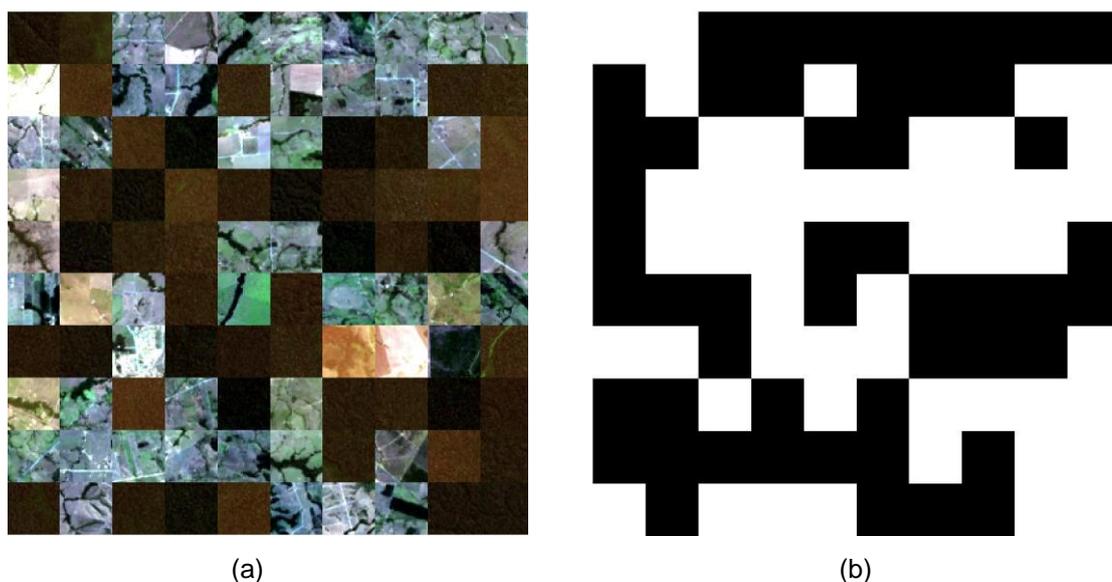


Figura 5.11 – Comparativo da Imagem mosaico original com a imagem segmentada do melhor modelo testado. a) Imagem mosaico original, e; b) Imagem segmentada da rede CNN2 + RMSProp + dropout.

5.5 Avaliação do Resultado com uso da Técnica Mosaico

A aplicação da técnica mosaico propiciou um balanceamento entre as duas classes definidas de desmatamento e de floresta, fazendo com que a rede obtivesse um equilíbrio no treinamento das duas classes, evitando a especialização da aprendizagem em apenas uma das classes. Os resultados obtidos foram satisfatórios, evitando uma alta sensibilidade e baixa especificidade ou vice-versa, tanto na validação quanto no teste. A sensibilidade representa a identificação correta das áreas desmatadas classificadas como desmatamento e a especificidade representa a identificação das áreas classificadas como não desmatamento, que são as áreas de floresta.

A imagem mosaico é composta por pequenos fragmentos de amostras de áreas desmatadas e florestas extraídas das imagens LANDSAT. Esses fragmentos de amostras na classe floresta possuem várias texturas de coloração devido às tonalidades da vegetação refletidas e captadas pelo sensor em diferentes condições climáticas.

5.6 Comparação com trabalhos da literatura

Nesta seção apresenta-se uma comparação dos resultados obtidos nesse trabalho com os três principais trabalhos da literatura que utilizaram redes convolucionais, com diferentes abordagens. Conforme apresentado na Tabela 3.1, para avaliar o desempenho dos seus métodos, Ortega et al. (2019) e Adarme et al. (2020) utilizaram as métricas F1-Score e

Acurácia. Já no trabalho De Bem et al. (2020), os autores fizeram uso não somente das métricas de F1-Score e Acurácia, mas também das métricas de sensibilidade e precisão.

No trabalho de Ortega et al. (2019), a aplicação da Rede Convolutacional Siamesa (S-CNN) resultou no valor de F1-Score de 65% e acurácia de 98%. No trabalho de Adarme et al. (2020) a acurácia e F1-Score obtidos foram de 98% e 63%, respectivamente, para a Rede Convolutacional Siamesa. No trabalho de De Bem et al. (2020) os autores alcançaram uma acurácia de 99,93%, F1-Score de 94,32%, sensibilidade de 96,19% e precisão de 92,52%.

Em nosso trabalho, com o melhor modelo selecionado da metodologia proposta, alcançamos um valor de acurácia de 99,973%, F1-Score de 99,973%, sensibilidade de 99,954% e precisão de 99,991%. Ou seja, obtivemos melhores valores para essas métricas do que os registrados na literatura. Observamos que todas as métricas utilizadas para avaliar este trabalho alcançaram valores superiores a 99%, mostrando a excelência da metodologia proposta.

A Tabela 5.10 apresenta a comparação dos resultados obtidos com os trabalhos da literatura com os resultados obtidos com a metodologia proposta.

Tabela 5.10 – Tabela de comparação entre os resultados obtidos com a metodologia proposta e os resultados obtidos em trabalhos da literatura

Literatura	Sensibilidade	Especificidade	Acurácia	Precisão	F1-Score
Ortega et al. (2019)	-	-	98%	-	65%
Adarme et al. (2020)	-	-	98%	-	63%
De Bem et al. (2020)	96,19%	-	99,93%	92,52%	94,32%
Metodologia Proposta	99,95%	99,99%	99,97%	99,99%	99,97%

5.7 CONSIDERAÇÕES

Neste capítulo foi mostrado os resultados obtidos das 36 simulações realizadas. Foi feito uma análise e discutido os resultados. Por fim, realizamos uma comparação entre os resultados obtidos com a metodologia proposta e os resultados obtidos em trabalhos da literatura, mostrando que a mesma obteve desempenho superior.

A seguir, no Capítulo 6, apresentaremos as conclusões e sugestões de trabalhos futuros.

6. CONCLUSÕES E TRABALHOS FUTUROS

Esta dissertação teve como objetivo a segmentação de áreas desmatadas em imagens de satélites Landsat-8 na região Sul da Amazônia Legal, utilizando redes neurais convolucionárias. Para realização deste trabalho foram avaliadas três arquiteturas de redes convolucionárias com aplicação de diferentes métodos de otimização e diferentes métodos para a melhoria da generalização. Além disso, foi feito o uso da técnica de imagem mosaico, devido ao grande desbalanceamento entre áreas desmatadas e áreas de floresta na região de estudo, o que provoca uma grande dificuldade em treinar as redes CNN, pois as mesmas podem se especializar no reconhecimento da classe mais frequente, ou seja, a área de floresta.

Os resultados obtidos deste trabalho foram comparados com os resultados dos trabalhos dos autores Ortega et al. (2019), Adarme et al. (2020) e De Bem et al. (2020), que fizeram uso de redes CNNs. Com a metodologia proposta, os resultados obtidos para as métricas acurácia, F1-Score, sensibilidade, especificidade e precisão com as arquiteturas CNN1 e CNN2 foram todos superiores aos resultados obtidos por esses autores.

De acordo com a análise e discussões dos resultados empreendida no capítulo anterior, concluímos que os três algoritmos de treinamento: SGDM, RMSProp e ADAM, tiveram resultados muito próximos, porém o que apresentou melhor resultado foi o algoritmo RMSProp para as arquiteturas CNN1 e CNN2. Os métodos para melhoria da generalização aplicados nas três arquiteturas tiveram uma pequena influência nos resultados, mostrando uma leve melhoria nos resultados obtidos no conjunto de teste para as três arquiteturas CNN1, CNN2 e CNN3. Além disso, podemos concluir que os blocos de camadas convolutivas, *batch normalization* e ReLU, e as camadas de *max pooling* retiradas da CNN1 tiveram pouca influência ao segmentar as áreas desmatadas nas arquiteturas CNN2 e CNN3, mostrando que as três arquiteturas são capazes de realizar a tarefa de segmentação com resultado satisfatório. Porém, com a metodologia proposta, a arquitetura que apresentou o melhor resultado foi a CNN2 com algoritmo de otimização RMSProp e com método de regularização *dropout*.

A principal contribuição deste trabalho foi avaliar a aplicação dessas três arquiteturas em imagens satelitais para segmentação de áreas desmatadas junto com a abordagem de imagem mosaico. Tal abordagem é precursora na área de sensoriamento remoto, permitindo a obtenção de resultados superiores aos trabalhos previamente publicados na literatura, com valores acima de 99% para todas as métricas estudadas. Constatamos que houve um avanço em relação ao estado da arte na área de sensoriamento remoto. A avaliação das três

arquiteturas com os diferentes algoritmos de otimização e diferentes métodos para a melhoria da generalização contribuíram para esse aprimoramento. Embora os ganhos obtidos nesse estudo tenham sido pequenos, a melhoria incremental contribuiu para o estado da arte.

Em trabalhos futuros, propõe-se avaliar os resultados obtidos das três arquiteturas com redes pré-treinadas como Google-Net, ResNet-18, ResNet-50 e ResNet-101, expandir a área de estudo para outras regiões da Amazônia e regiões do Cerrado, utilizando outras classes de uso de solo, como pasto e agricultura. A expansão da região de estudo é importante, pretende-se utilizar amostras de diferentes regiões (Estados amazônicos) ou de diferentes anos, uma vez que as amostras de diferentes regiões levam em conta a variabilidade espacial e espectral de uma região para a outra. Nesse ponto, deve-se considerar fatores que podem reduzir a acurácia da classificação como o fenômeno do deslocamento de domínio (mudança na distribuição dos dados na amostra). O deslocamento de domínio pode ser associado às mudanças nas condições ambientais, variabilidade geográfica, diferentes propriedades do sensor, nas quais as imagens foram adquiridas, entre outros aspectos (Soto et al., 2020). Além disso, é possível fazer uso de outros tipos de imagens de monitoramento de paisagem como as imagens de radar de abertura sintética (SAR – *Synthetic Aperture Radar*) que são perfeitamente indicadas para a região amazônica, pois não sofrem com a limitação da cobertura de nuvem.

REFERÊNCIAS

- ACHARYA, T. D.; and YANG, I. (2015). **Exploring Landsat 8**. International Journal of IT, Engineering and Applied Sciences Research. Vol. 4.pp. 4-10.
- ADARME, M. O.; FEITOSA, R. Q.; HAPP, P. N.; De ALMEIDA, C. A.; GOMES, A. R. (2020). **Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery**. Remote Sensing. Vol. 12, 910. 10.3390/rs12060910.
- AGGARWAL, C. C. **Neural Networks and Deep Learning**. Springer International Publishing AG, part of Springer Nature, Yorktown Heights, New York-USA, 2018.
- AWAD, M.; KHANNA, R. **Efficient learning machines: Theories, concepts, and applications for engineers and system designers**. New York: Apress Open, 2015.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. (2015). **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation**. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615).
- BARRETO, T. L. M.; ROSA, R. A. S.; WIMMER, C.; NOGUEIRA, J. B.; ALMEIDA, J.; and CAPPABIANCO, F. A. M. **Deforestation change detection using high-resolution multi-temporal X-Band SAR images and supervised learning classification**. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 5201-5204.
- BLANC, L.; GOND, V.; MIN, H. T. D. (2016). **Remote Sensing and Measuring Deforestation**. 10.1016/B978-1-78548-105-5.50002-5.
- BULLOCK, E.; WOODCOCK, C.; and OLOFSSON, P. (2018). **Monitoring Tropical Forest Degradation Using Spectral Unmixing and Landsat Time Series Analysis**. Remote Sensing of Environment. 10.1016/j.rse.2018.11.011.
- COOPS, N. C., TOOKE, T.R. (2017) Introduction to Remote Sensing. In: Gergel S., Turner M. (eds) Learning Landscape Ecology. Springer, New York, NY.
- COSTA, O.; APARECIDO, E.; MATRICARDI, E.; PEDLOWSKI, M.; MIGUEL, E.; and GASPAS, R. (2019). **Selective Logging Detection in the Brazilian Amazon**. Floresta e Ambiente. 26. 1-10. 10.1590/2179-8087.063417.
- DE BEM, P. P.; DE CARVALHO JUNIOR, O. A.; FONTES GUIMARÃES, R.; TRANCOSO GOMES, R. A. ARNALDO. **Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks**. Remote Sensing, Vol. 12 Issue 6, p901. 1p., mar, 2020.
- DINIZ, C. G *et al.*. **DETER-B: The New Amazon Near Real-Time Deforestation Detection System**, in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3619-3628, July 2015.
- FEARNSIDE, P. M. **Desmatamento na Amazônia: dinâmica, impactos e controle**. Acta Amaz., Manaus, v. 36, n. 3, p. 395-400, 2006. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0044-59672006000300018&lng=en&nrm=iso > Acessado em 02 de abril de 2020.

FLORENZANO, T. G. **Iniciação em Sensoriamento Remoto**. 3ª Ed. São Paulo: Oficina de Textos, 2011, Versão Digital 2013. Disponível em: <<https://books.google.com.br/books?id=18GkH5X81XcC&lpg=PT6&lr&hl=pt-BR&pg=PT2#v=onepage&q&f=false>>. Acessado em 20 de março de 2020.

GLASSNER, A. **Deep Learning: From Basics to Practice**. Kindle Edition, Vol. 2, 2018. Disponível em: <https://pt.b-ok2.org/book/3631907/92f63f>.

HAGAN, M.; DEMUTH, H. B.; BEALE, M. H.; JESUS, O. **Neural Network Desing**. 2nd, 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE (2014). **Amazônia Legal Brasileira**. Disponível em: <<https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapas-regionais/15819-amazonia-legal.html>>. Acessado em 20 de abril de 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE): **Shapefiles Malhas Municipais do Estado Amazonas**. Disponível em: <ftp://geofpt.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2015/UFs/AM/am_microrregioes.zip>. Acessado em: 20 de fevereiro de 2020. Brasil, 2019.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE (2019). **Metodologia Utilizada nos Projetos PRODES e DETER**. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/Metodologia_Prodes_Deter_revisada.pdf>. Acessado em 20 de fevereiro de 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE (2020a). **Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite**. Disponível em: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>>. Acessado em 20 de fevereiro de 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE (2020b). **A Estimativa da Taxa de Desmatamento por Corte Raso para Amazônia Legal em 2019 é de 9.762 km²**. Disponível em: <<http://www.obt.inpe.br/OBT/noticias-obt-inpe/a-estimativa-da-taxa-de-desmatamento-por-corte-raso-para-a-amazonia-legal-em-2019-e-de-9-762-km2>>. Acessado em 10 de abril de 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE (2020c). **TerraBrasilis**. Disponível em: <<http://terrabilis.dpi.inpe.br/app/map/deforestation?hl=pt-br>>. Acessado em 7 de abril de 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE (2020d). **Dados Geospaciais da Amazônia Legal Brasileira**. Disponível em: <<http://www.dpi.inpe.br/prodesdigital/dadosn/>>. Acessado em 12 de março de 2020.

JENSEN, R. J. **Sensoriamento Remoto do Ambiente: Uma perspectiva em Recursos Terrestres**. Ed.2, São José dos Campos - São Paulo, 2011.

JOSHI, N.; MITCHARD, E.T.; WOO, N.; TORRES, J.; MOLL-ROCEK, J.; EHAMMER, A.; COLLINS, M.; JEPSEN, M.R.; FENSHOLT, R. **Mapping Dynamics of Deforestation and Forest Degradation in Tropical Forests using Radar Satellite Data**. *Environ. Res. Lett.* 2015, 10, 034014

KHAN, A.; SOHAIL, A.; ZAHOORA, U.; SAEED, A. (2019). **A Survey of the Recent Architectures of Deep Convolutional Neural Networks**.

KIM, K. (2016). **Book Review: Deep Learning**. Healthcare Informatics Research. 22. 351. 10.4258/hir.2016.22.4.351.

KINTISCH, E. (2007). **Carbon emissions. Improved Monitoring of Rainforests Helps Pierce Haze of Deforestation**. *Science*, 316 5824, 536-7 .

LONG, J.; SHELHAMER, E.; and Darrell, T. **Fully Convolutional Networks for Semantic Segmentation**. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3431-3440.

LOUSSAIEF, S.; and ABDELKRIM, A. (2018). **Convolutional Neural Network Hyper-Parameters Optimization based on Genetic Algorithms**. International Journal of Advanced Computer Science and Applications. 9. [10.14569/IJACSA.2018.091031](https://doi.org/10.14569/IJACSA.2018.091031).

MENESES, P.; e ALMEIDA, T. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Brasília, 2012. Disponível em: https://www.researchgate.net/publication/332292728_INTRODUCAO_AO_PROCESSAMENTO_DE_IMAGENS_DE_SENSORIAMENTO_REMOTO/citation/download. Acessado em 12 de fevereiro de 2020.

MIYAGAWA, M. **Segmentação do Lúmen e Identificação de Região de Bifurcações em Imagens de Tomografia por Coerência Óptica Intravascular Utilizando Redes Neurais Convolutivas**. 2019. 88 f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Tecnologia, Universidade Federal do Amazonas, Manaus, 2019.

MINISTÉRIO DO MEIO AMBIENTE - MMA (2018). **Plano de Ação para Prevenção e Controle do Desmatamento e das Queimadas no Cerrado (PPCerrado) e Plano de Ação para Prevenção e Controle do Desmatamento na Amazônia Legal (PPCDAm) Fase 2016-2020**. Volumes I, II e III, Brasília, 2018.

MUKHERJEE, P. **Convolution Neural Networks vs Fully Connected Neural Networks**. Disponível em: <<https://medium.com/datadriveninvestor/convolution-neural-networks-vs-fully-connected-neural-networks-8171a6e86f15>>. Acessado em: 05 de fevereiro de 2020.

OLIVEIRA, J. P.; COSTA, M. G. F.; e COSTA FILHO, C. F. F. **Detecção de Áreas Desmatadas Utilizando Redes Neurais Artificiais Associados a Métodos para Melhoria da Generalização**. Em Anais do Encontro Anual de Computação de 2017 - ENACOMP 2017. Catalão, GO, Brasil.

ORTEGA, M.; CASTRO, J.; NIGRI, H. P.; GOMES, A.; and FEITOSA, R. (2019). **Evaluation of Deep Learning Techniques for Deforestation Detection in the Amazon Forest**. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. IV-2/W7. 121-128. 10.5194/isprs-annals-IV-2-W7-121-2019.

ROSA, R. A. S.; FERNANDES, D.; BARRETO, T. L. M.; WIMMER, C.; and NOGUEIRA, J. B. **Deforestation detection in Amazon rainforest with multitemporal X-band and p-band sar images using cross-coherences and superpixels**. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 3015-3018.

RUDER, S. (2016). **An Overview of Gradient Descent Optimization Algorithms**. Computer Science Machine Learning, Cornell University, Disponível em: <https://arxiv.org/abs/1609.04747>. Consultado em 28 de maio de 2020.

SERRÃO, M. K. M. **Detecção Automática de Bacilos em Baciloscopia de Campo Claro usando Aprendizado Profundo e Técnica de Imagem Mosaico**. 2020. 82 f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Tecnologia, Universidade Federal do Amazonas, Manaus, 2020.

SANTOS, L. **Pooling Layer Introduction**. Disponível em: <https://leonardoaraujosantos.gitbooks.io/artificialintelligence/content/pooling_layer.html>. Acessado em: 05 de fevereiro de 2020.

SETIAWAN, W; UTOYO, M. I.; and RULANINGTYAS, R. **Vessels Semantic Segmentation with Gradient Descent Optimization**. International Journal of Engineering & Technology, [S.l.], v. 7, n. 4, p. 4062-4067, dec. 2018. ISSN 2227-524X. Disponível em: <<https://www.sciencepubco.com/index.php/ijet/article/view/18104/12157>>. Acessado em 03 de março de 2020. doi:<http://dx.doi.org/10.14419/ijet.v7i4.18104>.

SHIMABUKURO, Y. E.; ARAI, E.; DUARTE, V.; JORGE, A.; DOS SANTOS, E. G.; GASPARINI, K. A. C. G; and Dutra, A. C. **Monitoring Deforestation and Forest Degradation Using Multi-Temporal Fraction Images Derived from Landsat Sensor Data in the Brazilian Amazon**. International Journal of Remote Sensing, 40:14, pgs. 5475-5496, 2019, DOI: [10.1080/01431161.2019.1579943](https://doi.org/10.1080/01431161.2019.1579943).

SILVA, I. V.; SPATTI, D. H.; & FLAUZINO, R. A. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas - Curso Prático**. Artliber Editora Ltda, São Paulo, SP, Brasil, 2010.

SOTO, P.; COSTA, G.; FEITOSA, R.; NIGRI HAPP, P.; ORTEGA, M.; NOA TURNES, J.; ALMEIDA, C. and HEIPKE, C. (2020). **Domain Adaptation With CycleGAN For Change Detection In The Amazon Forest**. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIII-B3-2020. pgs.1635;

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; and SALAKHUTDINOV, R. (2014). **Dropout: A Simple Way to Prevent Neural Networks from Overfitting**. Journal of Machine Learning Research. 15. 1929-1958.

UNITED STATES GEOLOGICAL SURVEY – USGS (2019). EarthExplorer. Disponível em: <https://earthexplorer.usgs.gov/>. Acessado em 10 de novembro de 2019.

UNITED STATES GEOLOGICAL SURVEY – USGS (2020). **What are the best spectral bands to use for my study?**. Disponível em: https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research?qt-news_science_products. Acessado em 10 de fevereiro de 2020.

WU, J. 2020. **Convolutional Neural Networks**. Disponível em:<https://cs.nju.edu.cn/wujx/teaching/15_CNN.pdf>. Acessado em 15 de abril de 2020.

ZHANG, A.; LIPTON, Z.C.; Li, M.; SMOLA, A.J. **Dive into Deep Learning**. Dive into Deep Learning Book By: Reference: Introduction to Deep Learning STAT 157, University of California, Berkeley, Spring, UC Berkeley, Release 0.7.1, february 2020. Disponível em: <<https://d2l.ai/d2l-en.pdf>>. Consultado em 03 de março de 2020.