



**PODER EXECUTIVO  
MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO AMAZONAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**



JOEL PARENTE DE OLIVEIRA

METODOLOGIA DE FUSÃO DE DADOS USANDO APRENDIZADO PROFUNDO  
PARA SEGMENTAÇÃO SEMÂNTICA DE USOS DE SOLO NA AMAZÔNIA

MANAUS  
2021

UNIVERSIDADE FEDERAL DO AMAZONAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

JOEL PARENTE DE OLIVEIRA

METODOLOGIA DE FUSÃO DE DADOS USANDO APRENDIZADO PROFUNDO  
PARA SEGMENTAÇÃO SEMÂNTICA DE USOS DE SOLO NA AMAZÔNIA

Tese apresentada ao Curso de Doutorado em  
Informática, área de concentração Ciência da  
Computação do Programa de Pós-Graduação  
em Informática da Universidade Federal do  
Amazonas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho  
CoOrientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Marly Guimarães Fernandes Costa

MANAUS  
2021

## Ficha Catalográfica

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

O48m Oliveira, Joel Parente de  
Metodologia de fusão de dados usando aprendizado profundo para segmentação semântica de usos de solo na Amazônia / Joel Parente de Oliveira . 2021  
125 f.: il. color; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho  
Coorientadora: Marly Guimarães Fernandes Costa  
Tese (Doutorado em Informática) - Universidade Federal do Amazonas.

1. aprendizado profundo. 2. redes neurais convolucionais. 3. sensoriamento remoto. 4. segmentação de imagem. 5. segmentação semântica. I. Costa Filho, Cícero Ferreira Fernandes. II. Universidade Federal do Amazonas III. Título



PODER EXECUTIVO  
MINISTÉRIO DA EDUCAÇÃO  
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



## FOLHA DE APROVAÇÃO

**“METODOLOGIA DE FUSÃO DE DADOS USANDO  
APRENDIZADO PROFUNDO PARA SEGMENTAÇÃO  
SEMÂNTICA DE USOS DE SOLO NA AMAZÔNIA”**

**JOEL PARENTE DE OLIVEIRA**

Tese de Doutorado defendida e aprovada pela banca examinadora constituída pelos Professores:

  
Prof. Cícero Ferreira F. Costa Filho – PRESIDENTE

  
Prof. José Raimundo Gomes Pereira – MEMBRO EXTERNO

  
Dr. Clahildek Matos Xavier – MEMBRO EXTERNO

  
Dr. Carlos Benedito Santana da Silva Soares - MEMBRO EXTERNO

  
Dr. Luis Fernando Bueno - MEMBRO EXTERNO

Manaus, 29 de Janeiro de 2021

## AGRADECIMENTOS

Primeiro a Jeová Deus, que conforme **Apocalipse 4:11**, Ele merece toda glória, honra e poder pois criou todas as coisas e tudo veio a existência segundo a sua vontade.

A minha querida esposa Evelyn Oliveira pelo apoio, incentivos e compreensão durante todo o período de realização do meu doutorado.

Aos meus pais, José Parente e Izabel Oliveira que sempre incentivaram minha formação e sempre me deram o apoio necessário.

Ao meu orientador, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e a Prof. Marly Guimarães Fernandes Costa pelo apoio e grande ajuda prestada durante todo o desenvolvimento do trabalho. Meus sinceros agradecimentos.

Ao meu irmão Prof. Dr. Jozias Parente de Oliveira que sempre esteve disponível para tirar dúvidas e sugerir correções no trabalho.

Aos membros da banca examinadora: Clahildek Xavier, Luis Fernando Bueno, Carlos Soares e José Raimundo.

À UFAM, ao PPGI e ao PPGEE/CETELI pela oportunidade e estrutura disponibilizada.

E a todos que contribuíram de alguma forma para a realização deste trabalho.

## RESUMO

Este estudo propõe uma metodologia que utiliza aprendizado profundo e um algoritmo de segmentação multiresolução para realizar a segmentação semântica de imagens de sensoriamento remoto. O objetivo da segmentação semântica é classificar o uso do solo em três regiões: floresta, pasto e agricultura. Inicialmente, a imagem é segmentada usando uma rede convolucional. Em seguida, uma imagem com regiões homogêneas é gerada usando um algoritmo de segmentação de multiresolução. Finalmente, um processo de fusão de dados é proposto para fundir as informações provenientes desses dois processos de segmentação. O campo de estudo foram áreas da região amazônica brasileira. Os dados de entrada utilizados foram imagens LANDSAT-8/OLI. Os dados de referência foram extraídos dos resultados do projeto TerraClass em 2014. Dois conjuntos de dados foram avaliados: o primeiro com seis bandas e o segundo com três bandas. Três arquiteturas de redes convolucionais foram avaliadas juntamente com três métodos de otimização, SGDM, ADAM, e RMSProp, e com dois métodos para melhoria da generalização: dropout e regularização  $L_2$ . O melhor modelo, definido como a associação de uma arquitetura da rede, de um método de otimização e de um método de generalização, que teve o melhor desempenho no conjunto de validação, foi submetido a uma metodologia de validação cruzada de 5 pastas. Os resultados obtidos com os modelos propostos foram comparados com redes pré-treinadas usando a metodologia de transferência de conhecimento. Com esse objetivo foram utilizadas as seguintes redes pré-treinadas: ResNet50, InceptionResnetv2, MobileNetv2 e Xception. Finalmente, a metodologia proposta foi avaliada em regiões utilizadas por outros autores. Os valores de acurácia obtidos para as imagens avaliadas foram superiores a 99%, o que mostra a excelência da técnica de classificação de uso de solo desenvolvida nesse trabalho para classificação de imagens de sensoriamento remoto.

**Palavras-chave:** aprendizado profundo; redes neurais convolucionais; sensoriamento remoto; segmentação de imagem

## ABSTRACT

This study proposes a methodology that uses deep learning and a multiresolution segmentation algorithm to perform the semantic segmentation of remote sensing images. The objective of the semantic segmentation is to classify the land use in three regions: forest, pasture and agriculture. Initially, the image is segmented using a convolutional network. Then, an image with homogeneous regions is generated using a multiresolution segmentation algorithm. Finally, a data fusion process is proposed to merge the information from these two segmentation processes. The field of study were areas of the Brazilian Amazon region. The input data used were LANDSAT-8/OLI images. The reference data were extracted from the results of the TerraClass project in 2014. Two sets of data were evaluated: the first with six bands and the second with three bands. Three convolutional network architectures were evaluated along with three optimization methods, SGDM, ADAM, and RMSProp, and two methods for generalization improvement: dropout and L2 regularization. The best model, defined as the association of a network architecture, an optimization method and a generalization method, which had the best performance in the validation set, was submitted to a 5-folder cross validation methodology. The results obtained with the proposed models were compared with pre-trained networks using the knowledge transfer methodology. For this purpose the following pre-tested networks were used: ResNet50, InceptionResnetv2, MobileNetv2 and Xception. Finally, the proposed methodology was evaluated in regions used by other authors. The accuracy values obtained for the images evaluated were higher than 99%, which shows the excellence of the land use classification technique developed in this work for the classification of remote sensing images.

**Keywords:** Deep learning; convolutional neural networks; remote sensing; image segmentation.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Região avaliada por Lee, Cardille e Coe, (2020). Fonte: Lee, Cardille e Coe, (2020) .....	25
Figura 2 – Exemplo de uma imagem avaliada por Lee, Cardille e Coe, (2020). Em (a) temos a imagem LANDSAT-5/TM com uma composição RBG utilizando as bandas 4,5,7. Em (b) temos a imagem classificada. A acurácia obtida foi de 80%. Fonte: Lee, Cardille e Coe, (2020) .....	25
Figura 3 – Degradação e Desmatamento calculado por Bullock, Woodcock e Olofsson, (2020) para os anos de 1994 a 2013. Fonte: Bullock, Woodcock e Olofsson, (2020).....	26
Figura 4 – Região avaliada por Shimabukuro <i>et al.</i> , (2019). Em (a) é mostrado as unidades da Federação do Brasil e em destaque o estado de Mato Grosso. Em (b) temos a localização da área de estudo no estado de Mato Grosso. Em (c), tem uma imagem LANDSAT-8/OLI numa composição RGB utilizando as bandas 6, 5 e 4. A região laranja corresponde a região utilizada para o treinamento e a região em amarelo corresponde a região de validação. Fonte: Shimabukuro <i>et al.</i> , (2019).....	27
Figura 5 – Imagem de 2013 classificada. A acurácia obtida foi de 94,93%. Fonte: Shimabukuro <i>et al.</i> , (2019).....	27
Figura 6 – Área de estudo de Lima <i>et al.</i> , (2019). Fonte: Lima <i>et al.</i> , (2019).....	28
Figura 7 – Área de estudo de Nascimento <i>et al.</i> , (2020). Fonte: Nascimento <i>et al.</i> , (2020) ....	29
Figura 8 – Resultados da classificação da região de estudo de Maretto <i>et al.</i> , (2020). A acurácia obtida foi de 95% aproximadamente. Fonte: Maretto <i>et al.</i> , (2020) .....	29
Figura 9 – Localização das regiões de estudo na região amazônica com os locais de treinamento (b, c) A e B e (d) local de teste C como imagens compostas LANDSAT-8/OLI em cores verdadeiras, tiradas de junho e julho de 2018.. Fonte: Bem <i>et al.</i> (2020).....	30
Figura 10 – Região do bioma Amazônia brasileira, localizada no estado do Pará, Brasil. Em (a,b) são apresentadas uma composição colorida (Infravermelho próximo, verde, azul) da área de estudo em duas datas diferentes T1 e T2; (c) Referência do processo de desmatamento de 2016 a 2017. Fonte: Adarme <i>et al.</i> (2020).....	32
Figura 11 – Região do bioma Cerrado Brasileiro, localizada no estado do Maranhão, Brasil. Em (a,b) são apresentadas uma composição colorida (Infravermelho próximo, verde, azul) da área de estudo em duas datas diferentes T1 e T2; (c) Referência do processo de desmatamento de 2017 a 2018. Fonte: Adarme <i>et al.</i> (2020).....	32
Figura 12 – Obtenção de imagem por sensoriamento remoto. Fonte: (FLORENZANO, 2011) .....	40
Figura 13 – Espectro eletromagnético. Fonte: (MUNDOGEO, 2004).....	41
Figura 14 – Exemplo do efeito da resolução espectral no contraste do rio com a vegetação. Imagens da região amazônica nas bandas: (a) visível e (b) infravermelho próximo. Fonte: (MENESES e ALMEIDA, 2012) .....	44
Figura 15 – Exemplos de imagens com diferentes níveis de quantização ou de resolução radiométrica. Fonte: (MENESES e ALMEIDA, 2012).....	45
Figura 16 – Limite da Amazônia Legal Brasileira e área não mapeada pelo PRODES. Fonte: (INPE, 2019b).....	48
Figura 17 – Exemplo de uma região de Floresta. Em (a) temos uma região da cena 230/065 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de floresta. Fonte: (INPE, 2016). .....	49

Figura 18 – Exemplo de uma região de Agricultura. Em (a) temos uma região da cena 227/068 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de agricultura. Fonte: (INPE, 2016).....	49
Figura 19 – Exemplo de uma região de Pasto. Em (a) temos uma região da cena 227/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de pasto. Fonte: (INPE, 2016).....	50
Figura 20 – Exemplo de uma área não-observada. Em (a) temos uma região da cena 226/062 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma não-observada. Fonte: (INPE, 2016). ....	50
Figura 21 – Exemplo de uma região de área urbana. Em (a) temos uma região da cena 232/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área urbana. Fonte: (INPE, 2016).....	51
Figura 22 – Exemplo de uma região de mineração. Em (a) temos uma região da cena 232/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de mineração. Fonte: (INPE, 2016). ....	51
Figura 23 – Exemplo de uma região caracterizada como outros. Em (a) temos uma região da cena 230/062 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área chamada de outros. Fonte: (INPE, 2016). ....	52
Figura 24 – Exemplo de uma convolução. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016) .....	55
Figura 25 – Exemplo de operação de <i>pooling</i> utilizando o valor máximo. A entrada possui um tamanho de 7x7 e a saída foi calculada utilizando o valor máximo e os passos 1 e 2. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016).....	56
Figura 26 – Exemplo de convolução transposta com passo 1. Fonte: (SERRÃO, 2020) .....	58
Figura 27 – Exemplo de uso da camada de <i>dropout</i> . A rede da direita corresponde a aplicação do <i>dropout</i> na rede da esquerda. Fonte: (SRIVASTAVA <i>et al.</i> , 2014).....	58
Figura 28 – Exemplo de segmentação semântica. Fonte: (MATHWORKS, 2018).....	60
Figura 29 – Efeito do momento na suavização de oscilações: (a) método SGD (b) método SGDM. Fonte: Adaptado de Aggarwal (2018).....	61
Figura 30 – Diagrama em blocos da metodologia empregada .....	65
Figura 31 – Amazônia Legal com as cenas LANDSAT-8/OLI avaliadas neste trabalho. ....	66
Figura 32 – Fluxograma da metodologia utilizada para a obtenção das Imagens do Mosaico	68
Figura 33 – Retalhos de imagens apresentados composição colorida utilizando as bandas B6-B5-B4. (a) floresta, (b) pasto e (c) agricultura .....	68
Figura 34 – Exemplo de imagem em mosaico. Em (a) é mostrada uma imagem em mosaico na composição colorida utilizando as bandas B6-B5-B4, composta de retalhos de agricultura, floresta e de pasto, em (b) a imagem padrão ouro, e em (c), (d) e (e) temos o padrão ouro para cada retalho de imagem, correspondente à área agricultura, de pasto e floresta, respectivamente. ....	70
Figura 35 – Etapas da Metodologia de Segmentação Semântica utilizada neste trabalho. ....	71
Figura 36 – Arquiteturas CNNs utilizadas neste trabalho. Os números abaixo de algumas camadas correspondem ao número de filtros. Este número representa o número de neurônios na camada convolucional que se ligam à mesma região na entrada.....	73
Figura 37 – Fluxograma do processo de refinamento da classificação da imagem. ....	79
Figura 38 – Ilustração do processo de fusão: (a) imagem original; (b) imagem resultante do algoritmo de segmentação de região homogênea. Dentro de pequenos retângulos, três pequenas regiões são selecionadas e ampliadas. A região inferior tem mais pixels cinza, por	

isso é classificada como pasto. A região central tem mais pixels pretos, portanto é classificada como agricultura, e como há mais pixels brancos na região superior, é classificada como floresta. ....	79
Figura 39 – Gráfico com o desempenho (acurácia) de cada modelo CNN proposto neste trabalho no conjunto de validação, com imagens de 6 bandas. ....	83
Figura 40 – Gráfico com o desempenho (acurácia) de cada modelo CNN proposto neste trabalho no conjunto de validação, com imagens de 3 bandas. ....	86
Figura 41 – LANDSAT-8/OLI imagens das regiões amazônicas. (a) imagem original exibida em composição colorida das bandas B6-B5- B4; (b) padrão ouro; (c) classificação semântica feita pela CNN2, com uma acurácia de 96,84%. (d) imagem de saída resultante da segmentação multiresolução; (e) resultado da classificação da imagem após processo de fusão de dados, com uma acurácia de 99,80%. ....	89
Figura 42 – LANDSAT-8/OLI imagens da região Amazônica. (a) imagens originais exibidas em composição colorida B6-B5-B4; (b) imagem padrão em ouro; (c) imagens classificadas pelo modelo CNN2/ADAM/L <sub>2</sub> e processo de fusão. As acurácias obtidas para as imagens 1, 2, 3, 4 e 5 foram de 99,63%, 99,08%, 99,66%, 99,83% e 99,85% respectivamente. ....	90
Figura 43 – Exemplo de imagem-mosaico para a região utilizada por Bem <i>et al.</i> (2020). Em (a) é a imagem-mosaico mostrada em uma composição colorida B6-B5-B4 e (b) o padrão ouro correspondente. ....	95
Figura 44 – Exemplo de imagem-mosaico para a região utilizada por Adarme <i>et al.</i> (2020). Em (a) é a imagem-mosaico mostrada em uma composição colorida B6-B5-B4 e (b) o padrão ouro correspondente. ....	96
Figura 45 – Imagem LANDSAT-8/OLI da região amazônica. (a) Imagem de uma região (3200x3200pixels) da cena LANDSAT-8/OLI 230/65 avaliada por Bem <i>et al.</i> (2020) exibida em composição colorida B6-B5-B4. (b) mostra imagem classificada, com acurácia de 99,91%. ....	97
Figura 46 – Imagem LANDSAT-8/OLI da região amazônica. Em (a), imagem de uma região (2000x1200pixels) da cena LANDSAT-8/OLI 225/62 avaliada por Adarme <i>et al.</i> (2020) exibida em composição colorida B6-B5-B4. Em (b), a imagem é classificada com acurácia de 99,94%. ....	97

## LISTA DE TABELAS

Tabela 1 – Valores médios de acurácia para a região 227/68 (Barbosa <i>et al.</i> , 2015).....	24
Tabela 2 – Datas de aquisição para cada local e cenas correspondentes do LANDSAT-8/OLI Fonte: (Bem <i>et al.</i> (2020)) .....	31
Tabela 3 – Bandas LANDSAT-8/OLI. Fonte: (U.S. GEOLOGICAL SURVEY, 2019).....	46
Tabela 4 – Algumas das arquiteturas mais difundidas FONTE: (GUO <i>et al.</i> , 2016).....	60
Tabela 5 – Lista de Imagens utilizadas.....	65
Tabela 6 – Divisão dos retalhos de imagens.....	69
Tabela 7 – Banco de Imagens-mosaicos gerados com a quantidade para os conjuntos de treinamento, validação e teste.....	70
Tabela 8 – Parâmetros de Treinamento das CNNs.....	74
Tabela 9 – Experimentos da metodologia de validação cruzada com 5-pastas.....	76
Tabela 10 – Características do computador utilizado.....	80
Tabela 11 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, com imagens de 6 bandas. ....	82
Tabela 12 – Desempenho dos melhores modelos de cada CNN no conjunto de validação, com imagens de 6 bandas. ....	83
Tabela 13 – Desempenho do melhor modelo CNN1/RMSProp utilizando a metodologia de validação cruzada com 5-pastas e imagens de 6 bandas. ....	84
Tabela 14 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, com imagens de 3 bandas. ....	84
Tabela 15 – Desempenho dos melhores modelos de cada CNN, no conjunto de validação, com imagens de 3 bandas. ....	86
Tabela 16 – Desempenho do modelo CNN1/RMSProp/Dropout utilizando uma metodologia de validação cruzada com 5-pastas, com imagens de três bandas.....	87
Tabela 17 – Acurácia utilizando transferência de conhecimento para o conjunto de validação. .....	87
Tabela 18 – Resultados obtidos para a segmentação semântica das imagens da Figura 42 utilizando os modelos CNN2/ADAM/L <sub>2</sub> e Fusão de Dados .....	91
Tabela 19 – Divisão dos retalhos de imagens 20x20 <i>pixels</i> .....	91
Tabela 20 – Divisão dos retalhos de imagens 10x10 <i>pixels</i> .....	92
Tabela 21 – Banco de Imagens-mosaicos gerados com a quantidade para os conjuntos de treinamento e validação .....	92
Tabela 22 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, treinadas com imagens de 6 bandas e imagens mosaico com retalhos 20x20 <i>pixels</i> .....	92
Tabela 23 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, treinadas com imagens de 6 bandas e imagens mosaico com retalhos 10x10 <i>pixels</i> .....	92
Tabela 24 – Resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelos CNNs treinados com imagens mosaico compostas de retalhos 20x20 <i>pixels</i> .....	93
Tabela 25 – Resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelos CNNs treinados com imagens mosaico compostas de retalhos 10x10 <i>pixels</i> .....	93

Tabela 26 – Comparação dos resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelo CNNs treinados com imagens mosaico compostas de retalhos e utilizando o processo de fusão de dados .....	94
Tabela 27 – Bancos de imagens montados utilizando as imagens avaliadas por Bem <i>et al.</i> (2020) e Adarme <i>et al.</i> (2020) .....	95
Tabela 28 – Desempenho de três melhores modelos de cada CNN usando as mesmas regiões utilizadas por Bem <i>et al.</i> (2020) e Adarme <i>et al.</i> (2020) para os respectivos conjuntos de validação .....	96
Tabela 29 – Resultados obtidos para a segmentação semântica das imagens da Figura 45 e Figura 46 utilizando os modelos CNN2/ADAM/L <sub>2</sub> e Fusão de Dados .....	98

## LISTA DE ABREVIATURAS E SIGLAS

ACCG	Acurácia Global
ACCM	Acurácia Média
ADAM	Estimativa de Dinâmica Adaptativa ( <i>Adaptive Moment Estimation</i> )
B2	Banda Azul
B3	Banda Verde
B4	Banda Vermelha
B5	Banda Infravermelho próximo
B6	Banda Infravermelho de ondas curtas 1
B7	Banda Infravermelho de ondas curtas 2
CNN	Redes Convolucionais
EVI	Índice de Vegetação Melhorado
F1	<i>Score F1</i>
FN	Falso Negativo ( <i>False Negative</i> )
FP	Falso Positivo ( <i>False Positive</i> )
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
J	Coefficiente de similaridade de Jaccard
JP	Coefficiente de similaridade Jaccard ponderado
LANDSAT	<i>Land Remote Sensing Satellite</i>
LANDSAT-5/TM	<i>Land Remote Sensing Satellite / Thematic Mapper</i>
LANDSAT-7/ETM	<i>Land Remote Sensing Satellite / Enhanced Thematic Mapper Plus</i>
LANDSAT-8/OLI	<i>Land Remote Sensing Satellite / Operational Terra Imager</i>
L <sub>2</sub>	Regularização L <sub>2</sub>
LULC	Uso e Cobertura da Terra ( <i>Land Use and Land Cover</i> ).
MLP	Perceptron Multicamadas
NDVI	Índice de Vegetação de Diferença Normalizada
OBT	Observação da Terra
OLI	<i>Operational Land Imager</i>
PRODES	Programa de Monitoramento do Desflorestamento na Amazônia Legal

RMSProp <i>Propogation</i> )	Propagação da Raiz Média Quadrática ( <i>Root Mean Square</i> )
RNA	Rede Neural Artificial
SAR	Synthetic Aperture Radar – Radar de Abertura Sintética
SENS	Sensibilidade
SGDM <i>Descent with Momentum</i> )	Gradiente Descendente Estocástico com Momento ( <i>Stochastic Gradient</i> )
SPRING	Sistema de Processamento de Informações Georeferenciadas
SSP	<i>Spatial pyramid pooling</i>
SVM	<i>Support Vector Machine</i>
TIRS	<i>Thermal InfraRed Sensor</i>
TN	Verdadeiro Negativo ( <i>True Negative</i> )
TP	Verdadeiro Positivo ( <i>True Positive</i> )

## **LISTA DE QUADROS**

Quadro 1 – Quadro Comparativo da Revisão Bibliográfica.....	35
---	----

# SUMÁRIO

<b>CAPÍTULO 1</b> .....	18
<b>1 INTRODUÇÃO</b> .....	18
<b>1.1 OBJETIVOS</b> .....	21
<b>1.1.1 GERAL</b> .....	21
<b>1.1.2 ESPECÍFICOS</b> .....	21
<b>1.2 ORGANIZAÇÃO</b> .....	22
<b>CAPÍTULO 2</b> .....	23
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	23
<b>CAPÍTULO 3</b> .....	39
<b>3 FUNDAMENTAÇÃO TEÓRICA</b> .....	39
<b>3.1 Sensoriamento Remoto</b> .....	39
<b>3.2 LANDSAT</b> .....	45
<b>3.3 Projeto TerraClass</b> .....	46
<b>3.4 Aprendizado profundo</b> .....	52
<b>3.5 Métodos de otimização</b> .....	61
<b>3.5.1 SGDM</b> .....	61
<b>3.5.2 RMSProp</b> .....	62
<b>3.5.3 Adam</b> .....	62
<b>3.6 Transferência de conhecimento (<i>Transfer Learning</i>)</b> .....	63
<b>CAPÍTULO 4</b> .....	64
<b>4 MATERIAIS E MÉTODOS</b> .....	64
<b>4.1 Imagem LANDSAT-8/OLI</b> .....	65
<b>4.2 Seleção de Bandas</b> .....	67
<b>4.3 Segmentação Semântica com Redes Convolucionais</b> .....	70
<b>4.4 Segmentação de Regiões Homogêneas</b> .....	77
<b>4.5 Fusão de Dados</b> .....	78
<b>4.6 Ambiente de Desenvolvimento</b> .....	79
<b>CAPÍTULO 5</b> .....	81
<b>5 RESULTADOS E DISCUSSÕES</b> .....	81
<b>5.1 Resultados das Simulações (Imagens com seis bandas)</b> .....	82
<b>5.2 Resultados das Simulações (Imagens com três bandas)</b> .....	84
<b>5.3 Resultados das Simulações utilizando transferência de conhecimento</b> .....	87
<b>5.4 Classificação/Segmentação de imagens da Região Amazônica</b> .....	87

<b>5.5 Avaliação da Metodologia em Regiões usadas por Adarme et al. (2020) e Bem et al. (2020)</b> .....	94
<b>CAPÍTULO 6</b> .....	99
<b>6 CONCLUSÕES</b> .....	99
<b>7 REFERÊNCIAS</b> .....	104
<b>APÊNDICE 1</b> .....	108
<b>APÊNDICE 2</b> .....	114

---

## 1 INTRODUÇÃO

Sensoriamento remoto consiste na coleta de informações sobre um objeto ou área geográfica sem contato direto. Estas informações são coletadas através de várias tecnologias, tais como sensores instalados a bordo de aeronaves, satélites, ou outras plataformas (JENSEN, 2014). Os dados gerados a partir de sistemas de sensoriamento remoto são de grande utilidade para várias aplicações, dentre as quais podemos citar: urbanas (interferência demográfica, cadastro, planejamento urbano, suporte ao setor imobiliário); agrícolas (condição de culturas, previsão de safras, erosão de solos); geológicas (minerais, petróleo, gás natural); ecológicas (regiões alagadas, solos, florestas, oceanos, águas continentais); florestais (produção de madeira, estimativa de biomassa, monitoramento de desmatamento) (NOVO, 2008).

Com relação ao monitoramento do desmatamento na Amazônia por sensoriamento remoto, o Instituto Nacional de Pesquisas Espaciais (INPE) é uma referência mundial. Entre os vários projetos desenvolvidos pelo INPE para este monitoramento, destacam-se o Programa de Monitoramento do Desmatamento na Amazônia Legal (PRODES) e o TerraClass (INPE, 2019b, a).

O PRODES fornece dados por meio de mapas anuais de desmatamento na região Amazônica. O projeto TerraClass utiliza os dados gerados pelo PRODES para realizar uma classificação de uso e cobertura de solo. Neste último caso, o objetivo é identificar os vários usos de solo, como: floresta, pasto, agricultura, áreas urbanas, mineração e outros. Esse tipo de informação pode ajudar os órgãos a desenvolverem políticas públicas de prevenção para conter o avanço do desmatamento. Em 2018 e 2019, o desmatamento na Amazônia foi de 7.535 km<sup>2</sup> e 10.129 km<sup>2</sup>, respectivamente. Portanto, a área desmatada apresentou um aumento de 34% de 2018 a 2019. Em 2019, 84,56% de todo o desmatamento observado na Amazônia ocorreu nos estados do Pará, Mato Grosso, Amazonas e Rondônia (INPE, 2019c). Embora PRODES e TerraClass sejam grandes projetos e forneçam dados muito confiáveis, eles ainda dependem de uma parcela significativa de trabalho realizado pela intervenção humana. A fim de acelerar e tornar o processo mais eficiente, vários estudos têm sido feitos utilizando ferramentas de aprendizagem de máquinas (ADARME *et al.*, 2020; BARBOSA *et al.*, 2015; BEM *et al.*, 2020; BULLOCK; WOODCOCK; OLOFSSON, 2018; CASSOL *et al.*, 2020; LEE; CARDILLE; COE, 2020; LESEUX *et al.*, 2019; LIMA *et al.*, 2019; MARETTO *et al.*, 2020; NASCIMENTO *et al.*, 2020; SHIMABUKURO *et al.*, 2019).

Em Lima *et al.*, (2019), Nascimento *et al.*, (2020) e Shimabukuro *et al.*, (2019), os autores realizaram segmentação de imagem utilizando classificação por limiar. Lima *et al.*, (2019) comparam o desempenho dos dados LANDSAT-8/OLI e Sentinel-2 para a detecção de corte seletivo em uma área localizada na Amazônia brasileira. As imagens de ambos os satélites tiveram desempenho muito próximos em termos de acurácia, com valores de 96,7% e 95,7% para Sentinel-2 e LANDSAT-8/OLI, respectivamente. Nascimento *et al.*, (2020) classificaram imagens de sensoriamento remoto nas classes de cerrado, floresta, água, área de mineração e áreas de reflorestamento. A acurácia obtida pelos autores foi de aproximadamente 96%, em média. Em Shimabukuro *et al.*, (2019), os autores tinham como objetivo classificar regiões de desmatamento, áreas de degradação e regiões de queimada. A acurácia obtida pelos autores foi de 94,93%, em média.

Nos trabalhos de Barbosa *et al.*, (2015), Lee, Cardille e Coe, (2020) e Cassol *et al.*, (2020), os autores desenvolveram uma metodologia para classificação de regiões de agricultura, floresta e pasto. Barbosa *et al.*, (2015) obteve uma acurácia média de 86,04% utilizando o classificador Máquina de Vetores de Suporte (SVM). Em Lee, Cardille e Coe, (2020), os autores utilizaram um método Bayesiano de classificação não Supervisionado (BULC-U) e obtiveram uma acurácia de 80%. Cassol *et al.*, (2020) classificaram imagens utilizando o método *random forests* e alcançaram uma acurácia de 88%.

Segundo Aggarwal, (2018), a partir da primeira década deste século, a rede neural renasceu sob o novo rótulo conhecido como aprendizado profundo. O aprendizado profundo faz uso de modelos computacionais com arquiteturas hierárquicas compostas de múltiplas camadas de processamento, a fim de "aprender" representações de dados em formatos muito diferentes: áudio, imagens e texto (LECUN, BENGIO e HINTON, 2015). De acordo com Aggarwal (2018), a grande quantidade de dados disponíveis nos últimos anos, juntamente com o aumento do poder computacional, com as unidades gráficas de processamento (*Graphical Processing Units – GPUs*), permitiu o uso de arquiteturas mais profundas que antes não eram possíveis.

Os trabalhos de Adarme *et al.* (2020), Bem *et al.* (2020) e Maretto *et al.* (2020), utilizaram técnicas de aprendizado profundo para detectar áreas desmatadas em regiões da Amazônia brasileira. Os resultados obtidos por estes autores foram bastante satisfatórios. Os autores alcançaram acurácias acima de 98% na detecção de desmatamento.

Por essa breve revisão da literatura, que é confirmada mais adiante, no capítulo de revisão bibliográfica, onde foi realizada uma revisão mais profunda, constatou-se que, diferentemente da proposta desta tese, os trabalhos encontrados na literatura não apresentam

um modelo em aprendizado profundo para a classificação de uso/cobertura da terra para a região Amazônica Brasileira, mas tão somente para detecção de desmatamento. Portanto, isso abre novas possibilidades para pesquisas neste campo.

Um grande desafio no treinamento de redes convolucionais (*Convolutional Neural Networks* – CNNs) com imagens de sensoriamento remoto da região Amazônica é que as classes são desequilibradas. Em outras palavras, em uma região capturada por uma imagem de satélite, a maior parte corresponde a floresta e poucas áreas correspondem a outros usos de solo, como pasto e agricultura, de forma que a região de floresta é predominante. Este problema pode levar, no treinamento da CNN, a que o método de otimização tenha melhor desempenho nas classes com uso de solo mais presentes (floresta).

Para resolver o problema citado no parágrafo anterior, referente a distribuição desigual de uso de solo nas imagens de sensoriamento remoto, propomos a técnica de imagem mosaico para o treinamento da CNN. Nesta técnica, pequenas amostras retangulares de agricultura, floresta e pasto são extraídas de imagens de satélite. A partir destas amostras, é criada uma imagem maior, denominada de imagem mosaico, com aproximadamente o mesmo número de pequenos retângulos de floresta, agricultura e pasto.

Outra contribuição deste trabalho é a melhoria dos resultados obtidos com a CNN na classificação do uso de solo, por meio de um processo de fusão de dados. Esse processo consiste em combinar a imagem de saída da CNN com uma imagem gerada por meio de um algoritmo de segmentação de regiões homogêneas. Os resultados da fusão superam os resultados obtidos apenas com a CNN.

Neste trabalho, outra contribuição importante é a comparação efetuada entre o desempenho de arquiteturas CNN customizadas (*scratch networks*) com os resultados obtidos através da técnica de transferência de conhecimento (*Transfer Learning* - TL), utilizando treinamento raso das seguintes arquiteturas: ResNet50 (RN50), InceptionResnetv2 (IR2), MobileNetv2 (MN2), e Xception (XC) (CHOLLET, 2017; HE *et al.*, 2016; SANDLER *et al.*, 2018; SZEGEDY *et al.*, 2017). Ressalta-se que, no treinamento raso, apenas a última camada de classificação é treinada na transferência de conhecimento.

Por fim, a metodologia proposta inicialmente para classificação do uso do solo, é também aplicada para detecção de desmatamento, obtendo resultados promissores em ambas as aplicações. Dentro desse contexto mais amplo de aplicações, que incluem tanto a classificação do uso do solo, quanto a detecção de áreas desmatadas, procede-se uma comparação do desempenho da metodologia proposta com o desempenho de outros métodos

publicados anteriormente na literatura. Para tanto, a metodologia proposta foi aplicada nas mesmas regiões avaliadas por outros autores.

## 1.1 OBJETIVOS

### 1.1.1 GERAL

O objetivo geral deste trabalho é propor uma metodologia por meio de técnicas de aprendizado profundo e um algoritmo de segmentação multiresolução para realizar a segmentação semântica de imagens de sensoriamento remoto em classes de uso do solo na região Amazônica.

### 1.1.2 ESPECÍFICOS

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

- 1) Avaliar o desempenho de arquiteturas de redes convolucionais para segmentação semântica de imagens de sensoriamento remoto em classes de uso do solo em regiões dos seguintes estados da Amazônia Legal Brasileira: Amazonas, Rondônia, Mato Grosso e Pará;
- 2) Caracterizar o desempenho dessas arquiteturas em função de métodos utilizados para melhoria da generalização, como: regularização  $L_2$  e *dropout*;
- 3) Comparar o desempenho da arquitetura proposta com o desempenho de arquiteturas pré-treinadas, por meio da técnica de transferência de conhecimento;
- 4) Utilizar o método desenvolvido para segmentação semântica de uso do solo em problemas de detecção de áreas desmatadas e não desmatadas;
- 5) Comparar o desempenho da metodologia proposta com o desempenho de outros métodos publicados na literatura
- 6) Propor a melhoria do desempenho das redes convolucionais na classificação do uso do solo e na detecção do desmatamento, através da fusão de dados entre a segmentação semântica das redes convolucionais com o método de segmentação de regiões homogêneas;
- 7) Propor a técnica de imagem mosaico para o pré-processamento dos dados de entrada da CNN;

- 8) Avaliar o melhor conjunto de bandas (3 ou 6) para classificação de imagens de sensoriamento em classes de uso do solo na região Amazônica.

## 1.2 ORGANIZAÇÃO

Este trabalho está organizado em 6 capítulos descritos a seguir:

- Introdução (Capítulo 1);
- Revisão Bibliográfica (Capítulo 2);
- Fundamentação Teórica (Capítulo 3);
- Materiais e Métodos (Capítulo 4);
- Resultados e Discussões (Capítulo 5);
- Conclusões (Capítulo 6);
- Referências Bibliográficas;
- Apêndice.

No Capítulo 1, é apresentada a introdução deste trabalho.

No Capítulo 2, identificam-se os principais sistemas de classificação de imagens de satélite até então publicados, por meio de uma pesquisa bibliográfica nas principais bases de dados literárias.

No Capítulo 3, apresenta-se a fundamentação teórica necessária ao desenvolvimento do presente trabalho por meio da descrição das ferramentas utilizadas.

No Capítulo 4, apresenta-se o conjunto de dados e os passos metodológicos utilizados.

No Capítulo 5, exibem-se os resultados obtidos com o método proposto.

No Capítulo 6, são apresentadas as conclusões do trabalho e as sugestões de trabalhos futuros.

No Apêndice 1 encontra-se uma cópia do artigo publicado no congresso Geoinfo 2020 (*XXI Brazilian Symposium on GeoInformatics*)

No Apêndice 2 encontra-se uma cópia do artigo publicado no periódico IEEE (*Institute of Electrical and Electronics Engineers*) Access.

---

## 2 REVISÃO BIBLIOGRÁFICA

Imagens de sensoriamento remoto registram a radiação eletromagnética da superfície da Terra em um determinado momento. Atualmente, os dados de imagem de satélite normalmente utilizados são LANDSAT, Spot, NOAA, Quickbird, IKONOS, ASTER, etc. Os dados da imagem de sensoriamento remoto possuem diferentes resoluções espaciais, espectrais, de radiação e de tempo. Por exemplo, em relação ao satélite meteorológico NOAA, a sua resolução espacial é baixa, no entanto a resolução temporal é alta, por isso é frequentemente utilizado na cobertura do uso / ocupação do solo e na alteração do sensoriamento remoto do nível do mar na escala continental ou global. Por outro lado, a série de imagens de satélite *Landsat*, com resolução espacial mínima de 30m, traz vantagens econômicas e técnicas para a investigação e monitoramento abrangente no ambiente com recursos de média escala e benefícios ecológicos (YAO *et al.*, 2017).

O avanço tecnológico permitiu que uma grande quantidade de imagens de sensoriamento remoto esteja disponível. Esses dados podem ser utilizados para inúmeras aplicações como mapear o uso da terra, planejamento urbano, vigilância, etc. No entanto, realizar uma análise manual muitas vezes requer tempo e muita mão de obra especializada. Portanto, a necessidade de extração automática de informações nessas imagens incentivou o desenvolvimento e criação de métodos de classificação utilizando aprendizado de máquina (SEVO e AVRAMOVIC, 2016). Uma das metodologias que tem sido empregada para esse propósito são as redes de aprendizado profundo.

A seguir é feita uma breve revisão de trabalhos encontrados nas bases *IEEE Xplore*, *Web of Science* e do INPE que realizam a classificação de imagens de sensoriamento na região Amazônica.

Em Barbosa *et al.*, (2015), Cassol *et al.*, 2020 e Lee, Cardille e Coe, (2020), os autores realizaram a classificação das imagens em diferentes regiões do Estado de Mato Grosso, Brasil. As classes utilizadas pelos autores foram agricultura, floresta e pasto. Além disso, Barbosa *et al.*, (2015), utilizou também a classe de não-floresta.

Barbosa *et al.* (2015) desenvolveram uma metodologia para classificação do uso da terra por meio da utilização de índices de vegetação EVI disponíveis nas séries temporais MODIS e imagens LANDSAT-5/TM. Segundo o autor, nesta região existem extensas áreas com concentração de agricultura, pasto e floresta. O conjunto de dados de treinamento e testes foi

construído a partir de informações do projeto TerraClass2008. Os autores utilizaram cinco classificadores: Árvore de Decisão (DT), Bayes Ingênuo (NB), Vizinho Mais Próximo (NN), Máquina de Vetores de Suporte (SVM), Floresta de Caminhos Ótimos (OPF). Em relação a região correspondente a imagem 227/68 (Imagem LANDSAT-5/TM), os resultados mostraram que o classificador SVM obteve uma maior taxa de acerto, conforme mostrado na Tabela 1. Para esta região, foi obtida uma acurácia média de 86,04%.

**Tabela 1 – Valores médios de acurácia para a região 227/68 (Barbosa *et al.*, 2015)**

Ano	Classificador	Agricultura	Floresta	Não-Floresta	Pasto	Média
2008	NB	80,46	87,36	47,59	86,10	75,38
	NN	77,70	79,77	62,18	84,14	75,95
	DT	75,06	79,43	60,58	83,79	74,72
	OPF	74,83	78,85	62,41	83,79	74,97
	SVM	<b>87,47</b>	<b>93,45</b>	<b>72,07</b>	<b>91,15</b>	<b>86,04</b>
2010	NB	82,22	86,91	49,60	83,73	75,62
	NN	75,32	79,13	61,19	79,52	73,79
	DT	74,76	79,60	59,52	78,02	72,98
	OPF	73,25	78,10	61,35	78,57	72,82
	SVM	84,68	90,64	65,56	91,67	83,14

Cassol *et al.*, (2020) utilizaram imagens LANDSAT-8/OLI e PROBA-V C1 para Uso e Cobertura da Terra (LULC - *Land Use and Land Cover*). Para as imagens LANDSAT-8/OLI as bandas utilizadas foram: verde, vermelho e infravermelho próximo. As bandas das imagens PROBA-V C1 foram: azul, vermelho e infravermelho próximo. O padrão-ouro utilizado foram os resultados do projeto MapBiomass. Os autores apresentaram um novo método para produzir mapas de LULC. A ideia inovadora envolve a análise do comportamento temporal (anual) das proporções de vegetação, solo e sombra das classes LULC para classificá-las de acordo com seus picos máximos e desvios padrão, que estão intimamente relacionados aos comportamentos biofísicos e fenológicos das classes de floresta, agricultura e pasto. O classificador utilizado pelos autores foi o *random forest*, e a acurácia obtida foi de 88%.

Lee, Cardille e Coe, (2020) avaliaram uma região do estado de Mato Grosso que incluiu os municípios de Alto Boa Vista, Querência, Ribeirão Cascalheira, e Canarana. As imagens utilizadas foram LANDSAT-5/TM dos anos de 1986 a 2000. As bandas utilizadas foram as bandas quatro, cinco e sete. Os dados do projeto GlobCover de 2009 foram utilizados como padrão ouro. A Figura 1 apresenta a área de estudo avaliada pelos autores. Os autores utilizaram o classificador não supervisionado chamado BULC-U. A Figura 2 apresenta uma imagem do ano 2000 avaliada pelos autores. A acurácia obtida para essa imagem foi de 80%.

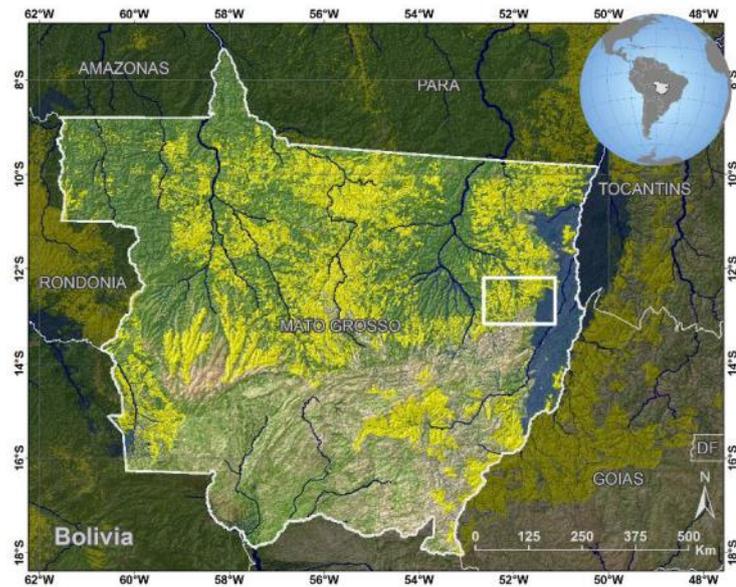


Figura 1 – Região avaliada por Lee, Cardille e Coe, (2020). Fonte: Lee, Cardille e Coe, (2020)

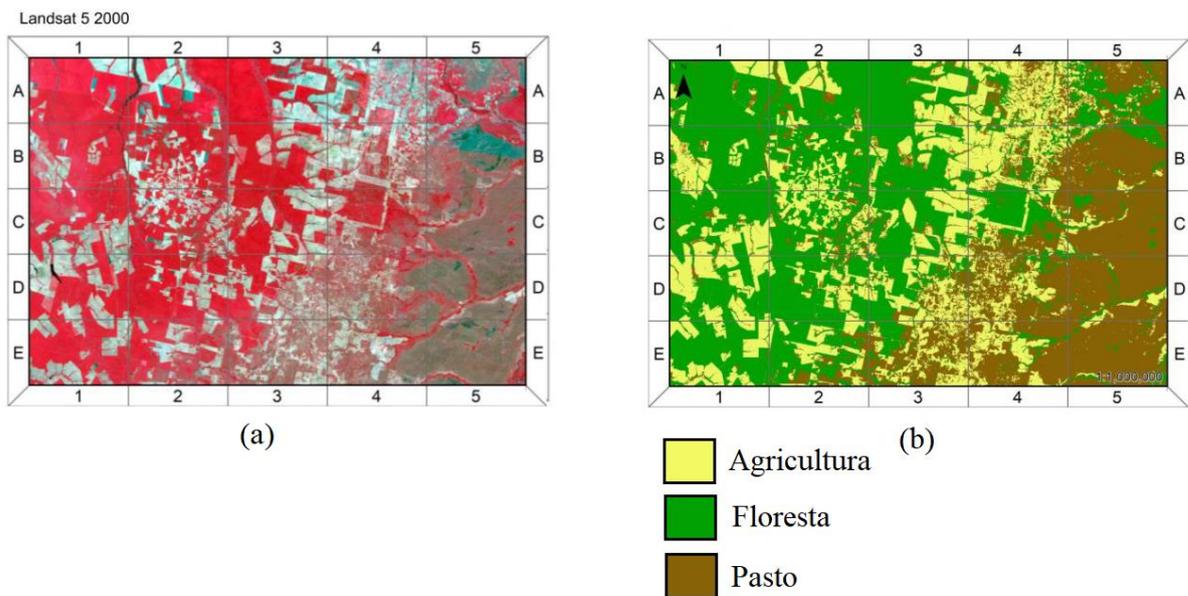
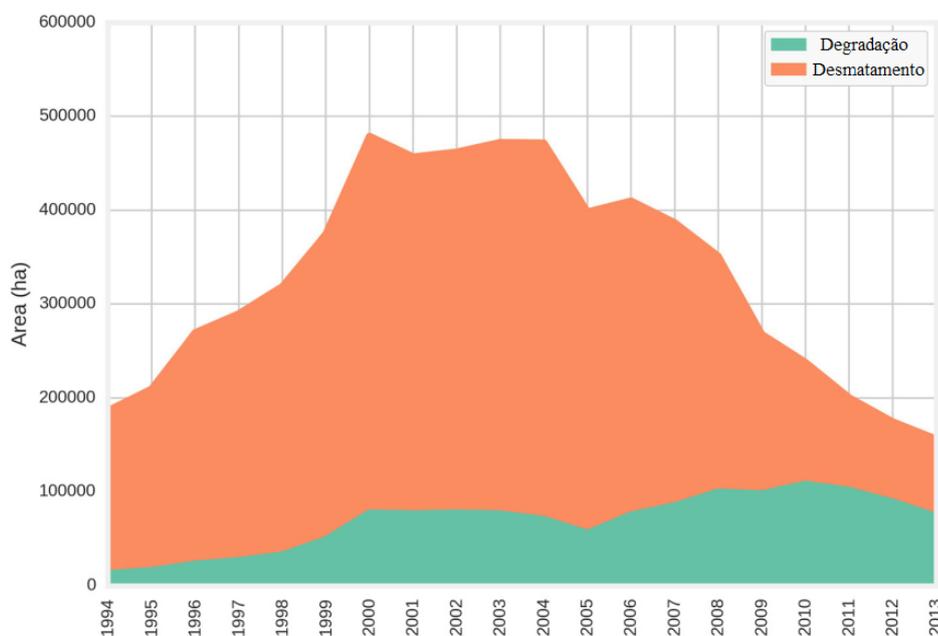


Figura 2 – Exemplo de uma imagem avaliada por Lee, Cardille e Coe, (2020). Em (a) temos a imagem LANDSAT-5/TM com uma composição RGB utilizando as bandas 4,5,7. Em (b) temos a imagem classificada. A acurácia obtida foi de 80%. Fonte: Lee, Cardille e Coe, (2020)

Bullock, Woodcock e Olofsson, (2020) desenvolveram uma metodologia capaz de mapear e estimar a degradação, desmatamento e floresta. A região avaliada foi o estado de Rondônia. Foram utilizadas imagens LANDSAT-5/TM e LANDSAT-8/OLI dos anos de 1990 a 2013. Os dados de treinamento para classificação foram coletados na interface do navegador Google Earth Engine. Os autores utilizaram *random forest* para classificar regiões. A acurácia obtida foi de 85%. Ao fim do estudo realizado por Bullock, Woodcock e Olofsson, (2020), os autores concluíram que durante o período de 24 anos de seus estudos, houve claros padrões

temporais de desmatamento e degradação. As taxas anuais de desmatamento e degradação aumentaram rapidamente no final dos anos 90 e início dos anos 2000. Após 2004, no entanto, o desmatamento caiu rapidamente e a degradação subiu para seu nível mais alto no período do estudo. A Figura 3 apresenta um gráfico que mostra a degradação e o desmatamento entre os anos de 1994 e 2013.

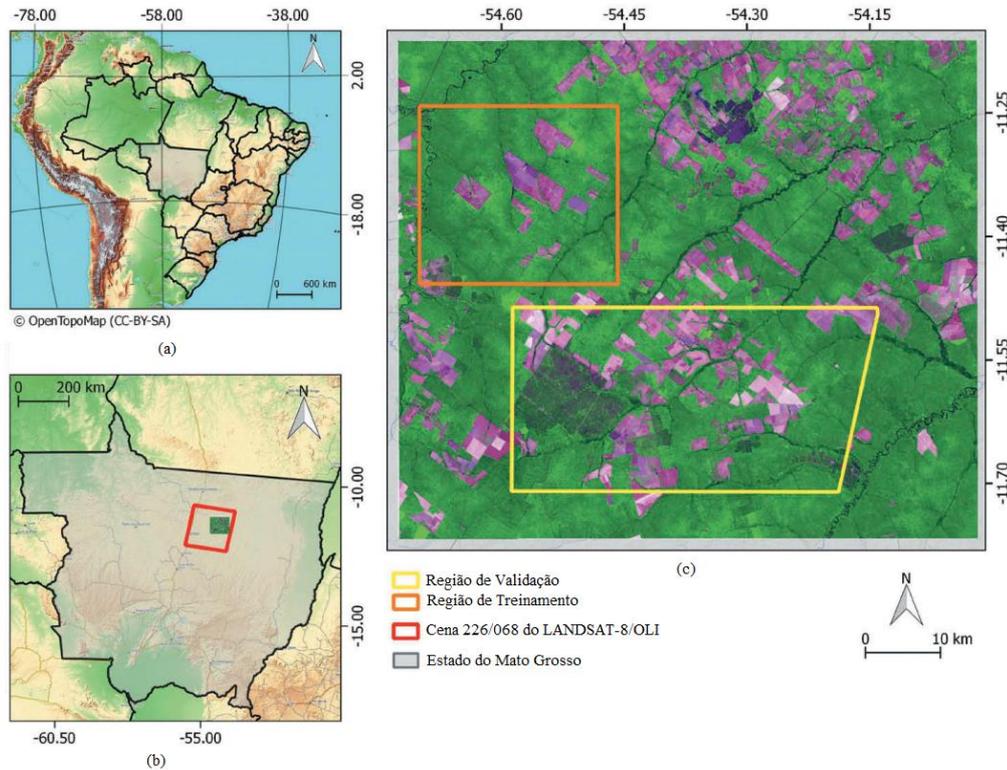


**Figura 3 – Degradação e Desmatamento calculado por Bullock, Woodcock e Olofsson, (2020) para os anos de 1994 a 2013. Fonte: Bullock, Woodcock e Olofsson, (2020)**

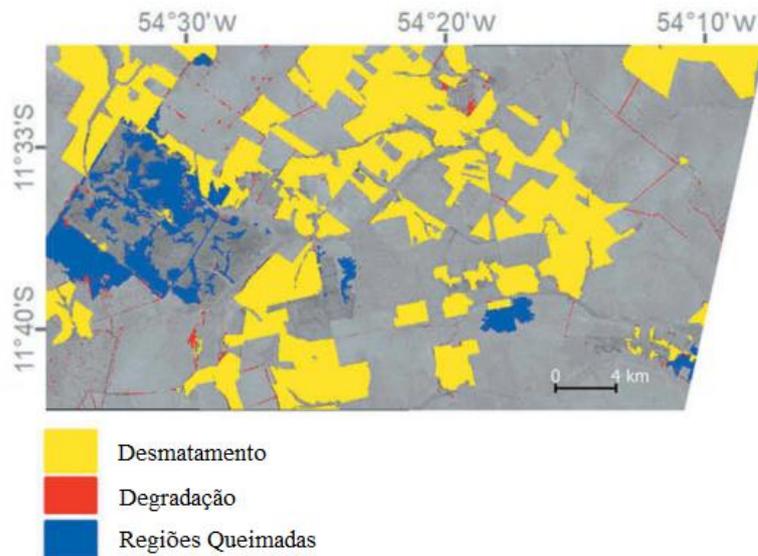
Em Lima *et al.*, (2019), Nascimento *et al.*, (2020) e Shimabukuro *et al.*, (2019) descritos a seguir, os autores empregaram uma segmentação por limiar para a classificação da imagem de sensoriamento remoto.

Em Shimabukuro *et al.*, (2019), os autores avaliaram uma região do estado de Mato Grosso com o objetivo de classificar o desmatamento, áreas de degradação e regiões de queimada. A região de estudo corresponde a uma área da cena 226/68 da LANDSAT. A Figura 4 apresenta a região avaliada pelos autores. Foram utilizadas 21 imagens do LANDSAT-5/TM de 2005 a 2011, 15 imagens do LANDSAT-8/OLI de 2013 a 2017 e duas imagens RapidEye do ano de 2013. As bandas utilizadas das imagens LANDSAT-5/TM foram as bandas 5 (Infravermelho de ondas curtas), 4 (Infravermelho) e 3 (vermelho). Para as imagens do LANDSAT-8/OLI, os autores utilizaram as bandas 6 (Infravermelho de ondas curtas 1), 5 (Infravermelho próximo) e 4 (Vermelho). Os autores selecionaram imagens do período da estação seca (maio a setembro). Segundo os autores, esse período possui uma alta probabilidade

de ter imagens sem nuvens e as atividades de desmatamento e degradação florestal ocorrem durante este período. A Figura 5 apresenta uma imagem de 2013 classificada. A acurácia obtida pelos autores foi de 94,93%.

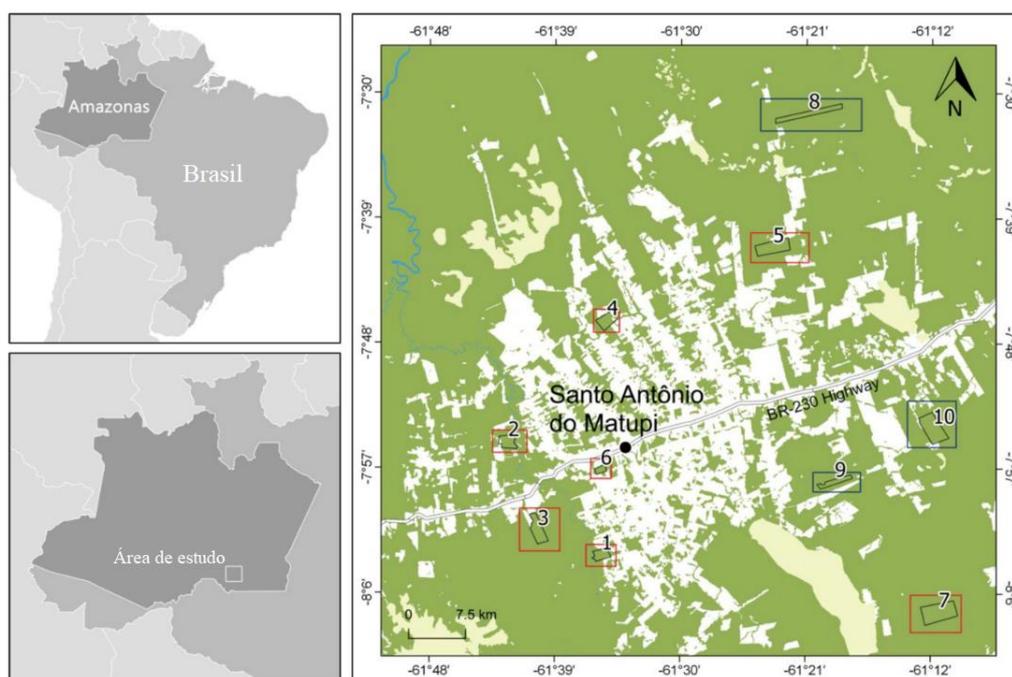


**Figura 4 – Região avaliada por Shimabukuro *et al.*, (2019). Em (a) é mostrado as unidades da Federação do Brasil e em destaque o estado de Mato Grosso. Em (b) temos a localização da área de estudo no estado de Mato Grosso. Em (c), tem uma imagem LANDSAT-8/OLI numa composição RGB utilizando as bandas 6, 5 e 4. A região laranja corresponde a região utilizada para o treinamento e a região em amarelo corresponde a região de validação. Fonte: Shimabukuro *et al.*, (2019).**



**Figura 5 – Imagem de 2013 classificada. A acurácia obtida foi de 94,93%. Fonte: Shimabukuro *et al.*, (2019).**

Em Lima *et al.*, (2019), os autores utilizaram imagens LANDSAT-8/OLI e Sentinel-2 para classificar o corte seletivo em uma região do sul do Estado do Amazonas, perto da vila de Santo Antônio do Matupi. As bandas utilizadas foram: azul, verde, vermelho, infravermelho próximo, infravermelho de ondas curtas 1 (SWIR-1) e infravermelho de ondas curtas 2. A região avaliada pelos autores possui dez locais de manejo florestal sustentável. Todas as áreas de manejo florestal compreendem florestas privadas licenciadas nos anos 2016 e 2017. No Brasil, os estados têm a responsabilidade de controlar o processo de licenciamento e emitir licenças de exploração madeireira. No Amazonas, o Instituto de Proteção Ambiental do Estado do Amazonas é responsável pela emissão dessas licenças. A Figura 6 apresenta as áreas de manejo florestal (numeradas de 1 a 10). Os autores avaliaram as áreas de 1 a 7, pois são nessas regiões onde ocorre a atividade de extração de madeira. A acurácia obtida pelos autores foi de 95,7% para as imagens LANDSAT-8/OLI e 96,7% para as imagens Sentinel-2.



**Figura 6 – Área de estudo de Lima *et al.*, (2019). Fonte: Lima *et al.*, (2019)**

Em Nascimento *et al.*, (2020), os autores realizaram a classificação das imagens na região das minas de Carajás, localizadas no sul do estado do Pará. A Figura 7 apresenta a área de estudo dos autores. As classes definidas pelos autores foram cerrado, floresta, água, área de mineração e áreas de reflorestamento. Os autores utilizaram imagens ikonos, geoeye e world view dos anos de 2011 a 2015. O padrão ouro foi gerado pelos próprios autores através de interpretação visual. A acurácia obtida pelos autores foi de aproximadamente 96%.

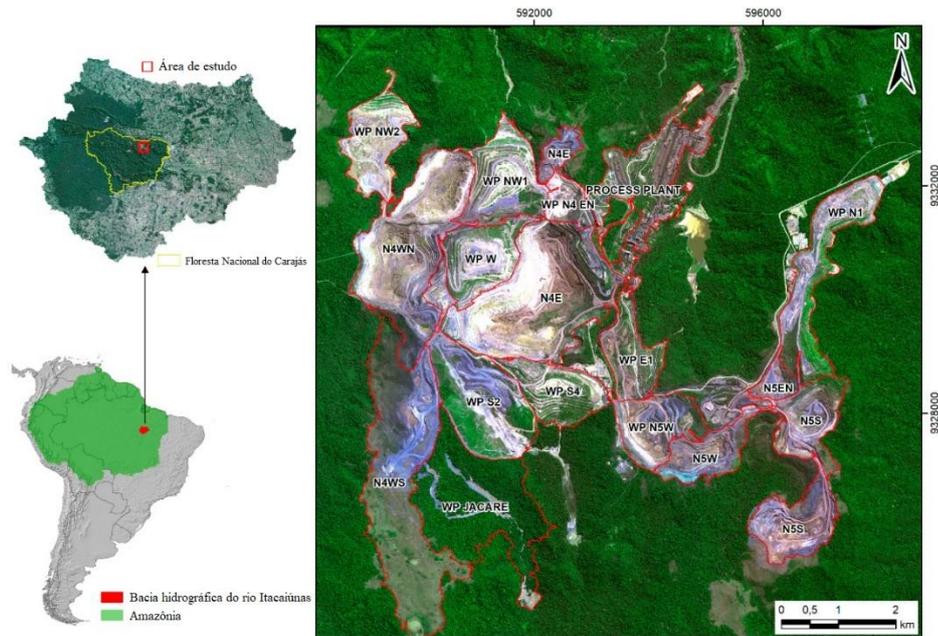


Figura 7 – Área de estudo de Nascimento *et al.*, (2020). Fonte: Nascimento *et al.*, (2020)

Em Maretto *et al.*, (2020), os autores utilizaram a CNN para realizar a classificação das áreas desmatadas em uma região do sul do estado do Pará. Os autores utilizaram imagens Landsat-8/OLI e dados PRODES como padrão ouro. O conjunto de dados de treinamento é composto por imagens de cinco anos, de 2013 a 2017. O conjunto de dados de teste foi composto por imagens de 2018. O projeto PRODES classifica as regiões em quatro classes principais, não-floresta, floresta, hidrografia e desmatamento. Para simplificar, uma vez que a hidrografia e a não-floresta são invariantes no tempo, os autores agruparam com a classe floresta em uma única classe, denominada não desmatamento. O classificador utilizado pelos autores foi a CNN U-NET. A Figura 8 apresenta a imagem classificada da região avaliada pelos autores. A acurácia obtida foi aproximadamente 95%.

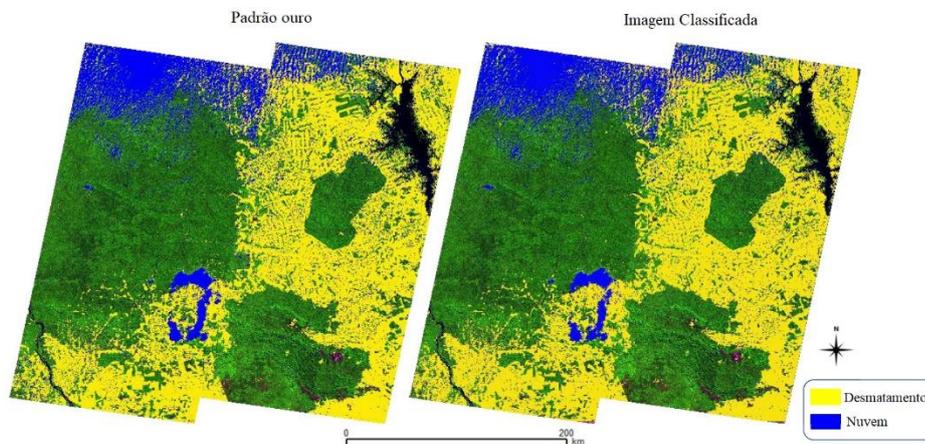
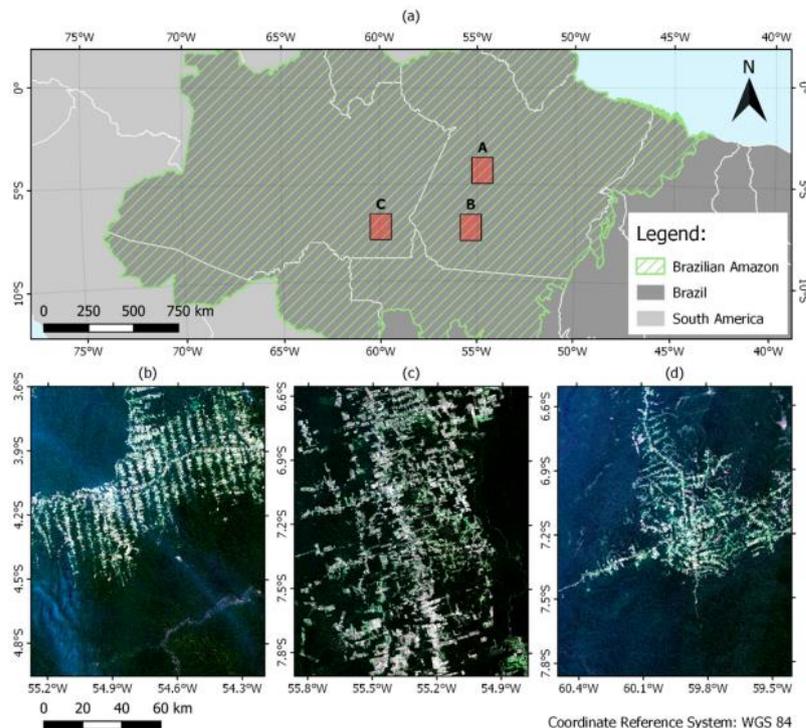


Figura 8 – Resultados da classificação da região de estudo de Maretto *et al.*, (2020). A acurácia obtida foi de 95% aproximadamente. Fonte: Maretto *et al.*, (2020)

Bem *et al.*, (2020) mapearam o desmatamento entre imagens com aproximadamente um ano de diferença, especificamente entre 2017 e 2018 e entre 2018 e 2019. Para isso, os autores utilizaram três arquiteturas da CNN disponíveis na literatura - SharpMask, U-Net e ResUnet - e depois compararam o desempenho com dois algoritmos de aprendizado de máquina clássico: RF e perceptron de multicamadas (MLP). Nos experimentos realizados, foram utilizadas as bandas 1 a 7 de imagens LANDSAT-8/OLI de três regiões da Amazônia Brasileira. Segundo os autores, essas regiões abrangem os principais centros de desmatamento que se desenvolveram ao longo da Transamazônica (BR-230) e rodovias “Cuiabá – Santarém”. A Figura 9 apresenta a localização da área de estudo. O conjunto de treinamento utilizou duas imagens (Sites A e B) e o conjunto de validação utilizou a imagem restante (Site C), conforme apresentado na Tabela 2. As imagens obtidas foram da estação seca para minimizar a cobertura de nuvens e reduzir o ruído. Os dados de referência utilizados foram os dados de desmatamento ocorridos apenas nos anos de 2017, 2018 e 2019 extraídos do projeto PRODES do ano de 2017, 2018 e 2019. Os melhores resultados foram obtidos com o modelo ResUnet, nesse caso a acurácia e o F1-Score foram de 99,93% e 94,65%, respectivamente.

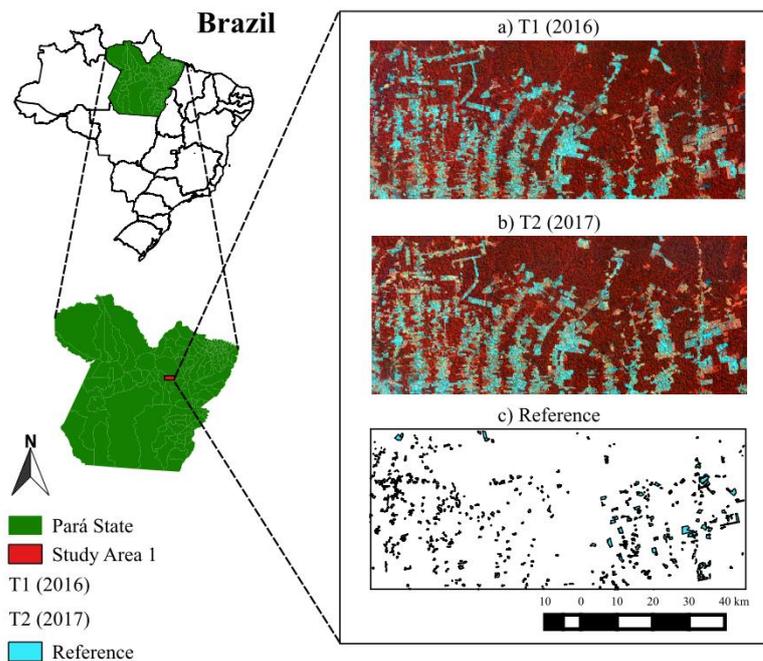


**Figura 9 – Localização das regiões de estudo na região amazônica com os locais de treinamento (b, c) A e B e (d) local de teste C como imagens compostas LANDSAT-8/OLI em cores verdadeiras, tiradas de junho e julho de 2018.. Fonte: Bem *et al.* (2020)**

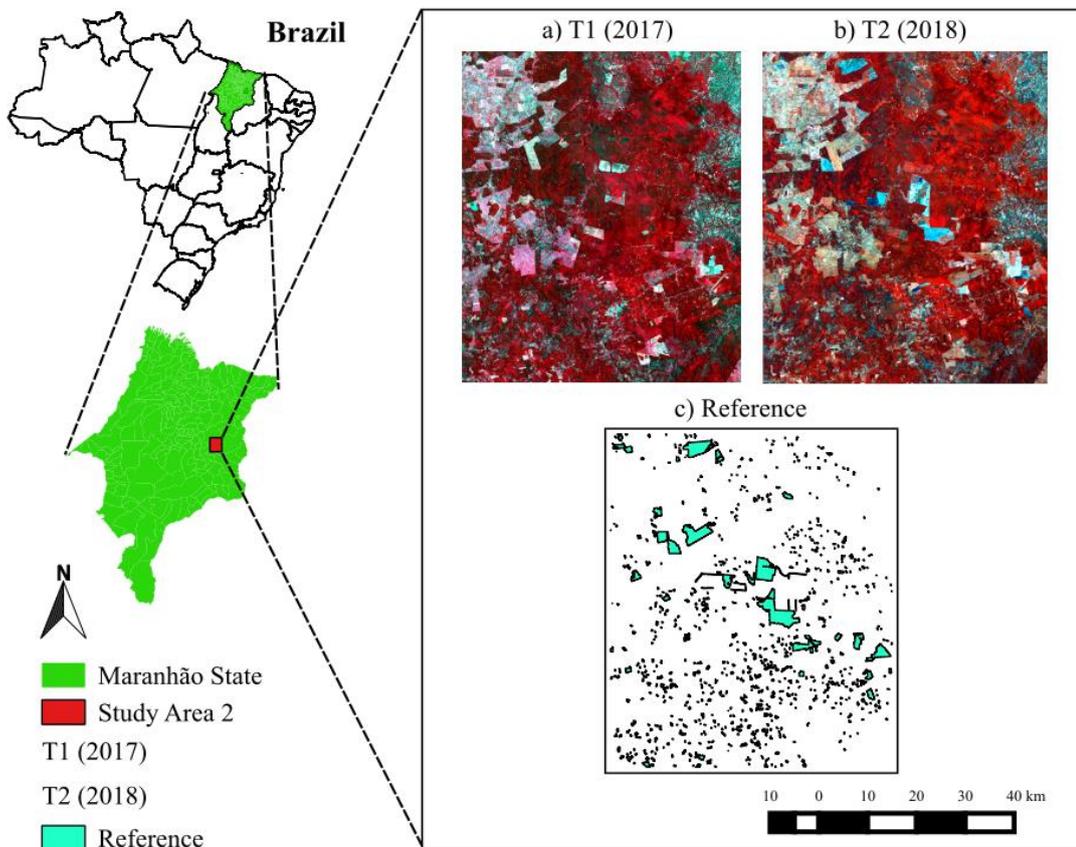
**Tabela 2 – Datas de aquisição para cada local e cenas correspondentes do LANDSAT-8/OLI Fonte: (Bem *et al.*(2020))**

LOCAL	CENA LANDSAT-8/OLI	DATA DA IMAGEM		
		2017	2018	2019
A	227/63	18/07	21/07	24/07
B	227/65	18/07	21/07	24/07
C	230/65	21/06	24/06	13/07

Adarme *et al.*, (2020) avaliaram estratégias baseadas em aprendizado profundo para detecção automática de desmatamento. As áreas de estudo foram duas regiões com dois padrões de desmatamento: os biomas Amazônia e Cerrado no Brasil. Nos experimentos realizados, foram utilizadas as bandas 1 a 7 de imagens LANDSAT-8/OLI de duas regiões da Amazônia Brasileira. A primeira área de estudo corresponde a uma região do bioma Amazônia, mais especificamente localizada no estado do Pará, Brasil, conforme mostrado na Figura 10. O Estado do Pará compreende 26% da Amazônia brasileira, e a maior parte é coberta por densa floresta tropical. Esta área tem enfrentado um processo contínuo de degradação, conforme indicado pelos relatórios do PRODES e DETER. Nesta área de estudo foram avaliadas duas imagens LANDSAT-8/OLI correspondente aos de 2016 e 2017. Os dados de referências utilizados nesse experimento referem-se ao desmatamento que ocorreu entre agosto de 2016 e julho de 2017 extraídos do projeto PRODES. A segunda área de estudo pertence ao bioma Cerrado brasileiro, localizado no estado do Maranhão, Brasil. A Figura 11 ilustra essa área de estudo. O estado do Maranhão está em uma área de transição entre três biomas diferentes: Cerrado (64%), Amazônia (35%) e Caatinga (1%), com predominância de formações de savana no Cerrado. Para esse caso, foram avaliadas duas imagens LANDSAT-8/OLI do ano de 2017 e 2018. Com respeito aos dados de referências, foram utilizadas informações de desmatamento ocorridos apenas nos anos de 2017 e 2018. Os autores obtiveram os melhores resultados utilizando redes neurais Siamesas. Essas redes podem ser consideradas como uma extensão de uma CNN convencional. Os autores obtiveram uma acurácia de 95% e de 63% em termos de F1-Score na Amazônia, e de 97% em termos de acurácia e 78% em termos de do F1-Score para o Cerrado, mostrando que os resultados para o banco de dados do Cerrado alcançaram percentuais mais altos que o da Amazônia. A razão para isso, segundo os autores está no padrão de desmatamento no bioma Cerrado que nesse caso vegetação é completamente removida e a maior parte do solo é exposta, ao contrário da região Amazônica, onde é comum haver restos de vegetação no processo de desmatamento, o que dificulta a detecção.



**Figura 10 – Região do bioma Amazônia brasileira, localizada no estado do Pará, Brasil. Em (a,b) são apresentadas uma composição colorida (Infravermelho próximo, verde, azul) da área de estudo em duas datas diferentes T1 e T2; (c) Referência do processo de desmatamento de 2016 a 2017. Fonte: Adarme *et al.* (2020)**



**Figura 11 – Região do bioma Cerrado Brasileiro, localizada no estado do Maranhão, Brasil. Em (a,b) são apresentadas uma composição colorida (Infravermelho próximo, verde, azul) da área de estudo em duas datas diferentes T1 e T2; (c) Referência do processo de desmatamento de 2017 a 2018. Fonte: Adarme *et al.* (2020)**

## CONSIDERAÇÕES FINAIS

A classificação de imagens por sensoriamento remoto desempenha um papel importante na tecnologia de observação da Terra, tendo sido amplamente explorada nos campos militar e civil. No entanto, devido às características dos dados de sensoriamento remoto, como alta dimensionalidade e poucas amostras rotuladas disponíveis, a classificação de imagens de sensoriamento remoto enfrenta grandes desafios científicos e práticos. Nos últimos anos, à medida que surgem novas técnicas de aprendizado profundo, abordagens para a classificação de imagens de sensoriamento remoto com aprendizado profundo alcançaram avanços significativos, oferecendo novas oportunidades para a pesquisa e desenvolvimento da classificação de imagens de sensoriamento remoto.

Li *et al.*, (2018) realizaram uma pesquisa bibliográfica onde apresentaram alguns modelos típicos de aprendizado profundo que podem ser utilizados na classificação de imagens de sensoriamento remoto. Os modelos típicos de aprendizado profundo abordados foram *Convolutional neural networks* (CNN), *Stacked autoencoders* (SAE), *Deep belief networks* (DBF). Segundo os autores, o desempenho das técnicas de classificação de imagens de sensoriamento remoto baseado em aprendizado profundo mostrou boa eficácia na solução de problemas do mundo real. No entanto, Li *et al.*, (2018), acrescentam que ainda existem questões importantes em aberto como pequenas quantidades de dados rotulados e estrutura de rede adequada para classificação de imagens de sensoriamento remoto. A quantidade de imagens de sensoriamento remoto rotuladas ainda é muito pequena, isso restringe as abordagens de classificação de imagem. Como construir uma rede eficiente e treiná-la com um pequeno número de amostras de treinamento é algo desafiador e interessante. Investigar novos modelos que podem explorar amostras não rotuladas é claramente uma direção desejável para trabalhos futuros. Em relação ao tópico de estrutura de rede adequada, os autores mencionam que, em geral os modelos de rede existentes possuem um bom desempenho para a execução de uma tarefa dedicada e em geral a região de análise é pequena e limitada e a exploração de estruturas de rede apropriadas para um determinado problema de classificação de imagens de Sensoriamento Remoto ainda é um tópico aberto.

Os trabalhos apresentados nesta revisão mostraram que as técnicas de aprendizagem de máquinas têm grande potencial para classificar imagens de sensoriamento remoto e para realizar segmentação automática nos mais variados usos de cobertura vegetal na região amazônica. Em Adarme *et al.*, (2020), Bem *et al.*, (2020) e Maretto *et al.*, (2020), os autores utilizaram técnicas de aprendizado profundo para classificar as regiões de desmatamento na Amazônia. Os

resultados alcançados foram bastante satisfatórios. No entanto, diferentemente da proposta desta tese, os trabalhos encontrados na literatura não apresentam um modelo em aprendizado profundo para a classificação de uso do solo para a região amazônica brasileira. Portanto, isso abre novas possibilidades para pesquisas neste campo. Neste trabalho, disponibilizamos o conjunto de dados de imagens utilizado, que cobre diferentes áreas da Amazônia. Estas áreas abrangem os estados brasileiros do Amazonas, Mato Grosso, Pará e Rondônia.

No Quadro 1, ao final desta seção, é apresentado um resumo dos artigos citados, esclarecendo suas características e os resultados alcançados em cada um.

Quadro 1 – Quadro Comparativo da Revisão Bibliográfica

Ano/autor	Título	Objetivos	Materiais	Classificador	Medida de Desempenho e Resultado		
2015 / David Pereira Barbosa Alexandre Noma Thales Sehn Korting Leila Maria Garcia Fonseca	<i>Um Estudo Experimental com Classificadores baseados em Regiões e Perfis EVI</i>	Classificar cobertura/uso da terra nas seguintes classes: Agricultura, Floresta, Não-Floresta, Pasto	<b>Banco de Dados:</b> Imagens LANDSAT-5/TM, Imagens MODIS, TerraClass  <b>Padrão ouro:</b> TerraClass2008, TerraClass2010 (imagem segmentada pelo algoritmo Baaz)	DT NB NN SVM OPF	Acurácia		
					<b>Ano</b>	<b>Classificador</b>	<b>Acurácia Média</b>
					2008	NB	75,38
						NN	75,95
						DT	74,72
						OPF	74,97
						SVM	<b>86,04</b>
					2010	NB	75,62
						NN	73,79
						DT	72,98
OPF	72,82						
SVM	83,14						
2020 / Henrique Luis Godinho Cassol, Egidio Arai, Edson Eyji Sano, Andeise Cerqueira Dutra, Tânia Beatriz Hoffmann e Yosio Edemir Shimabukuro	<i>Maximum Fraction Images Derived from Year-Based Project for On-Board Autonomy-Vegetation (PROBA-V) Data for the Rapid Assessment of Land Use and Land Cover Areas in Mato Grosso State, Brazil</i>	Classificar imagens de sensoriamento remoto  <b>Classes:</b> Agricultura, Floresta e pasto.	<b>Banco de Imagens:</b> LANDSAT-8/OLI e PROBA-V C1  <b>Padrão ouro:</b> MapBiome	<i>Random forest (RF)</i>	Acurácia: 88%		
2020 / Jacky Lee, Jeffrey A. Cardille, e Michael T. Coe	<i>Agricultural Expansion in Mato Grosso from 1986–2000: A Bayesian Time Series Approach to Tracking Past Land Cover Change</i>	Classificar imagens de sensoriamento remoto  <b>Classes:</b> Agricultura, Floresta e pasto.	<b>Banco de Imagens:</b> LANDSAT-5/TM  <b>Padrão ouro:</b> Gloobcover 2009	Classificador não supervisionado: BULC-C	Acurácia: 80%		

Ano/autor	Título	Objetivos	Materiais	Classificador	Medida de Desempenho e Resultado
2018 / Eric L. Bullock, Curtis E. Woodcock, Pontus Olofsson	<i>Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis</i>	<p>Mapear e estimar a degradação, floresta e desmatamento</p> <p><b>Classes:</b> degradação, floresta e desmatamento</p>	<p><b>Banco de Imagens:</b> LANDSAT-5/TM e LANDSAT-8/OLI</p> <p><b>Padrão ouro:</b> google earth engine</p>	<i>Random forest</i>	Acurácia: 85%
2019 / Thaís Almeida Lima, René Beuchle, Andreas Langner, Rosana Cristina Grecchi, Verena C. Griess e Frédéric Achard	<i>Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian Amazon</i>	<p>Comparar a classificação de regiões de corte seletivo com imagens LANDSAT-8/OLI e Sentinel</p>	<p><b>Banco de Imagens:</b> LANDSAT-8/OLI e Sentinel-2</p>	Classificação por limiar	<p>Acurácia LANDSAT-8/OLI: 95,7% Sentinel-2: 96,7%</p>
2020 / Filipe Silveira Nascimento, Markus Gastauer, Pedro Walfir M. Souza-Filho, Wilson R. Nascimento Jr., Diogo C. Santos e Marlene F. Costa	<i>Land Cover Changes in Open- Cast Mining Complexes Based on High- Resolution Remote Sensing Data</i>	<p>Classificar imagens na região de minas de Carajás, sul do Pará</p> <p><b>Classes:</b> cerrado, floresta, água, área de mineração e áreas de reflorestamento</p>	<p><b>Banco de Dados:</b> iknos, geoeeye e world view</p>	Classificação por limiar	Acurácia: 96%

Ano/autor	Título	Objetivos	Materiais	Classificador	Medida de Desempenho e Resultado
2019 / Yosio Edemir Shimabukuro, Egidio Arai, Valdete Duarte, Anderson Jorge, Erone Ghizoni dos Santos, Kaio Allan Cruz Gasparini e Andeise Cerqueira Dutra	<i>Monitoring deforestation and forest degradation using multi-temporal fraction images derived from Landsat sensor data in the Brazilian Amazon</i>	Classificar imagens de sensoriamento remoto <b>classes:</b> desmatamento, degradação e regiões de queimada.	<b>Banco de Imagens:</b> LANDSAT-5/TM, LANDSET-8/OLI e Rapideye	Classificação por limiar	Acurácia: 94,93%
2020 / Raian V. Mareto, Leila M. G. Fonseca, Nathan Jacobs, Thales S. Körting, Hugo N. Bendini, E Leandro L. Parente	<i>Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest</i>	Classificar imagens de sensoriamento remoto <b>classes:</b> desmatamento e não-desmatamento.	<b>Banco de Imagens:</b> LANDSAT-8/OLI <b>Padrão ouro:</b> PRODES	CNN	Acurácia: 95%
2020 / Bem et. al (2020)	<i>Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks</i>	Mapear o desmatamento entre imagens com aproximadamente um ano de diferença, entre 2017 e 2018 e entre 2018 e 2019 <b>Classes:</b> Desmatamento e Não-Desmatamento	Banco de imagens: Imagens LANDSAT-8/OLI <b>Padrão ouro:</b> PRODES	CNN	Acurácia: 99,93%

Ano/autor	Título	Objetivos	Materiais	Classificador	Medida de Desempenho e Resultado
2020 / Adarme <i>et. al</i>	<i>Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery</i>	Avaliar estratégias baseadas em aprendizado profundo para detecção automática de desmatamento nos biomas Amazônico e Cerrado. <b>Classes:</b> Desmatamento e Não-Desmatamento	Banco de imagens: Imagens LANDSAT-8/OLI	CNN	Acurácia (Amazônia): 95% Acurácia (Cerrado): 97% F1-Score (Amazônia): 63% F1-Score (Cerrado): 78%

### 3 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo são apresentados os conceitos essenciais para uma melhor compreensão desta tese. Os assuntos apresentados versam sobre sensoriamento remoto, imagens LANDSAT-8/OLI. Por fim, será feita uma revisão conceitual das técnicas de aprendizado profundo, CNN, métodos de otimização e Transferência de Conhecimento.

#### 3.1 Sensoriamento Remoto

Sensoriamento remoto é uma técnica que consiste na aquisição de informações de um objeto ou fenômeno sem o contato físico com ele. Esses dados podem ser utilizados para a extração de informações, monitoramento dinâmico, mapeamento temático, classificação de imagens etc. Essa técnica desempenha um papel vital para interpretar e identificar objetos no solo por meio de classificação de imagens, atribuindo à cada pixel ou região da imagem uma determinada categoria (BHARATKAR; PATEL, 2013).

Segundo Meneses e Almeida (2012), sensoriamento remoto visa a obtenção de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres.

A radiação eletromagnética possui uma dualidade de comportamento: onda e energia. Segundo Meneses e Almeida (2012), a radiação eletromagnética que se propaga pelo espaço, como luz solar, é ao mesmo tempo uma forma de onda e uma forma de energia.

Na Figura 12, podemos observar um exemplo de obtenção de dados por meio de sensoriamento remoto. Nela, vemos que o Sol ilumina a superfície terrestre. A energia proveniente do Sol, refletida pela superfície em direção ao sensor, é captada e registrada em forma de imagem, com o posterior processamento das mesmas (FLORENZANO, 2011).

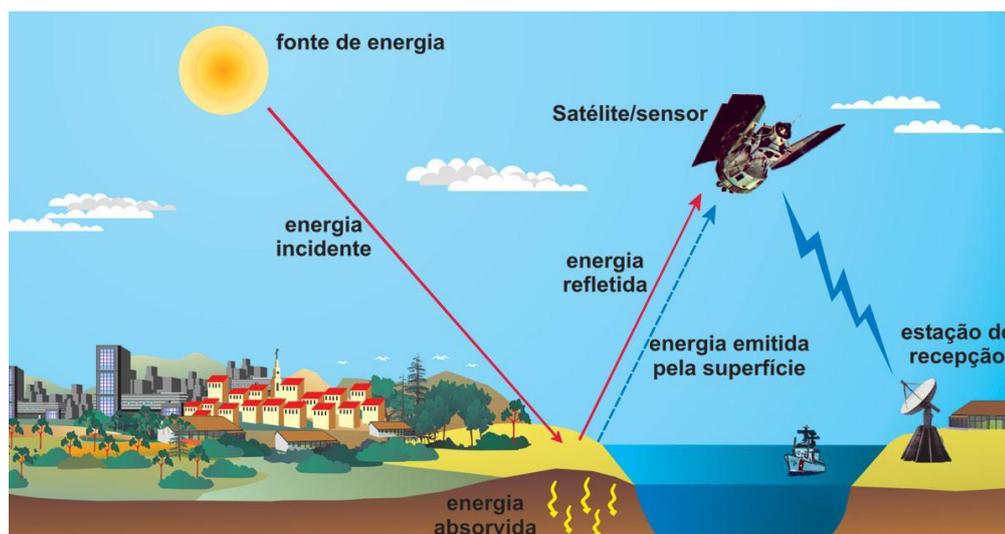


Figura 12 – Obtenção de imagem por sensoriamento remoto. Fonte: (FLORENZANO, 2011)

### 3.1.1 Fontes de Energia usadas em sensoriamento remoto

A obtenção de dados por sensoriamento remoto, como em qualquer outra atividade envolvendo equipamentos eletromagnéticos, requer o uso de energia. A energia com a qual operam os sensores remotos pode ser proveniente de uma fonte natural, como a luz do Sol ou o calor emitido pela superfície da Terra, ou pode ser de uma fonte artificial como, por exemplo, a energia do *flash* utilizado em uma máquina fotográfica ou o sinal produzido por um radar (FLORENZANO, 2011).

A energia utilizada em sensoriamento remoto para produção das imagens são ondas eletromagnéticas, que se propagam à velocidade da luz. A frequência das mesmas é expressa em unidades de *hertz* (Hz) e seus múltiplos, como quilohertz ( $1 \text{ kHz} = 10^3 \text{ Hz}$ ) e mega-hertz ( $1 \text{ MHz} = 10^6 \text{ Hz}$ ). O comprimento de onda ( $\lambda$ ) é medido em unidades de metro. A frequência de onda é o número de vezes que uma onda se repete por unidade de tempo, sendo diretamente proporcional à velocidade de propagação e inversamente proporcional ao comprimento da onda. Dessa maneira, como mostra a Figura 13, quanto menor for o comprimento da onda, maior será a frequência dela. O comprimento de onda é a distância entre dois picos de ondas sucessivas: quanto mais distantes esses picos, maior é o comprimento e, quanto menos distantes, menor será o comprimento de onda.

O espectro eletromagnético representa a distribuição da radiação eletromagnética, por regiões, segundo o comprimento de onda e a frequência. Na Figura 13, observa-se que o espectro eletromagnético abrange desde curtos comprimentos de onda, como os raios cósmicos e os raios gama ( $\gamma$ ), de alta frequência, até longos comprimentos de onda, como as ondas de

rádio e TV, de baixa frequência. Na região do espectro visível, o olho humano enxerga a energia (luz) eletromagnética, sendo capaz de distinguir as cores do violeta ao vermelho (FLORENZANO, 2011).

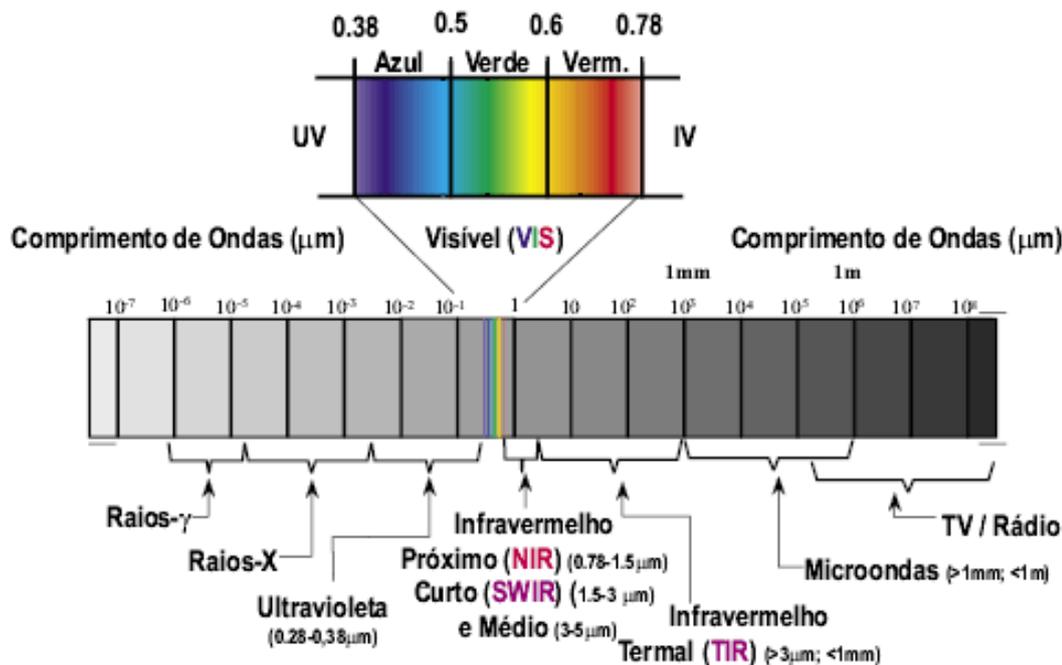


Figura 13 – Espectro eletromagnético. Fonte: (MUNDOGEO, 2004)

Segundo Meneses e Almeida (2012), o espectro eletromagnético foi arbitrariamente dividido pelo homem em intervalos de comprimentos de onda com base nos mecanismos físicos geradores da energia eletromagnética e nos mecanismos físicos de sua detecção. Segundo os autores, a nomenclatura de cada um dos intervalos foi realizada com base no uso que o homem encontrou para as suas aplicações. Meneses e Almeida (2012) acrescentam que é essencial saber os valores dos comprimentos de onda desses intervalos, as denominações que recebem, e quais comprimentos de ondas são possíveis de serem detectados por cada tipo de sensor.

A seguir é apresentado um resumo das características dos intervalos espectrais mostrados na Figura 13.

**Visível (0,45-0,76  $\mu\text{m}$ ):** Essa região do espectro apresenta a intensidade de fluxo radiante mais alta e é onde se verifica a melhor janela atmosférica, com a passagem de grande quantidade de radiação. É também responsável pela interação com os minerais, o que dá origem às suas cores, e com os pigmentos da vegetação. Porém, o problema dessa faixa espectral é o alto espalhamento da radiação solar incidente pelos gases atmosféricos. Isso pode provocar uma redução do contraste da refletância dos alvos terrestres (MENESES e ALMEIDA, 2012).

Infravermelho próximo (0,76 - 1,2  $\mu\text{m}$ ): Segundo Meneses e Almeida (2012), nessa região do espectro solar a atmosfera é bastante transparente, mas ocorrem algumas bandas de absorções, impedindo que todo o intervalo possa ser continuamente utilizado por sensoriamento remoto. Segundo os autores, esse intervalo espectral gera feições espectrais que são diagnósticas para identificar a natureza de vários tipos de rochas, principalmente as de composição mineral com metais de transição.

Infravermelho de ondas curtas (1,2 – 3,0  $\mu\text{m}$ ): É a região espectral geológica, pois é nesta faixa espectral que os vários minerais de alteração hidrotermal têm as suas diagnósticas feições de absorção. Nessa região os comprimentos de onda em 1,4  $\mu\text{m}$  e em 1,9  $\mu\text{m}$  são totalmente absorvidos pelas moléculas de vapor d'água da atmosfera, inviabilizando a utilização do sensoriamento remoto e, por consequência, de se determinar nos materiais terrestres a presença de água molecular nas suas estruturas (MENESES e ALMEIDA, 2012).

Infravermelho médio (3,0 - 5,0  $\mu\text{m}$ ): Nessa região o Sol e a Terra não emitem quantidades suficientes de energia que possam ser detectadas pelos sensores. Apenas alvos com elevadas temperaturas, como vulcões e incêndios, podem ser detectados, pois agem como fontes próprias de emissão de radiação. É uma região espectral pouco usada no sensoriamento remoto, à exceção de sensores meteorológicos ou atmosféricos (MENESES e ALMEIDA, 2012).

Infravermelho termal (5,0 – 1,0 mm): Segundo Meneses e Almeida (2012), essa região é conhecida como a região termal devido à radiação emitida pelos objetos terrestres em função das suas temperaturas de superfícies. Segundo os autores, a melhor janela atmosférica nessa região espectral para imageamento orbital é o intervalo de 8,0  $\mu\text{m}$  a 14,0  $\mu\text{m}$ , porque acima de 30 km, a camada de ozônio absorve toda a radiação além de 14  $\mu\text{m}$  emitida pela Terra.

Micro-ondas (3,0 - 100 cm) : Essa região, segundo Meneses e Almeida (2012) é de uso de sensores ativos (radar), que utilizam fontes artificiais para a geração da radiação eletromagnética. Segundo os autores, por causa do tamanho do comprimento de onda, o radar tem a habilidade de operar em condições atmosféricas adversas, com coberturas de nuvens ou chuvas, e pode operar tanto de dia como à noite. É utilizado para realizar o mapeamento geológico, pois segundo os autores a interação das micro-ondas com as rochas é controlada pelas texturas de relevo.

### 3.1.1 Resolução

Atualmente, o sensoriamento é constituído por uma razoável constelação de satélites que oferecem imagens para atender as necessidades de uma ampla demanda de usuários. Uma

forma de se abordar as potencialidades de um sensor é pelo dimensionamento de suas resoluções. A detecção ou identificação de um objeto nas imagens de sensoriamento remoto é determinada por quatro diferentes formas de medições: resolução espacial, espectral, radiométrica e temporal (MENESES e ALMEIDA, 2012).

#### 3.1.1.1 Resolução espacial

Segundo Meneses e Almeida (2012), resolução espacial determina o tamanho do menor objeto que pode ser identificado em uma imagem. Segundo os autores, um objeto somente pode ser detectado, quando o tamanho deste é, no mínimo, igual ou maior do que o tamanho do elemento de resolução no terreno. Ainda segundo Meneses e Almeida (2012), por exemplo, se uma casa tem 20 m x 20 m de tamanho, a resolução espacial da imagem deveria ser, no mínimo, de 20 metros para que essa casa possa ser identificada na imagem.

#### 3.1.1.2 Resolução espectral

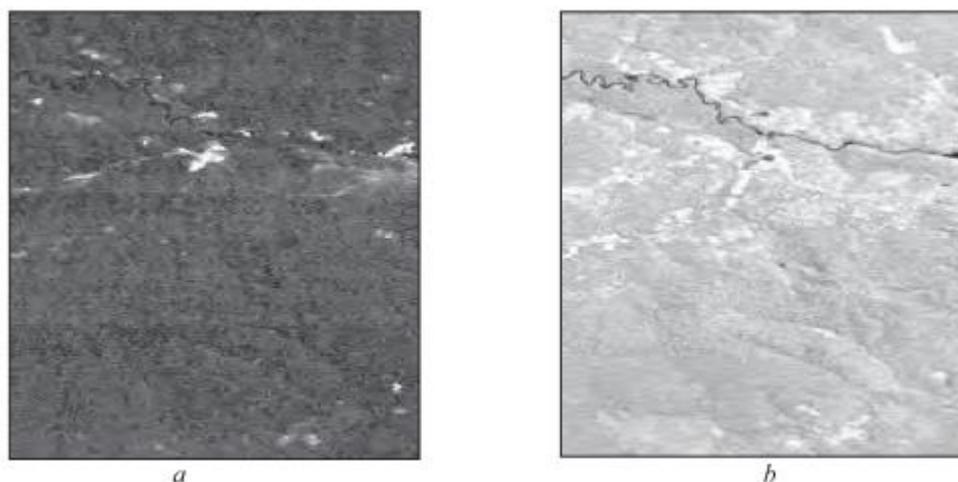
Uma outra propriedade de sensores de imageamento é a resolução espectral. Segundo Meneses e Almeida (2012), o termo resolução espectral envolve pelo menos três parâmetros de medida:

- i) o número de bandas que o sensor possui;
- ii) a largura em comprimento de onda das bandas;
- iii) as posições que as bandas estão situadas no espectro eletromagnético.

De forma comparativa, Meneses e Almeida (2012) afirmam que, um sensor tem melhor resolução espectral se ele possui maior número de bandas situadas em diferentes regiões espectrais e com larguras estreitas de comprimentos de onda. Os autores acrescentam que essa necessidade é devida às diferenças relativas de refletância entre os materiais da superfície da Terra, permitindo assim distinguir um material do outro, em determinados comprimentos de onda. Os autores também afirmam que a largura da banda deve ser dimensionada em concordância com as larguras das feições de absorção exibidas pelos espectros de refletância de cada material.

As feições de absorções são identificadores de composições dos tipos de rochas, solos, vegetação e água, e normalmente, são da ordem de 10 nm a 20 nm. Sensores que possuem bandas muito largas, além de 20 nm tendem a não ser capazes de diferenciar um objeto do outro em função de sua composição. Dessa forma, o objeto é detectado somente em razão da resolução espacial (MENESES e ALMEIDA, 2012).

Para ilustrar o efeito da resolução espectral na detecção ou identificação de um alvo, pode-se considerar o exemplo apresentado por Meneses e Almeida (2012). Para isso, considere a Figura 14. Nela é apresentada uma área de floresta da região amazônica. As duas imagens estão na resolução espacial de 30 metros. Na Figura 14(a), que corresponde a uma banda do visível na faixa entre  $0,63\mu\text{m}$  a  $0,69\mu\text{m}$ , os rios, apesar de possuírem dezenas de metros de largura, não são facilmente identificados, pois a refletância da água e da vegetação são, praticamente, iguais nesse comprimento de onda, não existindo contraste entre os dois tipos de alvos. Segundo Meneses e Almeida (2012), para a região amazônica, seria um erro selecionar esta banda espectral para se elaborar mapas de drenagem. Por outro lado, a Figura 14(b) mostra uma banda do infravermelho próximo. Segundo Meneses e Almeida (2012), nessa banda os rios são facilmente identificados, pois a água possui uma baixa refletância, enquanto a vegetação da floresta tem uma alta refletância. Nesse caso, ocorre uma razão de alto contraste entre os dois alvos.



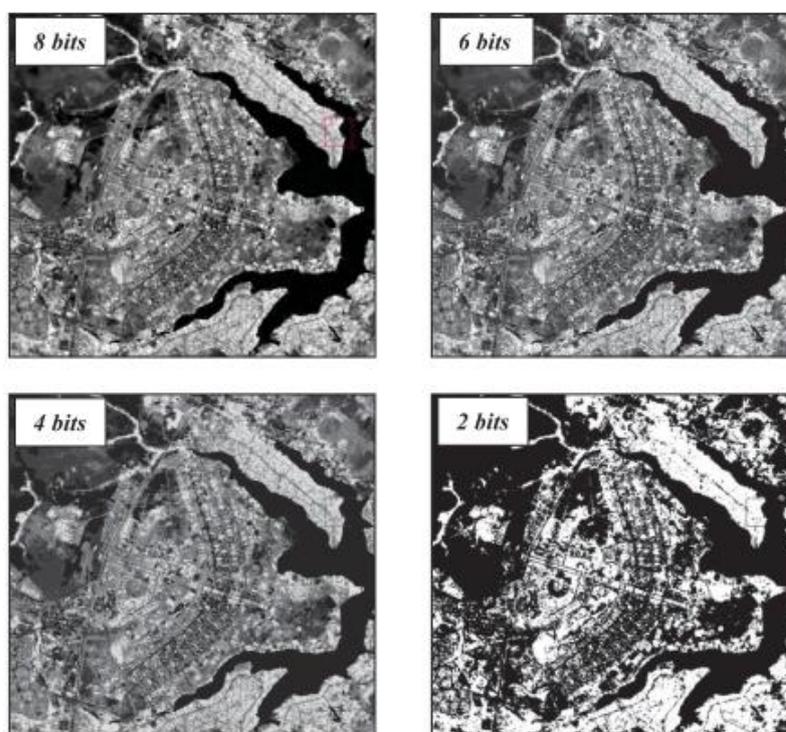
**Figura 14 – Exemplo do efeito da resolução espectral no contraste do rio com a vegetação. Imagens da região amazônica nas bandas: (a) visível e (b) infravermelho próximo. Fonte: (MENESES e ALMEIDA, 2012)**

### 3.1.1.3 Resolução radiométrica

Segundo Novo (2008), a resolução radiométrica de um sensor descreve a sua habilidade de distinguir variações no nível de energia refletida, emitida ou retro-espalhada que deixa a superfície do alvo. Novo (2008) acrescenta que quanto maior for a capacidade do sensor de distinguir diferenças de intensidade do sinal, maior será sua resolução radiométrica.

A medida radiométrica também pode ser chamada de quantização. De maneira prática, a quantização é medida pelo intervalo de número de valores digitais usados para expressar os

valores de radiância medidos pelo detector. Normalmente, é expressa em *bits*. Por exemplo, uma imagem discretizada em 8 bits pode ser representada com 256 níveis de cinza. Conforme exemplificado na Figura 15, quanto maior a quantização, melhor será a qualidade visual da imagem. Na Figura 15 podemos observar que a imagem de 8 *bits* tem detalhes visuais melhores que os das imagens de menores resoluções radiométricas, de 6, 4 e 2 bits. A imagem de 2 *bits* possui apenas 4 níveis de cinza. Em processamento digital de imagens, a medida radiométrica é também denominada de resolução de profundidade. (MENESES e ALMEIDA, 2012).



**Figura 15 – Exemplos de imagens com diferentes níveis de quantização ou de resolução radiométrica.**  
Fonte: (MENESES e ALMEIDA, 2012)

#### 3.1.1.4 Resolução temporal

O sensor possui ainda a resolução temporal. Essa resolução refere-se à frequência de imageamento sobre uma mesma área. Por exemplo, o sensor LANDSAT-5/TM possui uma resolução temporal de 16 dias, já o sensor instalado no satélite meteorológico Goes possui uma resolução temporal de meia hora (FLORENZANO, 2011).

## 3.2 LANDSAT

LANDSAT é um programa de satélite de observação da Terra de origem norte americana. As atividades iniciaram em 1972 com o lançamento do satélite ERTS-1. O programa

teve sequência com o LANDSAT 2, 3 4, 5, 6, 7 e atualmente encontra-se em funcionamento o LANDSAT-8/OLI. O objetivo do programa é fornecer dados multiespectrais da superfície da Terra (U.S. GEOLOGICAL SURVEY, 2019).

### 3.2.1 LANDSAT-8/OLI

As imagens geradas possuem um ciclo de repetição de 16 dias com resolução radiométrica de 16 *bits*. Os dados são coletados por meio de 11 bandas espectrais cuja resolução espacial é de 30m, com exceção da banda pancromática, que possui resolução de 15m e a banda infravermelho termal que possui resolução de 100 m. Na Tabela 3 são exibidas as bandas do LANDSAT-8/OLI com suas respectivas descrições, bem como, os dois tipos de sensores: Óptico, denominado *Operational Land Imager (OLI)* e o outro Termal, denominado de *Thermal InfraRed Sensor (TIRS)*.

**Tabela 3 – Bandas LANDSAT-8/OLI. Fonte: (U.S. GEOLOGICAL SURVEY, 2019)**

Sensor		Banda	Resolução Espacial	Comprimento de Onda (µm)
OLI	1	Aerosol	30 m	0,435 - 0,451
	2	Azul	30 m	0,452 - 0,512
	3	Verde	30 m	0,533 - 0,590
	4	Vermelho	30 m	0,636 - 0,673
	5	Infravermelho próximo	30 m	0,851 - 0,879
	6	Infravermelho de ondas curtas 1	30 m	1,5166 - 1,651
	7	Infravermelho de ondas curtas 2	30 m	2,107 - 2,294
	8	Pancromática	15 m	0,503 - 0,676
	9	<i>Cirrus</i>	30 m	1,363 - 1,384
TIRS	10	Infravermelho termal 1	100 m	10,60 - 11,19
	11	Infravermelho termal 2	100 m	11,50 - 12,51

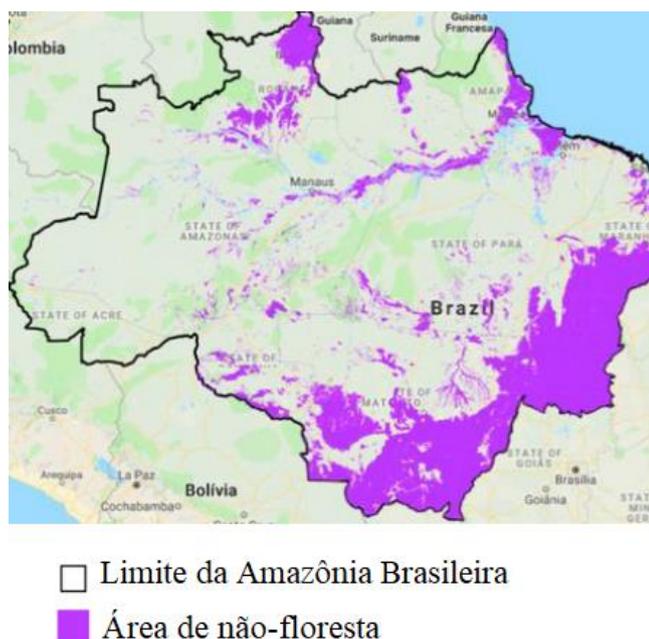
Cada imagem gerada pelo sistema LANDSAT tem um sistema de identificação das imagens composto de 2 números: o primeiro é o número da órbita e o segundo, o número da cena dentro daquela órbita, chamado de ponto.

### 3.3 Projeto TerraClass

A terra é um recurso natural muito importante. Devido ao crescimento da população e a disponibilidade de recursos limitados, um planejamento para a utilização adequada da terra é necessário. A informação sobre a taxa e o tipo de mudança no uso da terra é essencial para uma gestão e regularização de uso desses recursos. O levantamento de uso e cobertura da terra consiste na elaboração de mapas indicativos da distribuição geográfica dos usos através da

identificação de padrões homogêneos da cobertura terrestre. A consciência de que as modificações implementadas no uso e cobertura da terra implicam em impactos no clima local e global, e no funcionamento dos ecossistemas, trouxe uma nova linha de pesquisa para as ciências naturais e da terra. Sendo assim, o monitoramento de uso e cobertura do solo envolve capacitação multidisciplinar e serve de subsídio para diversos setores, sejam eles governamentais, privados, ou de pesquisa. Vale ressaltar que uso e cobertura do solo são termos que possuem significados distintos. Cobertura do solo está relacionado com os componentes que revestem a superfície da terra englobando solos, rochas, vegetação, florestas, água. Uso do solo refere-se às atividades humanas desenvolvidas diretamente na terra, descrevendo como a cobertura se modifica, como agricultura e pasto, construção e desflorestamento. A quantificação dos recursos do solo é possível com a utilização do sensoriamento remoto (AUGUSTO-SILVA *et al.*, 2013; BHARATKAR; PATEL, 2013).

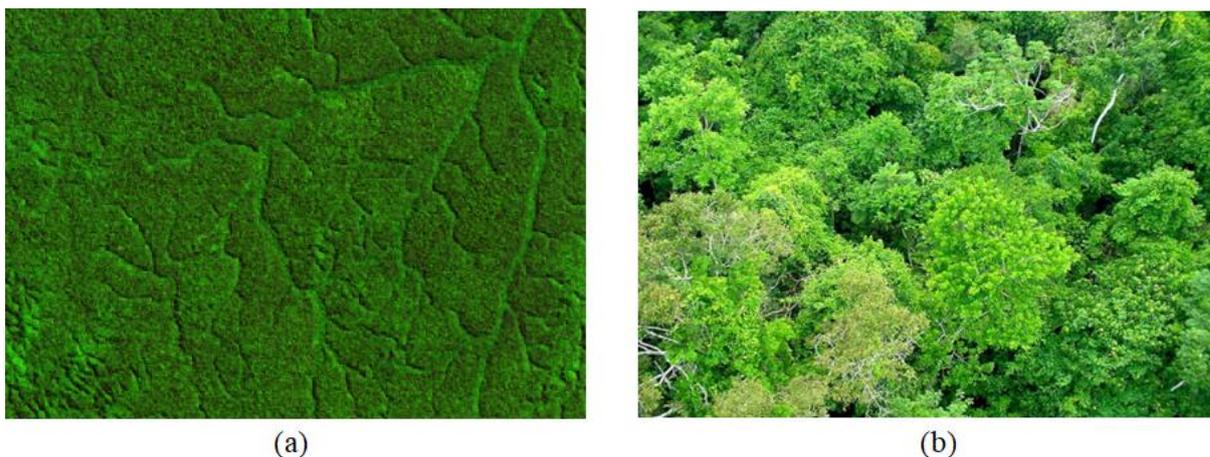
O Inpe possui diversos projetos de monitoramento da Amazônia, dentre esses existe o projeto chamado Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES). Esse monitoramento é realizado anualmente desde 1988. O objetivo do PRODES é estimar a taxa anual de desmatamento por corte raso da floresta primária na Amazônia Brasileira. No entanto, a região Amazônica possui outras áreas que não áreas de floresta e para excluir áreas que não pertencem a esse domínio, no início do PRODES, a equipe do INPE mapeou a chamada área de “não-floresta”, que não é considerada no mapeamento sistemático feito no PRODES até os dias de hoje. A área de não-floresta ocupa cerca de 961.000 km<sup>2</sup>, conforme apresentado na Figura 16. Outros tipos de regiões excluídas do mapeamento feito pelo PRODES são as áreas hidrográficas (INPE, 2019b).



**Figura 16 – Limite da Amazônia Legal Brasileira e área não mapeada pelo PRODES. Fonte: (INPE, 2019b)**

Outro projeto desenvolvido pelo INPE, juntamente com a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), é o projeto TerraClass. Esse projeto tem como objetivo produzir mapas sistêmicos de uso e cobertura das terras desflorestadas da Amazônia Legal brasileira. Esse projeto foi criado em 2010 para atender uma demanda do Governo Federal relacionada à qualificação dos desmatamentos observados na região da Amazônia Legal. Esse Projeto faz a quantificação das áreas de desmatamento mapeadas pelo projeto PRODES. Os dados gerados no projeto TerraClass delimitam as regiões nas seguintes classes: Floresta, Agricultura, pasto, área não observada, área urbana, mineração, outros, não-floresta e hidrografia (INPE, 2016).

As áreas de floresta são regiões que possuem vegetação natural e uma alta concentração de espécies de árvores. A Figura 24 mostra um exemplo que caracteriza esse tipo de região.



**Figura 17 – Exemplo de uma região de Floresta. Em (a) temos uma região da cena 230/065 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de floresta. Fonte: (INPE, 2016).**

As áreas de agricultura são regiões que apresentam culturas agrícolas, tais como plantações de café, cana-de-açúcar, soja, milho, algodão, entre outras. A Figura 18 ilustra esse tipo de região.



**Figura 18 – Exemplo de uma região de Agricultura. Em (a) temos uma região da cena 227/068 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de agricultura. Fonte: (INPE, 2016).**

As áreas de pasto são regiões que tiveram a vegetação original substituída por plantas usadas como fonte de alimento para animais, por exemplo para o gado. A Figura 18 apresenta um exemplo desse tipo de região.



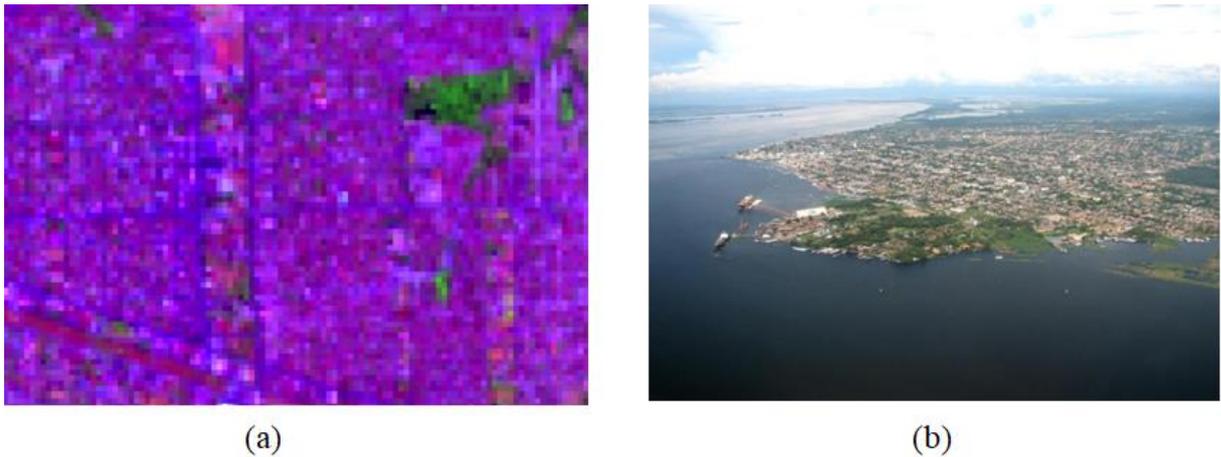
**Figura 19 – Exemplo de uma região de Pasto. Em (a) temos uma região da cena 227/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de pasto. Fonte: (INPE, 2016).**

As áreas não-observadas são regiões que não foram mapeadas devido à presença de nuvens, sombra de nuvens ou queimadas nas imagens de satélite utilizadas. A Figura 20 ilustra esse tipo de área.



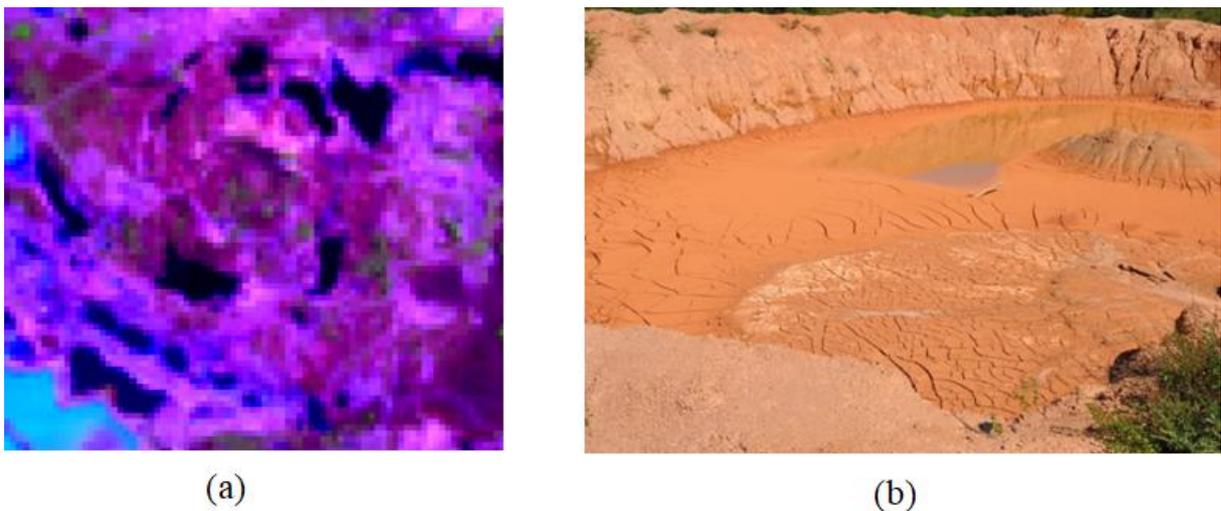
**Figura 20 – Exemplo de uma área não-observada. Em (a) temos uma região da cena 226/062 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma não-observada. Fonte: (INPE, 2016).**

As áreas urbanas são regiões de influência urbana como lugarejos, vilas, cidades ou regiões metropolitanas. Essas regiões apresentam arruamentos e infraestruturas residencial e industrial. A Figura 21 ilustra esse tipo de área.



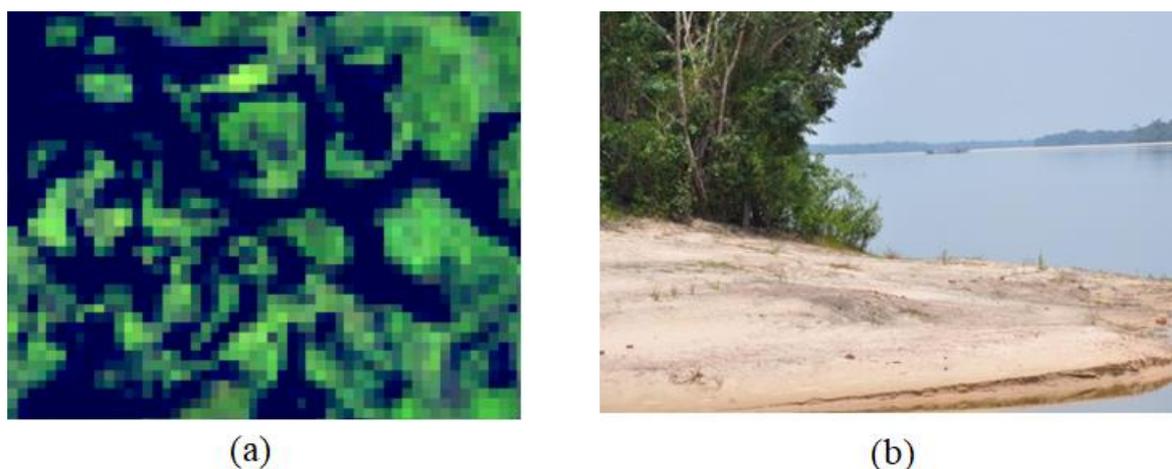
**Figura 21 – Exemplo de uma região de área urbana. Em (a) temos uma região da cena 232/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área urbana. Fonte: (INPE, 2016).**

As áreas de mineração são regiões de extração mineral que possuem solo exposto e alterações na paisagem local. A Figura 22 ilustra esse tipo de área.



**Figura 22 – Exemplo de uma região de mineração. Em (a) temos uma região da cena 232/067 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área de mineração. Fonte: (INPE, 2016).**

As áreas caracterizadas como outros são regiões que não se enquadram nas demais classes. Alguns exemplos desse tipo de região são: afloramentos rochosos, praias fluviais, bancos de areia. A Figura 23 apresenta um exemplo desse tipo de região.



**Figura 23 – Exemplo de uma região caracterizada como outros. Em (a) temos uma região da cena 230/062 do LANDSAT-8/OLI exibida em composição colorida B6-B5-B4. Em (b) temos uma foto de uma área chamada de outros. Fonte: (INPE, 2016).**

Segundo Inpe (2016), os dados do TerraClass para o ano de 2014 mostraram que aproximadamente 64% (3.178.580 km<sup>2</sup>) da Amazônia brasileira era coberta por áreas de floresta, 19% áreas de não-floresta (áreas correspondentes a outros usos de vegetação), 2% eram áreas de hidrografia e 15% eram áreas desmatadas. Entre esses 15% de áreas desmatadas, as áreas de agricultura correspondiam a 45.000 km<sup>2</sup> e as áreas de pasto correspondiam a 479.000 km<sup>2</sup>, isso representa 8% e 86% das áreas desmatadas da Amazônia, respectivamente.

O monitoramento da região amazônica realizado pelo INPE por meio do projeto TerraClass utiliza como metodologia a interpretação visual por meio de especialistas da área (INPE, 2019b). Dessa maneira, esse processo possui alta dependência de interação humana. O desenvolvimento de um modelo para segmentação automática pode melhorar o processo existente e inclusive possibilitar a criação de métodos mais eficientes. Neste caso, a detecção automática dos diversos usos do solo para a região amazônica traria uma importante contribuição metodológica.

### 3.4 Aprendizado profundo

A partir da primeira década deste século, a rede neural renasceu sob o novo rótulo conhecido como aprendizado profundo (AGGARWAL, 2018). O aprendizado profundo faz uso de modelos computacionais com arquiteturas hierárquicas compostas de múltiplas camadas de processamento, a fim de "aprender" representações de dados em formatos muito diferentes: áudio, imagens e texto (LECUN; BENGIO; HINTON, 2015). De acordo com Aggarwal, (2018), a grande quantidade de dados disponíveis nos últimos anos, juntamente com o aumento

do poder computacional, permitiu o uso de arquiteturas mais profundas que antes não eram possíveis.

Aprendizado Profundo refere-se a uma técnica de reconhecimento de padrões em que várias camadas de diferentes estágios de processamento da informação são usadas, não só para a classificação de padrões, mas também para o aprendizado da própria extração das características que serão utilizadas para a tarefa de classificação (DENG, 2014). É uma área multidisciplinar, que faz uma combinação de diversas áreas de pesquisa, tais como: redes neurais, modelos gráficos probabilísticos, otimização, reconhecimento de padrões e processamento de sinais.

A característica mais marcante desse modelo de aprendizado é justamente a arquitetura com muitas camadas para o reconhecimento de padrões, similarmente aos modelos atualmente existentes sobre o mecanismo de aprendizado do cérebro humano, principalmente em relação ao processo de aprendizagem visual (BENGIO, 2009; SERRE *et al.*, 2007). Diferentemente dos modelos de aprendizagem rasa (tradução livre para *Shallow Learning*), que são muito dependentes da intervenção humana para a extração de características dos dados a serem avaliados, os modelos baseados em aprendizado profundo podem aprender com menor envolvimento humano na construção do modelo pré-treinamento. O fato de integrarem os estágios de extração de característica e classificação na mesma estrutura possibilita menor custo computacional (BENGIO; LECUN, 2007).

Atualmente, existem diferentes modelos de aprendizado profundo, como por exemplo, Redes Neurais Recorrentes e CNNs. As Redes Neurais Recorrentes são projetadas para dados sequenciais como sentenças de texto, séries temporais e outras sequências discretas como sequências biológicas. Já as CNNs são utilizadas em aplicações de visão computacional para classificação de imagens e detecção de objetos (AGGARWAL, 2018). A metodologia proposta neste trabalho ora apresentado, avaliou diferentes modelos CNNs. A seguir é apresentada uma ideia geral sobre CNNs.

### **3.4.1 Redes Neurais Convolucionais (CNN)**

Segundo Aggarwal (2018), a motivação básica para a rede neural convolucional foi obtida do entendimento de Hubel e Wiesel (1959) sobre o funcionamento do córtex visual do gato, no qual porções específicas do campo visual pareciam excitar neurônios específicos. Este princípio mais amplo foi usado para projetar uma arquitetura esparsa para redes neurais convolucionais. A primeira arquitetura básica baseada nesta inspiração biológica foi o

neocognitron, que foi então generalizada para a arquitetura LeNet-5. Nas arquiteturas de CNNs, cada camada da rede é tridimensional, ou seja, possui uma dimensão espacial e uma profundidade correspondente ao número de características.

Outra característica das CNNs é o fato que os dados brutos podem ser aplicados diretamente à sua entrada. Não há necessidade de uma predefinição do conjunto de características que represente um dado modelo, visto que este papel é desempenhado pela própria rede. Segundo Albawi, Mohammed e Alzawi (2017) um aspecto benéfico das CNNs é reduzir o número de parâmetros na RNA. Essa conquista levou pesquisadores e desenvolvedores a abordar modelos maiores, a fim de resolver tarefas complexas, o que não era possível com RNAs clássicas. Por exemplo, em uma aplicação de detecção de face que faz uso de uma CNN, não é necessário se preocupar quanto ao local em que as faces estão localizadas nas imagens. O único objetivo é detectá-los, independentemente de sua posição nas imagens fornecidas. Outro aspecto importante da CNN é obtenção de recursos abstratos conforme a entrada se propaga para as camadas mais profundas da rede. Por exemplo, na classificação de imagens, a borda pode ser detectada nas primeiras camadas e, em seguida, as formas mais simples nas segundas camadas e, por fim, características mais complexas nas próximas camadas.

Para realizar tal feito, as CNN's são construídas basicamente com os seguintes conjuntos de camadas: camadas convolucionais; camadas de *polling*, camada ReLU (Unidades retificadoras lineares), camadas de convolução transposta, camada *dropout*, camada de normalização em lote. Ao final da rede, dependendo da aplicação, pode haver ainda camadas totalmente conectadas (*fully-connected layers*) e uma camada chamada *softmax* que tem como objetivo realizar a uma tarefa de classificação. A seguir, são apresentados os conceitos de cada camada que compõe a CNN.

#### **3.4.1.1 Camadas Convolucionais (*Convolutional Layers*)**

Para entender o conceito de camadas convolucionais, é necessário compreender o conceito de convolução.

Segundo Goodfellow, Bengio e Courville (2016), uma convolução pode ser definida como uma operação entre duas funções. A operação de convolução pode ser definida conforme a Equação (1), onde  $x$  é a entrada e  $w$  é o *kernel*. A saída  $s$  corresponde ao mapa de características.

$$s = (x * w) \quad (1)$$

Em aplicações que utilizam CNNs, onde os dados de entrada são imagens, é necessário realizar convoluções em mais de um eixo por vez, nesse caso um *kernel* bidimensional é utilizado. A expressão matemática para a convolução em CNNs pode ser representada por:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (2)$$

Onde  $S(i, j)$  corresponde às posições do mapa de características,  $K$  é o *kernel*,  $m$  e  $n$  se refere à dimensão  $K$ ,  $I$  é a imagem,  $i$  e  $j$  corresponde à dimensão de  $I$ .

Na Figura 24 é ilustrado graficamente uma convolução. Neste exemplo, a saída foi restringida às regiões onde o *kernel* encontra-se inteiramente dentro da imagem de entrada. Cada elemento da saída corresponde a soma dos produtos de cada elemento da entrada pelo parâmetro do *kernel*.

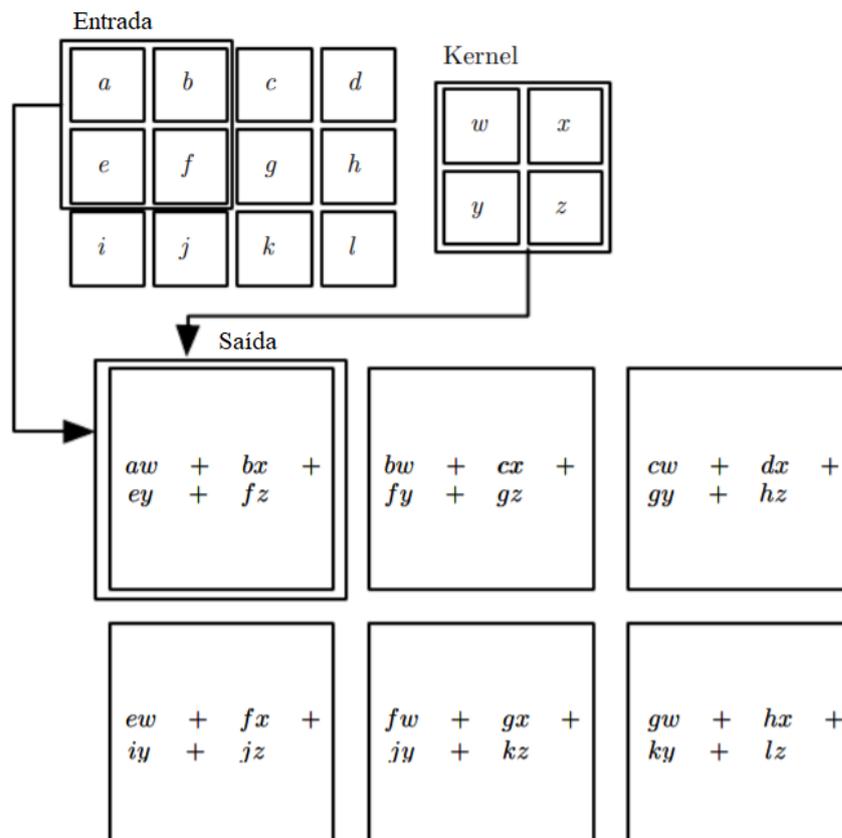


Figura 24 – Exemplo de uma convolução. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

Segundo Zeiler (2013), as camadas convolucionais constituem o primeiro conjunto de camadas de uma CNN e tem por objetivo obter, a partir da informação de entrada, mapas de características que permitem diferenciar as classes.

### 3.4.1.2 Camadas de *pooling* (subamostragem)

Esta camada pode ser utilizada para se reduzir o tamanho dos mapas de características obtidos pela camada convolucional (GUO *et al.*, 2016), e, assim como a primeira, também é invariante à translação.

As formas mais comuns de camada de amostragem são: operação de média e operação de máximo. Essa última operação é utilizada com mais frequência, devido a algumas vantagens, como uma convergência mais rápida no treinamento (SCHERER; MULLER; BEHNKE, 2010). A Figura 25 ilustra um exemplo de operação de *pooling* de tamanho 3x3. Nesse exemplo a entrada possui tamanho 7x7 e foi calculada a saída utilizando o valor máximo e passos 1 e 2. A saída com passo de 1 cria uma saída de 5x5 com elementos altamente repetitivos devido a maximização em regiões sobrepostas. Por outro lado, a saída gerada com o passo 2 possui menos sobreposição.

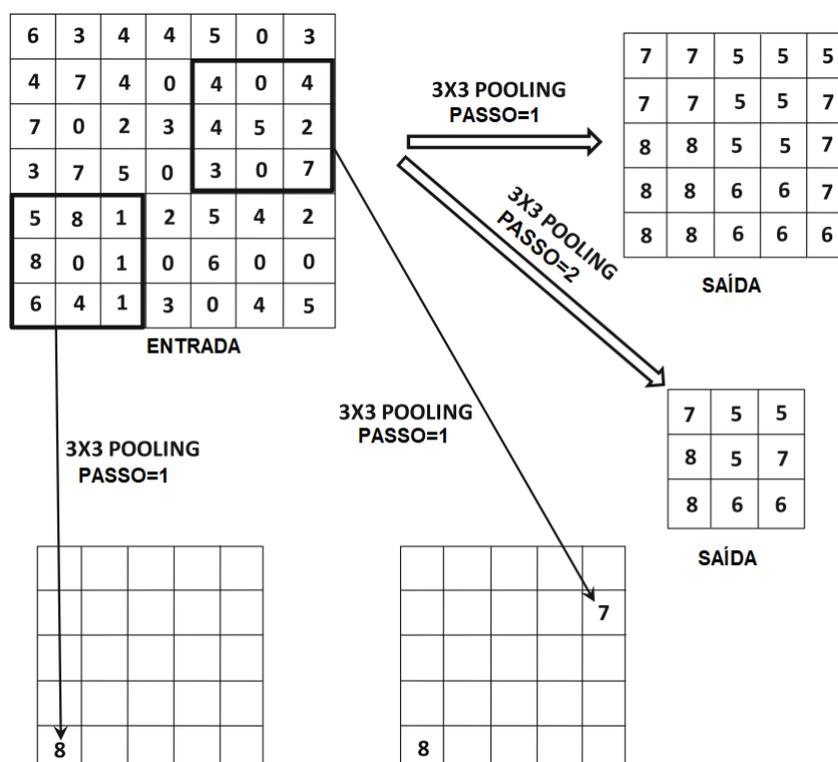


Figura 25 – Exemplo de operação de *pooling* utilizando o valor máximo. A entrada possui um tamanho de 7x7 e a saída foi calculada utilizando o valor máximo e os passos 1 e 2. Fonte: (GOODFELLOW; BENGIO; COURVILLE, 2016)

### 3.4.1.3 Camada ReLU

A rede CNN é constituída também de camadas não-lineares. Segundo Albawi, Mohammed e Alzawi (2017), a não linearidade pode ser usada para ajustar ou cortar a saída gerada. Essa camada é aplicada para saturar a saída ou limitar a saída gerada. Ainda segundo

os autores, a função sigmoide e tangente hiperbólica foram muito utilizadas na camada não-linear. No entanto, atualmente a função ReLU é mais indicada. A função ReLU é definida como:

$$ReLU(x) = \max\{0, x\}. \quad (3)$$

Conforme mostra a Eq. (3), a função ReLU possui uma definição simples, isso torna o processo de treinamento da rede mais rápido em relação a função sigmoide e tangente hiperbólica. Segundo Aggarwal (2018), a operação de convolução é intercalada com as operações de pooling e ReLU. A aplicação do ReLU não altera as dimensões de uma camada, pois é um simples mapeamento um a um dos valores de ativação. Em redes neurais tradicionais, a função de ativação é combinada com uma transformação linear com uma matriz de pesos para criar a próxima camada de ativações. Similarmente, uma ReLU normalmente segue uma operação de convolução que é o equivalente aproximado da transformação linear em redes neurais tradicionais.

#### 3.4.1.4 Camada de Convolução Transposta

A camada de convolução transposta tem como objetivo realizar uma operação de sobreamostragem na rede. Essa camada utiliza como entrada os mapas de características gerados pelas camadas *polling* e ReLU e a saída possui uma dimensão maior que a entrada da camada. Esse tipo de camada é empregado em aplicações que fazem uso de CNN's e que tem como objetivo realizar a segmentação semântica de uma imagem. Para realizar esse tipo de segmentação é necessário que as informações espaciais estejam preservadas. Desta maneira, cada pixel da imagem é classificado com sua classe correspondente gerando uma imagem do mesmo tamanho da imagem de entrada da rede. A Figura 26 ilustra um exemplo a operação de convolução transposta com passo 1. A entrada possui tamanho 2x2. A região em branco é preenchida com zeros e o *kernel* utilizado de tamanho 3x3. O resultado da convolução transposta é uma matriz de tamanho 4x4.

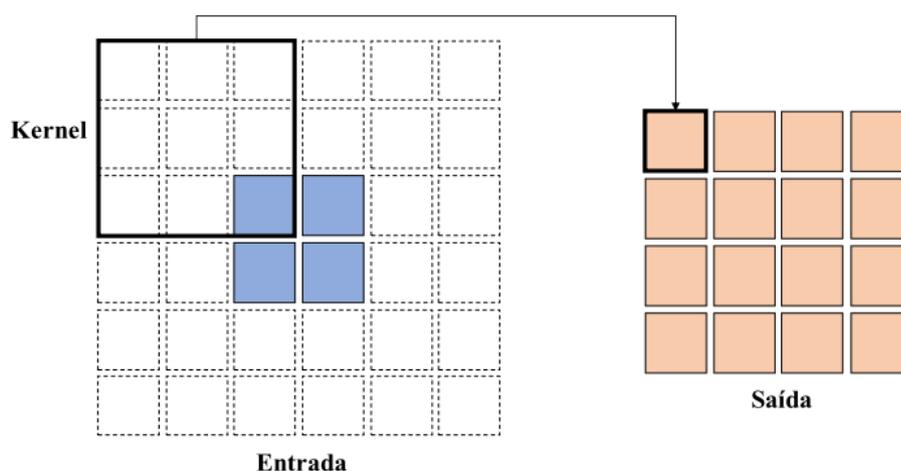


Figura 26 – Exemplo de convolução transposta com passo 1. Fonte: (SERRÃO, 2020)

### 3.4.1.5 Camada *dropout*

O método *dropout* é uma técnica que busca evitar o *overfitting*, fenômeno no qual a rede realiza um treinamento excessivo e que provoca um desempenho ruim quando a rede é submetida ao conjunto de testes.

Segundo Aggarwal (2018), *dropout* é um método que tem como objetivo criar um conjunto de redes neurais por meio de eliminação de neurônios. Caso um neurônio seja descartado, todas as conexões de entrada e saída desse neurônio são eliminadas. Os neurônios descartados são apenas das camadas de entrada e ocultas da rede. Durante o treinamento, os neurônios possuem uma probabilidade  $p$  de não serem eliminados. Os neurônios eliminados não participam da rede, como mostra a Figura 27.

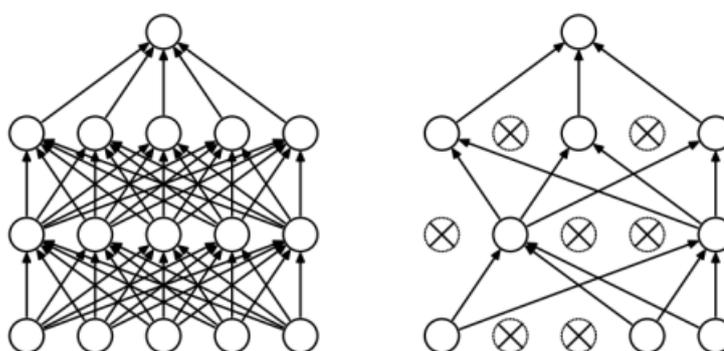


Figura 27 – Exemplo de uso da camada de *dropout*. A rede da direita corresponde a aplicação do *dropout* na rede da esquerda. Fonte: (SRIVASTAVA *et al.*, 2014)

### 3.4.1.6 Camada de normalização em lote

Segundo Aggarwal (2018), a normalização em lote é um método empregado para eliminar o problema da mudança de covariância interna. Esse problema ocorre durante o

treinamento da rede, onde as mudanças dos parâmetros da rede causam variações na distribuição das ativações da rede nas camadas ocultas. Ainda segundo o autor, as camadas de normalização em lotes são adicionadas entre as camadas ocultas, criando ativações com uma mesma faixa de variação de valores. A normalização em lote permite taxas de aprendizado maiores e contribui na generalização do modelo.

#### 3.4.1.7 Regularização $L_2$

A regularização é um método que busca melhorar a capacidade de generalização dos algoritmos de aprendizado por meio de alguma restrição durante a etapa de treinamento. A teoria de regularização foi proposta por Tikhonov e Arsenin (1977), para resolver problemas mal-condicionados (mal postos). Segundo os autores, a ideia básica da regularização é estabilizar a solução por meio de algum funcional não-negativo que incorpore informação prévia sobre a solução.

Neste trabalho foi utilizado a regularização  $L_2$  que consiste na adição de um termo à função de custo  $E$ . O termo adicionado à função de custo é a soma do quadrado de todos os pesos  $w$  da rede escalonado por um fator  $\frac{\lambda}{2n}$ , onde  $\lambda$  é o parâmetro de regularização e  $n$  é o tamanho do conjunto de treinamento. Com a regularização, a rede tende a minimizar a função de custo e a aprender pesos pequenos. Sendo assim, para pequenos valores de  $\lambda$ , a rede prioriza minimizar a função de custo original, e para altos valores de  $\lambda$ , a rede prioriza aprender pesos pequenos. A Eq. 4 mostra a função de custo regularizada  $E_R$  (SERRÃO, 2020).

$$E_R = E + \frac{\lambda}{2n} \sum_w w^2 \quad (4)$$

#### 3.4.1.8 Segmentação Semântica

A segmentação semântica é um dos avanços recentes em aprendizado profundo. Este tipo de segmentação fornece uma classificação a nível de pixel de uma imagem. Quando uma CNN realiza uma classificação semântica em uma imagem, é realizada uma associação de cada pixel da imagem a uma determinada classe (MATHWORKS, 2018). Na Figura 28 podemos observar um exemplo do processo de segmentação semântica de uma imagem. Na primeira metade da rede são utilizadas camadas *pooling* que reduzem o mapa de característica da imagem e na segunda metade são empregadas camadas de convolução transposta que realizam uma sobreamostragem. No final desse processo, o resultado é uma nova imagem onde cada pixel possui uma classe atribuída.



Figura 28 – Exemplo de segmentação semântica. Fonte: (MATHWORKS, 2018).

### 3.4.1.9 Camadas inteiramente conectadas (*Fully-connected layers*)

Essas camadas desempenham um papel similar ao das camadas de uma rede neural tradicional. Possuem cerca de 90% dos parâmetros da CNN e mudanças em sua estrutura não são comuns.

O ponto fraco dessa camada é justamente possuir um número muito grande de parâmetros, que resulta em uma alta complexidade computacional durante seu treinamento. No entanto, como se situam na saída da CNN, possui um menor número de ativações, o que contribui para uma redução do número de parâmetros.

Com o avanço nos últimos anos das redes neurais convolucionais em aplicações de visão computacional, surgiram algumas arquiteturas que se destacaram em tarefas de reconhecimento de objetos. Essas arquiteturas são listadas na Tabela 4.

Tabela 4 – Algumas das arquiteturas mais difundidas FONTE: (GUO *et al.*, 2016)

Método	Ano	Configuração
AlexNet	2012	5 camadas convolucionais + 3 inteiramente conectadas
Clarifai	2013	5 camadas convolucionais + 3 inteiramente conectadas
SPP	2014	5 camadas convolucionais + 3 inteiramente conectadas
VGG	2014	13 a 15 camadas convolucionais + 3 inteiramente conectadas
GoogLeNet	2014	22 camadas convolucionais + 1 inteiramente conectadas

### 3.5 Métodos de otimização

O processo de treinamento de uma rede neural é uma tarefa que demanda muito poder computacional. Com o objetivo de acelerar esse processo, vários métodos de otimização foram propostos. Neste trabalho, foram realizadas simulações com os métodos Gradiente Descendente Estocástico com Momento (SGDM, do inglês *Stochastic Gradient Descent with Momentum*), Propagação da Raiz Média Quadrática (RMSProp, do inglês *Root Mean Square Propagation*) e Estimativa de Dinâmica Adaptativa (ADAM, do inglês *Adaptive Moment Estimation*), objetivando verificar o método que resultou no melhor desempenho da rede. Nas seções seguintes são apresentados esses três métodos.

#### 3.5.1 SGDM

Segundo Goodfellow, Bengio e Courville (2016), o gradiente descendente estocástico (SGD, do inglês *Stochastic Gradient Descent*) e suas variantes são o algoritmo de otimização mais utilizado em aprendizado de máquina em geral e em aprendizado profundo em particular. No entanto, ainda segundo os autores, o treinamento utilizando o SGD às vezes pode ser lento. Com o objetivo de acelerar o aprendizado, uma alternativa é utilizar o SGDM.

O SGDM introduz uma variável que controla a velocidade na qual os parâmetros se movem pelo espaço de parâmetros. Esse método possibilita uma convergência mais rápida, pois a acelera taxa de aprendizado em regiões de gradiente constante e evita oscilações desnecessárias. Na Figura 29 podemos observar o efeito do momento na suavização das oscilações no movimento do gradiente descendente até o ponto ótimo: a) ilustra o método SGD e b) ilustra o método SGDM.



**Figura 29 – Efeito do momento na suavização de oscilações: (a) método SGD (b) método SGDM. Fonte: Adaptado de Aggarwal (2018)**

### 3.5.2 RMSProp

O algoritmo RMSProp é uma versão modificada de outro algoritmo, o AdaGrad. O AdaGrad realiza uma adaptação individual das taxas de aprendizado de todos os parâmetros do modelo por meio um escalonamento inversamente proporcional à raiz quadrada da soma de todos os valores históricos ao quadrado. No contexto de otimização convexa, o algoritmo AdaGrad possui propriedades teóricas desejáveis. No entanto, empiricamente, verificou-se que o acúmulo de gradientes quadrados desde o início do treinamento pode resultar em uma diminuição prematura e excessiva na taxa de aprendizado. Por outro lado, o RMSProp modifica o AdaGrad para ter um desempenho melhor na configuração não convexa, alterando a acumulação de gradiente em uma média móvel ponderada exponencialmente, de forma que valores históricos mais antigos da trajetória são descartados, tendo em vista que são atribuídos pesos menores aos dados mais antigos. Empiricamente, o RMSProp demonstrou ser um algoritmo de otimização prático e eficaz para redes neurais profundas. Atualmente, é um dos métodos de otimização mais empregados por profissionais de aprendizagem profunda (AGGARWAL, 2018).

### 3.5.3 Adam

O nome Adam vem da frase “momentos adaptativos, do inglês *“adaptive moments”*”. Adam é um algoritmo voltado para otimização de primeira ordem baseada em gradientes de funções objetivas estocásticas, com base em estimativas adaptativas de momentos de ordem inferior. O Adam é um método para otimização estocástica eficiente que requer apenas gradientes de primeira ordem com pouco requisito de memória. O método calcula as taxas de aprendizado adaptativo individual para diferentes parâmetros das estimativas do primeiro e do segundo momento dos gradientes. O Adam foi projetado para combinar as vantagens dos algoritmos AdaGrad e RMSProp (KINGMA; BA, 2015). Segundo Aggarwal (2018), existem duas principais diferenças entre o Adam e RMSProp. Primeiro, em Adam, o momento é incorporado diretamente utilizando uma ponderação exponencial do gradiente. Segundo, Adam realiza correções nas estimativas dos momentos de primeira ordem (termo do momento) e dos momentos de segunda ordem.

### 3.6 Transferência de conhecimento (*Transfer Learning*)

Segundo Goodfellow, Bengio e Courville (2016), transferência de conhecimento se refere à situação na qual o conhecimento adquirido para resolver determinado problema é explorado para melhorar a generalização em outro problema.

Segundo Pan e Yang (2010), pode-se classificar a transferência de conhecimento em transferência indutiva, transdutora e não-supervisionada. A transferência indutiva tem como característica o fato de a tarefa de destino ser diferente da tarefa de origem, independentemente de os domínios de origem e a tarefa de destino serem iguais ou não. Na transferência transdutora as tarefas de origem e destino são as mesmas, enquanto o domínio de aprendizado de origem e de destino são diferentes.

Por fim, a transferência não-supervisionada é semelhante à configuração de transferência de conhecimento indutiva. Na transferência não-supervisionada a tarefa de destino é diferente, mas relacionada à tarefa de origem. No entanto, a transferência não supervisionada concentra-se na resolução de tarefas de aprendizado não supervisionado no domínio de destino, como clusterização, redução de dimensionalidade e estimativa de densidade (PAN; YANG, 2010).

A técnica de transferência de conhecimento pode ser utilizada de três formas: ajuste raso (*shallow tuning*), ajuste fino (*fine tuning*) e ajuste profundo (*deep tuning*). No ajuste raso, apenas a camada de classificação da rede é treinada e todos os parâmetros das camadas anteriores são congelados. O ajuste fino, tem como objetivo treinar gradualmente mais camadas, ajustando os parâmetros da rede até que um aumento significativo do desempenho seja alcançado. Por fim, o ajuste profundo tem como objetivo treinar todos os parâmetros de todas as camadas da rede.

## CONSIDERAÇÕES FINAIS

Este capítulo apresentou conceitos essenciais para uma melhor compreensão deste trabalho. Os vários assuntos abordados, imagens ópticas, sensoriamento remoto, imagens LANDSAT-8/OLI, técnicas de aprendizado profundo, CNN, métodos de otimização e transferência de conhecimento, são essenciais para o entendimento desse trabalho de doutorado. O próximo capítulo explicará detalhadamente os métodos utilizados, que fazem uso dos conceitos ora apresentados.

## **4 MATERIAIS E MÉTODOS**

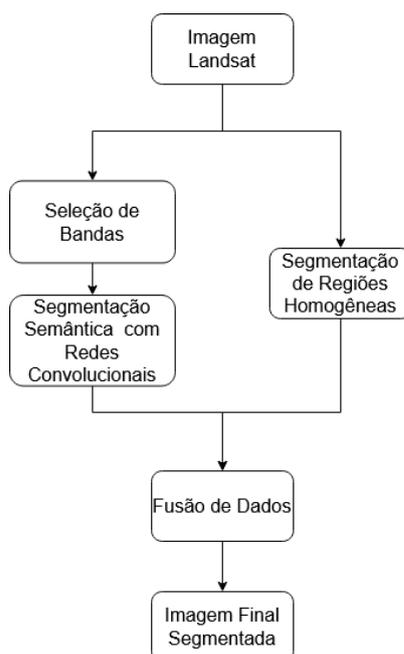
Este trabalho tem como objetivo propor um método para a segmentação de uso do solo na região amazônica. Neste Capítulo serão apresentados os materiais e a metodologia empregados neste trabalho.

A Figura 30 mostra um fluxograma da metodologia utilizada para a segmentação do uso do solo. Como mostrado, há dois caminhos: o esquerdo corresponde ao processamento da CNN, enquanto o direito, à segmentação com regiões homogêneas.

Ao iniciar o desenvolvimento deste trabalho, a metodologia proposta incluía apenas o processamento da CNN, que corresponde ao ramo esquerdo da Figura 30. No entanto, os resultados obtidos para segmentação de imagens reais da região de estudo utilizando apenas o modelo CNN não foram satisfatórios. Dessa forma, decidiu-se aprimorar os resultados obtidos com a CNN através da fusão das informações na sua saída com informações provenientes de um segundo método, o método de segmentação de regiões homogêneas, que corresponde ao ramo direito da Figura 30. Observou-se uma sensível melhoria das métricas utilizadas para avaliação de desempenho utilizadas nesse trabalho.

Do ponto de vista operacional, no ramo esquerdo da Figura 30, inicialmente é efetuada a seleção das bandas (3 ou 6 bandas) a serem apresentadas a entrada dela. Cada banda corresponde a uma imagem de entrada da CNN. Apresentando esse conjunto de imagens à entrada da CNN a mesma, através de segmentação semântica, realiza-se, conforme a aplicação para qual a mesma foi treinada, a identificação dos diferentes usos do solo ou de regiões desmatadas e não desmatadas. No ramo direito, realiza-se a segmentação das regiões homogêneas. No final, unindo os dois ramos, temos o bloco de fusão de dados.

Nas seções seguintes, procura-se explicar com detalhes as operações realizadas em cada um dos blocos mostrados na Figura 30.



**Figura 30 – Diagrama em blocos da metodologia empregada**

#### 4.1 Imagem LANDSAT-8/OLI

Neste trabalho foram utilizadas um total de 75 imagens LANDSAT-8/OLI de áreas da região da Amazônia Legal Brasileira. Estas imagens estão disponíveis gratuitamente em Usgs (2019b). A região de estudo está localizada no chamado “arco do desmatamento”. Essa região possui os maiores índices de desmatamento na Amazônia Legal Brasileira e uma grande expansão agropecuária (OVIEDO; LIMA; AUGUSTO, 2019).

As imagens utilizadas abrangem os estados do Amazonas, Mato Grosso, Pará e Rondônia, conforme mostrado na Tabela 5. Foram utilizadas imagens correspondentes a estação seca, devido a menor incidência de cobertura de nuvens. A Figura 31 apresenta um mapa com as cenas LANDSAT-8/OLI avaliadas neste trabalho.

**Tabela 5 – Lista de Imagens utilizadas.**

Orbita/Ponto	Data	Estado	Orbita/Ponto	Data	Estado	Orbita/Ponto	Data	Estado
229/65	09/08/2014	AM	227/70	11/08/2014	MT	224/62	22/08/2014	PA
229/66	09/08/2014	AM	228/70	18/08/2014	MT	228/65	02/08/2014	PA
230/64	16/08/2014	AM	224/69	06/08/2014	MT	228/66	18/08/2014	PA
230/65	16/08/2014	AM	226/69	04/08/2014	MT	227/66	26/07/2014	PA
230/66	15/07/2014	AM	227/69	11/08/2014	MT	226/66	04/08/2014	PA
231/64	08/09/2014	AM	228/69	01/07/2014	MT	225/66	13/08/2014	PA
231/65	23/08/2014	AM	225/71	13/08/2014	MT	223/63	12/06/2014	PA
231/66	23/08/2014	AM	226/71	20/08/2014	MT	225/65	28/07/2014	PA
232/64	30/08/2014	AM	227/71	11/08/2014	MT	224/64	05/07/2014	PA
232/65	30/08/2014	AM	224/68	21/07/2014	MT	224/65	22/08/2014	PA
001/65	12/08/2014	AM	226/68	04/08/2014	MT	224/66	06/08/2014	PA

Orbita/Ponto	Data	Estado	Orbita/Ponto	Data	Estado	Orbita/Ponto	Data	Estado
001/66	12/08/2014	AM	227/68	11/08/2014	MT	223/67	15/08/2014	PA
001/67	12/08/2014	AM	228/68	18/08/2014	MT	226/62	20/08/2014	PA
002/65	03/08/2014	AM	224/67	06/08/2014	MT	225/62	28/07/2014	PA
233/65	21/08/2014	AM	225/67	13/08/2014	MT	223/66	30/07/2014	PA
233/66	21/08/2014	AM	226/67	04/08/2014	MT	227/65	11/08/2014	PA
233/67	21/08/2014	AM	227/67	26/07/2014	MT	222/62	27/10/2014	PA
002/66	19/08/2014	AM	228/67	18/08/2014	MT	223/62	16/09/2014	PA
224/70	22/08/2014	MT	225/69	13/08/2014	MT	224/63	05/07/2014	PA
225/70	29/08/2014	MT	229/68	09/08/2014	MT	232/66	14/08/2014	RO
226/70	04/08/2014	MT	229/70	09/08/2014	MT	232/67	14/08/2014	RO
229/71	09/08/2014	MT	229/67	09/08/2014	MT	231/67	23/08/2014	RO
228/71	18/08/2014	MT	229/69	09/08/2014	MT	231/68	23/08/2014	RO
223/68	15/08/2014	MT	225/68	13/08/2014	MT	230/69	31/07/2014	RO
230/67	16/08/2014	MT	226/65	04/08/2014	PA	230/68	16/08/2014	RO

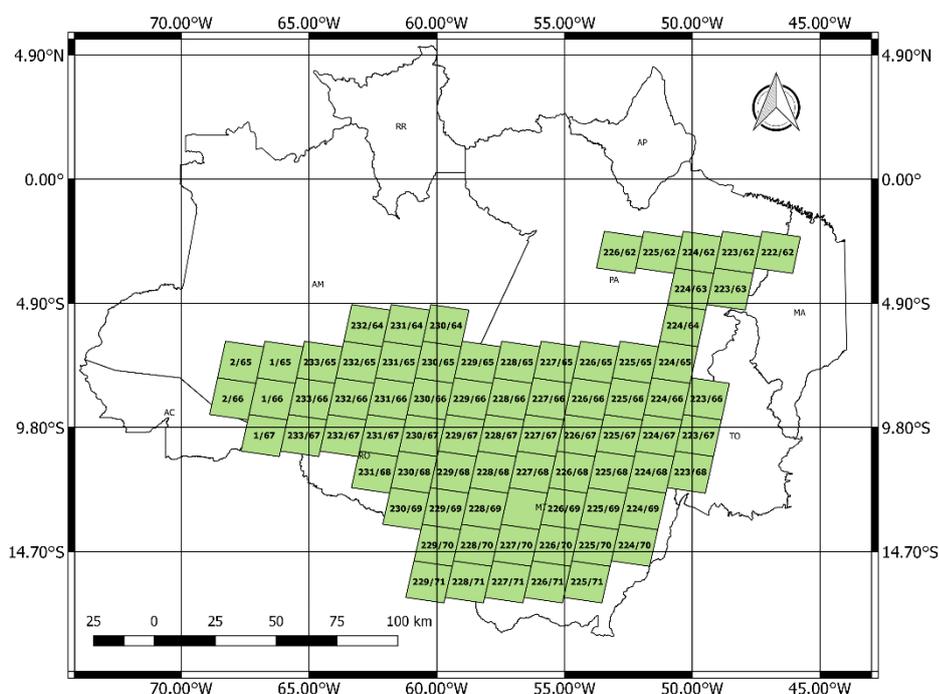


Figura 31 – Amazônia Legal com as cenas LANDSAT-8/OLI avaliadas neste trabalho.

#### 4.1.1 Dados de referências

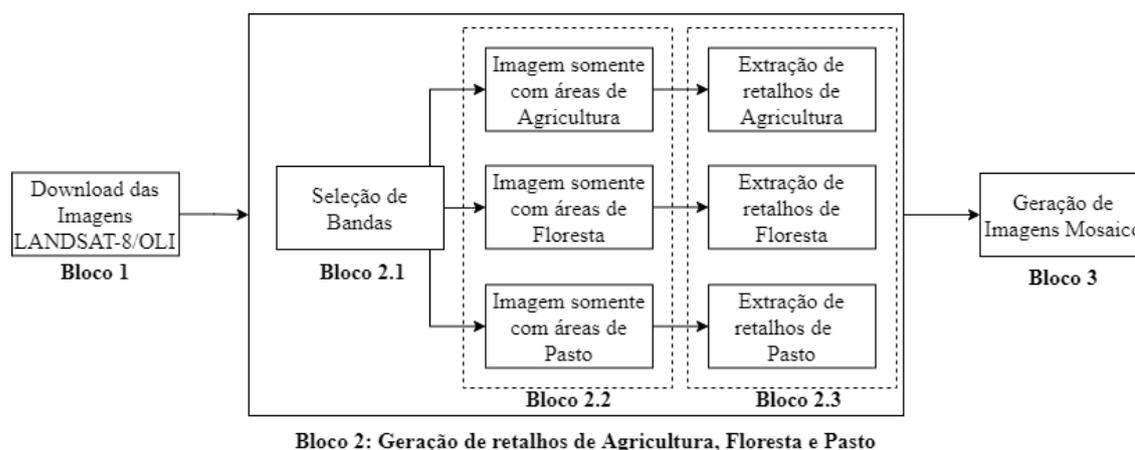
Para a geração do padrão ouro de treinamento da CNN, quando se trabalhou com a classificação do uso do solo, foram utilizados os dados do projeto TerraClass, descrito na Seção 3.3, do ano de 2014. Por outro lado, quando se trabalhou com identificação de desmatamento, foram utilizados os dados do projeto PRODES. Todas as informações de ambos os projetos foram adquiridas gratuitamente em (INPE, 2019a). Os dados gerados no projeto TerraClass delimitam as regiões da Amazônia brasileira nas seguintes classes: floresta, agricultura, pasto, área não observada, área urbana, mineração, outros, não floresta e hidrografia. Para este

trabalho, foram utilizadas informações sobre áreas de floresta, pasto e agricultura, correspondendo a um problema de reconhecimento de três classes. As classes área não observada, área urbana, mineração e outros não foram consideradas, pois representam uma área muito pequena e não seria possível extrair uma quantidade significativa de amostras para serem utilizadas no treinamento e uma ferramenta para reconhecimento de padrões, principalmente das redes CNNs, necessitam de uma maior quantidade de informações para o seu treinamento. Já as áreas de não-floresta e hidrografia não foram utilizadas pois são áreas que fazem parte de uma máscara de exclusão do INPE. Dessa forma, essas regiões não são consideradas no mapeamento sistemático.

#### **4.2 Seleção de Bandas**

Nesta seção é apresentado o procedimento utilizado para criar o banco de imagens utilizado neste trabalho. Segundo Usgs (2019b) as bandas Azul (B2), Verde (B3), Vermelho (B4), Infravermelho próximo (B5), infravermelho de ondas curtas 1 (B6) e Infravermelho de ondas curtas 2 (B7) são as mais adequadas para a análise de vegetação. Segundo Yu *et al.* (2019), B4, B5 e B6 é a melhor combinação de três bandas para aplicações de sensoriamento remoto em aplicações cujo objetivo seja realizar a classificação do uso do solo. Dessa maneira, foram montadas duas versões do banco de imagem, a primeira com as seis bandas (B2, B3, B4, B5, B6 E B7) e a segunda com três bandas (B4, B5, B6), com o objetivo de avaliar qual o conjunto de bandas mais adequado para realizar a segmentação de imagens de sensoriamento remoto para a região da Amazônia Legal Brasileira.

A Figura 32 apresenta um fluxograma da metodologia utilizada para criar o banco de dados de imagens. As etapas apresentadas neste fluxograma foram aplicadas para a criação do banco de imagens referente a aplicação de uso do solo e também para a montagem do banco de imagem de áreas desmatadas e não desmatadas. No entanto, detalhes em relação a criação do banco de imagem de desmatamento e não desmatamento serão fornecidos nas aplicações a serem mostradas no capítulo de resultados para detecção de áreas desmatadas. Doravante, nessa seção, far-se-á referência apenas à classificação do uso do solo.



**Figura 32 – Fluxograma da metodologia utilizada para a obtenção das Imagens do Mosaico**

Inicialmente foram baixadas todas as imagens LANDSAT-8/OLI, conforme apresentado no Bloco 1.

Para cada imagem mostrada na Tabela 5, os passos mostrados no Bloco 2 da Figura 32 foram aplicados. Este Bloco é responsável pela geração de retalhos de imagens de agricultura, floresta e de pasto. Inicialmente foi criada uma imagem resultante da composição das bandas B2, B3, B4, B5, B6 e B7 (Bloco 2.1). Em seguida, utilizando como referência os dados do Projeto TerraClass descritos na Seção 3.3, como mostrado no Bloco 2.2, são geradas três imagens, apenas com áreas agricultura, floresta ou de pasto. As etapas apresentadas nos Blocos 2.1 e 2.2 foram realizadas utilizando o software ENVI 5.5.

Finalmente, no Bloco 2.3 da Figura 32, utilizando o *software* MATLAB, foram gerados retalhos de imagens de tamanho 40x40 *pixels* de cada classe correspondente do uso do solo. A Figura 33 apresenta três exemplos desses retalhos, mostrando um retalho correspondente a floresta, outro correspondente a pasto e outro correspondente a agricultura. Foram gerados 4.000 retalhos de agricultura, 225.000 retalhos de floresta e 6.000 retalhos correspondentes a regiões de pasto. A opção de se utilizar retalhos com tamanho de 40x40 *pixels* carece de fundamentação teórica. Dessa forma, no capítulo de resultados, serão mostrados também resultados obtidos com imagens-mosaico construídas com retalhos de 20x20 e 10x10 *pixels*.



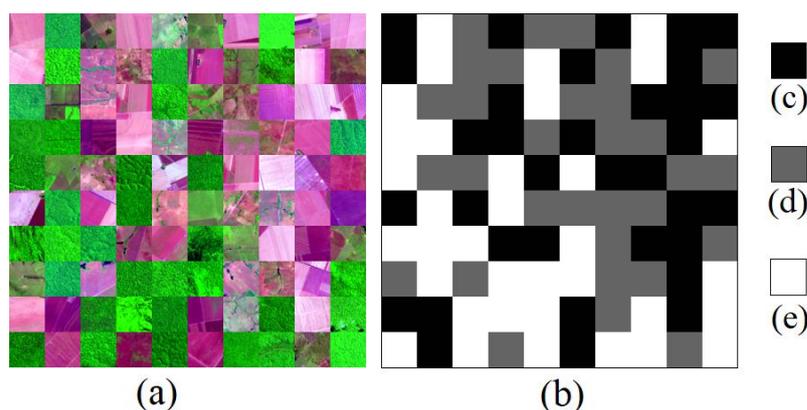
**Figura 33 – Retalhos de imagens apresentados composição colorida utilizando as bandas B6-B5-B4. (a) floresta, (b) pasto e (c) agricultura**

Para a montagem do conjunto de dados de treinamento das arquiteturas CNN propostas neste trabalho para classificação do uso do solo, os retalhos de imagens foram divididos em três partes, correspondentes a treinamento, validação e teste, conforme apresentado na Tabela 6. Dos 4.000 retalhos de agricultura, 2.000 foram separados para treinamento, 1.000 para validação e 1.000 para teste. Em relação aos 6.000 retalhos de pasto, 3.000 foram destinados para treinamento, 1.500 para validação e 1.500 para teste. Por fim, os 225.000 retalhos de floresta foram divididos aleatoriamente em 75.000 para treinamento, 75.000 para validação e 75.000 para teste.

**Tabela 6 – Divisão dos retalhos de imagens**

Classe	Treinamento	Validação	Teste
Agricultura	2000	1000	1000
Floresta	75000	75000	75000
Pasto	3000	1500	1500

No Bloco 3 da Figura 32, utilizando os conjuntos de retalhos de imagens de treinamento, validação e teste descritos anteriormente, foram geradas imagens definidas como imagens-mosaico. Cada uma dessas imagens-mosaico tem dimensão de 400x400 pixels e são compostas por retalhos de agricultura, floresta e pasto. Cada unidade da imagem-mosaico é gerada selecionando-se aleatoriamente um retalho de agricultura, floresta ou pasto. Para cada imagem-mosaico gerada, gera-se também uma imagem-mosaico correspondente ao padrão-ouro dela. Para a construção dessas imagens que constituíram o padrão-ouro, os pixels que correspondem a região de floresta foram marcados com o valor 255, os pixels correspondentes à região de pasto foram marcados com o valor 100 e os pixels correspondentes à região de agricultura foram marcados com o valor 1. Na Figura 34 mostra-se um exemplo de imagem-mosaico e o padrão-ouro correspondente. Pode ser observado que, dos 100 retalhos, 34 são de agricultura, 35 de floresta e 31 de pasto.



**Figura 34 – Exemplo de imagem em mosaico. Em (a) é mostrada uma imagem em mosaico na composição colorida utilizando as bandas B6-B5-B4, composta de retalhos de agricultura, floresta e de pasto, em (b) a imagem padrão ouro, e em (c), (d) e (e) temos o padrão ouro para cada retalho de imagem, correspondente à área agrícola, de pasto e floresta, respectivamente.**

Dois conjuntos de dados de imagens foram geradas com um total de 4.000 imagens-mosaico cada uma (imagens com 6 bandas e imagens com 3 bandas). Para cada versão, 70% das imagens (2800) foram destinadas para o conjunto de treinamento, 15% (600), para o conjunto de validação e 15% (600), para o conjunto de teste. O primeiro conjunto é composto por imagens com 6 bandas: B2, B3, B4, B5, B6 e B7. O segundo conjunto é composto por imagens com três bandas: B4, B5 e B6. A Tabela 7 apresenta o quantitativo de imagens dos dois conjuntos de dados de imagem gerados utilizados neste trabalho para classificação do uso do solo, mostrando a distribuição das imagens dos mesmos nos conjuntos de treinamento, validação e teste.

**Tabela 7 – Banco de Imagens-mosaicos gerados com a quantidade para os conjuntos de treinamento, validação e teste**

Versão	Bandas	Total de imagens-mosaico	Treinamento	Validação	Teste
1	B2, B3, B4, B5, B6, B7	4000	2800	600	600
2	B4, B5, B6	4000	2800	600	600

### 4.3 Segmentação Semântica com Redes Convolucionais

O objetivo dessa seção é apresentar a metodologia utilizada para classificação do uso do solo com redes convolucionais. No treinamento das redes CNNs para a classificação do uso do solo foram combinados 3 métodos de otimização (SGDM, ADAM e RMSProp) com 4 métodos para a melhoria da generalização (sem nenhuma técnica, regularização  $L_2$ , *dropout* e regularização  $L_2 + dropout$ ). Essas combinações, mostradas Bloco 2 da Figura 35, são utilizadas para o treinamento de cada uma das três arquiteturas de CNNs propostas nesse trabalho e que serão mostradas na seção a seguir. Cada arquitetura é selecionada através do Bloco 1 da Figura 35. Denomina-se um modelo de treinamento a junção de três elementos: arquitetura + método

de otimização + método para melhoria da generalização. Assim, temos um total de 36 modelos, que resulta da combinação de 3 arquiteturas x três métodos de otimização x 4 métodos para melhoria da generalização.

Nos Blocos 5 e 6 da Figura 35, são realizadas as operações de validação e teste, que ocorrem após o treinamento dos 36 modelos. Escolhe-se o melhor modelo como aquele que apresenta o melhor desempenho no conjunto de validação (Bloco 5). Escolhido o melhor modelo ele é testado utilizando o conjunto de teste, através da técnica de validação cruzada com 5 pastas (Bloco 6).

Através do Bloco 3 seleciona-se uma arquitetura pré-treinada. O treinamento raso da mesma (apenas a última camada de classificação) é feito no Bloco 4. Finalmente, no Bloco 7 é feita a comparação do desempenho do melhor modelo de CNN selecionado no Bloco 6 com a rede pré-treinada resultante do treinamento efetuado no Bloco 4.

Na próxima seção descreveremos as arquiteturas propostas nesse trabalho a serem utilizadas tanto para classificação do uso do solo como para identificação de áreas desmatadas.

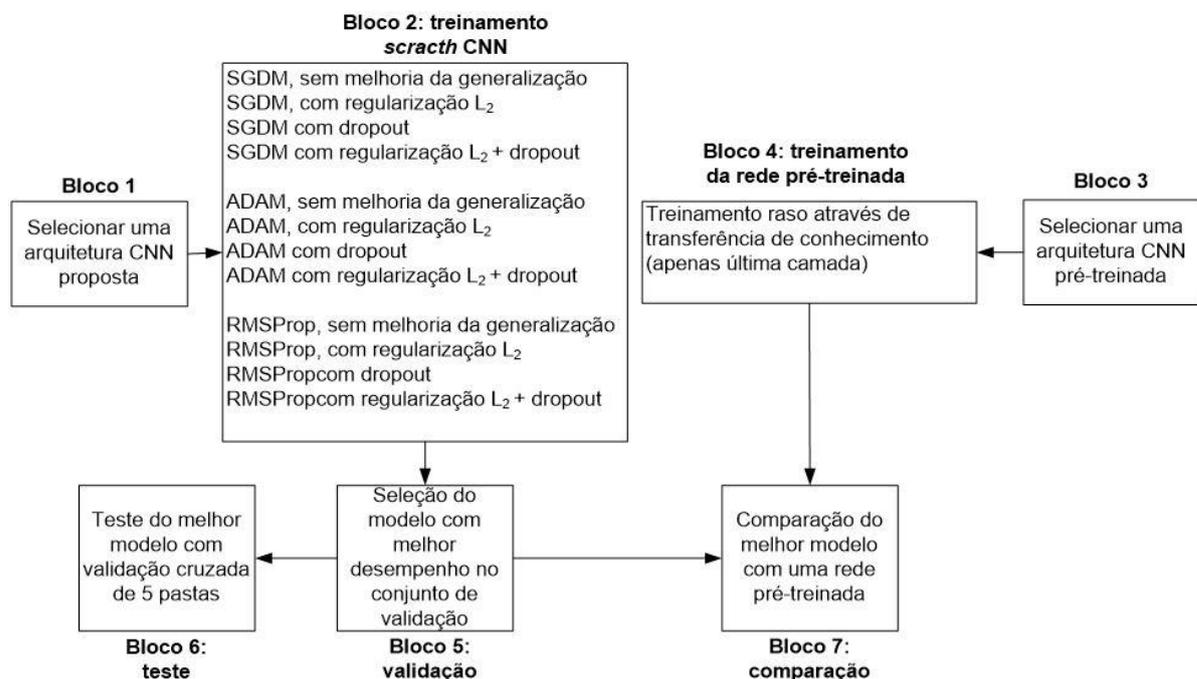


Figura 35 – Etapas da Metodologia de Segmentação Semântica utilizada neste trabalho.

#### 4.3.1 Avaliação de três arquiteturas CNNs

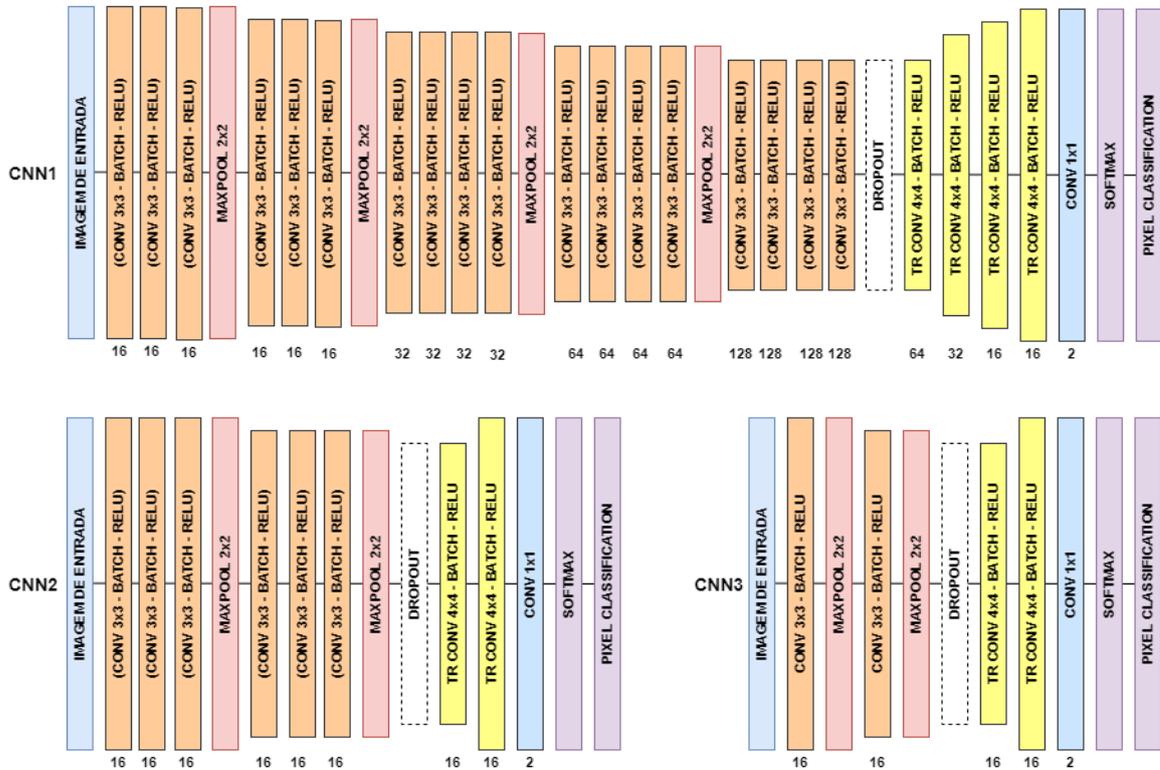
As arquiteturas CNNs utilizadas neste trabalho foram baseadas nas arquiteturas propostas no trabalho de Miyagawa *et al.* (2018). Neste trabalho, os autores tinham como

objetivo realizar a segmentação do lúmen em imagens de tomografia por coerência óptica intravascular (IVOCT). Embora o trabalho de Miyagawa *et al.* (2018) tenha objetivado uma aplicação diversa do trabalho ora apresentado, a escolha das CNNs deveu-se a dois motivos. Em primeiro lugar, as CNNs avaliadas pelos autores possuíam diversos tamanhos, com diferentes quantidades de camada convolucionais, o que permite avaliar a presença ou não do *overfitting*. Em segundo lugar, os autores alcançaram excelente desempenho com uma base de dados não muito volumosa, da mesma ordem de grandeza da base de dados utilizada neste trabalho.

Segundo Miyagawa *et al.* (2018), o IVOCT gera centenas de imagens que necessitam ser analisadas manualmente por um profissional com o objetivo de encontrar pontos de interesse que auxiliem no diagnóstico de doenças cardiovasculares. Os autores analisaram três arquiteturas CNNs para realizar a segmentação automática do lúmen. Na primeira arquitetura, a seção de subamostragem é composta por quatro etapas. Antes e depois da primeira etapa de subamostragem, existem três sequências de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU. Após a segunda, terceira e quarta etapas de subamostragem existem quatro sequências. A seção de sobreamostragem é composta de quatro camadas de convolução transposta com tamanho de filtro 4 x 4. A saída dessa rede é constituída por uma camada *softmax* e uma camada de classificação. Em relação a segunda e a terceira arquitetura, entre as camadas de convolução transposta foi inserido um bloco de camada convolucionais com filtro 3 x 3, camada *batch normalization*, camada ReLU e camada convolucionais com filtro 1 x 1. A terceira arquitetura corresponde a uma rede de grafos acíclicos dirigido (DAG, do inglês *Direct Acyclic Graph*). Esta terceira arquitetura possui conexões entre as saídas das camadas ReLU da etapa da subamostragem e as saídas das camadas de convolução transposta da etapa da sobreamostragem. No trabalho de MIYAGAWA *et al.* (2018), as CNNs foram avaliadas usando três versões do conjunto de dados de imagens. Nesse caso, foram utilizados dois tamanhos de imagem (768x768 pixels e 192x192 pixels) e também sistema de representação de coordenadas cartesianas e polar. Os melhores resultados para acurácia, valor de Dice e de Jaccard ficaram acima de 99%, 98% e 97%, respectivamente e foram obtidos com as imagens de menor tamanho representadas pelo sistema de coordenadas polares.

A Figura 36 apresenta as três arquiteturas CNNs (CNN1, CNN2 e CNN3) utilizadas neste trabalho para aplicações de sensoriamento remoto. Essas arquiteturas foram baseadas na primeira arquitetura do trabalho de MIYAGAWA *et al.* (2018). A camada de *dropout* foi

ilustrada com um retângulo tracejado pois ela é utilizada apenas nas simulações que utilizam essa técnica para melhoria da generalização.



**Figura 36 – Arquiteturas CNNs utilizadas neste trabalho. Os números abaixo de algumas camadas correspondem ao número de filtros. Este número representa o número de neurônios na camada convolucional que se ligam à mesma região na entrada.**

Conforme mostrado na Figura 36, a CNN1 possui quatro etapas de subamostragem (*maxpooling*). Antes e depois da primeira etapa de subamostragem, existem três sequências de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU. Após a segunda, terceira e quarta etapas de subamostragem existem quatro sequências. Além disso, a CNN1 é composta por quatro etapas de sobreamostragem.

A CNN2 possui duas etapas de subamostragem (*maxpooling*) e duas de sobreamostragem. Antes de cada subamostragem, existem três sequências de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU.

Por fim, a CNN3 da mesma forma que a CNN2, também possui duas etapas de subamostragem (*maxpooling*) e duas de sobreamostragem. No entanto, antes de cada subamostragem existem apenas uma sequência de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU.

O motivo de se avaliar três arquiteturas, conforme já dito, é permitir aquilatar a presença de *overffiting*. O *overfitting* é observado quando se tem um grande número de parâmetros numa

arquitetura de CNN e uma pequena quantidade de dados para treinamento. Pode-se afirmar que esse fenômeno está presente quando se tem um bom desempenho da CNN no conjunto de treinamento, mas um desempenho ruim nos conjuntos de validação e teste. A priori, não se pode afirmar que o *overfitting* estará presente em qualquer uma das 3 arquiteturas utilizadas. Porém, existe uma probabilidade maior que ele esteja presente na arquitetura maior, que apresenta uma maior quantidade de parâmetros a serem treinados.

Como dito anteriormente ao se comentar a Figura 35, cada uma das arquiteturas mostradas na Figura 36 foram avaliadas com 12 combinações de métodos de otimização e de técnicas para melhoria da generalização, mostrados no Bloco 1. É mister frisar na conclusão dessa seção que esse conjunto de 12 simulações de cada CNN foi realizada com o conjunto de 3 bandas de entrada e com o conjunto de 6 bandas de entrada. Assim, para cada CNN foram avaliados 24 modelos, e, para as três CNNs, 72 modelos. A seleção do melhor modelo é feita no Bloco 5 da Figura 35.

No processo de treinamento das CNNs é necessário realizar o ajuste de alguns parâmetros. Na Tabela 8 são mostrados os valores dos parâmetros utilizados para o treinamento das CNNs. Esses valores foram ajustados de maneira experimental. Destacamos que, para evitar oscilações em torno de um ponto de mínimo, a taxa de aprendizado decresce de 0,5 a cada 40 de épocas.

**Tabela 8 – Parâmetros de Treinamento das CNNs**

Parâmetros	Valor
Taxa de aprendizado inicial	0,001
Fator de queda de taxa de aprendizagem	0,5
Número de épocas	200
Tamanho do lote	2
Parâmetro da camada <i>dropout</i>	0,3
Fator de regularização	0,001

#### 4.3.2 Treinamento de redes pré-treinadas utilizando transferência de conhecimento

Conforme mostrado no Bloco 7 da Figura 34, o desempenho das CNNs treinadas a partir do *scratch*, que significa um treinamento feito a partir de parâmetros inicializados aleatoriamente, foi comparado com os desempenhos de redes pré-treinadas. O treinamento dessas redes, feito no Bloco 4 da Figura 35, utilizou a técnica de transferência de conhecimento com treinamento raso (*shallow tuning*) apresentado na Seção 3.6. Por treinamento raso, entende-se um treinamento em que todos os parâmetros da rede são congelados, à exceção dos

parâmetros da última camada, que sofrem um ajuste para se adaptar à tarefa da aplicação em mão. As redes pré-treinadas utilizadas neste trabalho foram ResNet50, InceptionResnetv2, MobileNetv2 e Xception, combinadas com métodos de otimização SGDM, ADAM e RMSProp. Essas redes são estruturas pré-treinadas em imagem RGB e necessitam ser adaptadas para lidar com imagens de satélite com múltiplas bandas. O conjunto de dados de treinamento utilizado foram imagens mosaico de seis bandas. Neste caso, a primeira e a última camada de cada rede pré-treinada precisou ser adaptada. A primeira camada foi alterada para aceitar como entrada uma imagem de tamanho 400x400x6. Por fim, a última camada foi alterada de forma que a saída da rede passasse a ser uma imagem de tamanho 400x400 *pixels*.

Um total de doze experimentos foram realizados. A escolha dessas redes deveu-se ao fato de serem arquiteturas propostas nos últimos anos, com ótimo desempenho no desafio “*ImageNet Large Scale Visual Recognition Competition (ILSVRC)*” (RUSSAKOVSKY *et al.*, 2015).

### **4.3.3 Validação cruzada**

Com o objetivo de avaliar a capacidade de generalização do modelo CNN, a partir de um conjunto de dados, foi empregada a técnica de validação cruzada, conforme mostrado no Bloco 6 da Figura 34. Para cada versão do conjunto de dados (3 ou seis bandas), selecionou-se, através do Bloco 5 da Figura 35, o modelo com melhor desempenho no conjunto de validação. Nesse modelo é que foi aplicada a metodologia de validação cruzada de 5 pastas que passamos a descrever. As pastas foram montadas com as imagens dos conjuntos de treinamento e teste. Conforme apresentado na Tabela 7, foram gerados 2800 imagens-mosaico para treinamento e 600 para teste. Dessa maneira, foram utilizados 3400 imagens-mosaico. Essas imagens foram divididas aleatoriamente em 5 pastas com 680 imagens cada uma, designadas por C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub> e C<sub>5</sub>. Foram então realizados cinco experimentos, conforme detalhado na Tabela 10. Em cada experimento, conjuntos de treinamento e testes diferentes são montados a partir das pastas C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub> e C<sub>5</sub>. Cada experimento tem o seu desempenho avaliado e um desempenho médio, mais confiável do ponto de vista estatístico, é obtido através de uma média dos desempenhos de cada experimento.

**Tabela 9 – Experimentos da metodologia de validação cruzada com 5-pastas.**

Experimento	Dados de Treinamento	Dados de Teste
1	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub>	C <sub>5</sub>
2	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub> , C <sub>5</sub>	C <sub>4</sub>
3	C <sub>1</sub> , C <sub>2</sub> , C <sub>5</sub> , C <sub>4</sub>	C <sub>3</sub>
4	C <sub>1</sub> , C <sub>5</sub> , C <sub>3</sub> , C <sub>4</sub>	C <sub>2</sub>
5	C <sub>5</sub> , C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub>	C <sub>1</sub>

#### 4.3.4 Métricas

Várias métricas têm sido utilizadas para avaliar o desempenho de métodos de segmentação. As mais comuns incluem taxas de sucesso e métricas de similaridade. Essas métricas comparam a consistência entre o padrão-ouro e a segmentação realizada pelo método de aprendizagem de máquina (KHOOSHABI, 2013).

Neste trabalho, foram utilizadas as seguintes métricas: a Acurácia Global (ACCG), Acurácia Média (ACCM), coeficiente de similaridade de Jaccard (J), coeficiente de similaridade Jaccard ponderado (JP) e score F1 (F1) (MATHWORKS, 2017). A seguir apresentam-se as definições dessas métricas.

Nas definições a seguir, G é o conjunto de pixels que pertence ao objeto segmentado manualmente pelos especialistas, chamado de padrão-ouro. A é o conjunto de pixels que pertence ao objeto segmentado por um método de segmentação automático. Quando um pixel pertence ao objeto de interesse, atribui-se o valor 1 ao mesmo. Caso contrário, atribui-se o valor 0 ao mesmo. O resultado da classificação de um pixel  $i$  é considerado como:

- Verdadeiro positivo (TP, do inglês *True Positive*) quando seu valor no método automático  $A_i$  e no padrão ouro  $G_i$  são iguais a 1.
- Falso positivo (FP, do inglês *False Positive*) quando  $A_i$  é 1 e  $G_i$  é 0.
- Falso negativo (FN, do inglês *False Negative*) quando  $A_i$  é 0 e  $G_i$  é 1.
- Verdadeiro negativo (TN, do inglês *True Negative*) quando  $A_i$  e  $G_i$  são iguais a 0.

A sensibilidade, especificidade, precisão e acurácia são definidas como:

$$\text{Sensibilidade} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Especificidade} = \frac{TN}{TN+FP} \quad (6)$$

$$\text{Precisão} = \frac{TP}{TP+FP} \quad (7)$$

$$\text{Acurácia} = \frac{TP+TN}{TP+FN+FP+TN} \quad (8)$$

A Acurácia Média indica a porcentagem de pixels identificados corretamente para cada classe. A Acurácia Global é a proporção de pixels classificados corretamente, independentemente da classe, para o número total de pixels.

O coeficiente Jaccard proposto em Jaccard (1912) é uma métrica de similaridade que mede quão bem duas segmentações se sobrepõem. O coeficiente Jaccard é uma medida estatística que penaliza os falsos positivos. É definido como a interseção entre duas regiões rotuladas similarmente,  $r$ , sobre suas uniões:

$$\sum Jaccard = \frac{(|Gr \cap Ar|)}{(|Gr \cup Ar|)} = \frac{(|TP|)}{((|TP|+|FP|+|FN|))} \quad (9)$$

O coeficiente Jaccard Ponderado é uma métrica que representa uma média ponderada do valor de Jaccard para cada classe. Nesse caso, os pesos utilizados nessa média são os números de pixels de cada classe.

Por fim, o F1-Score pode ser calculado por meio da combinação das métricas de precisão e sensibilidade. O F1-Score tem como objetivo indicar a qualidade geral do modelo e estabelece um equilíbrio entre a precisão e a sensibilidade.

$$F1 - Score = 2x \frac{Precisão \times Sensibilidade}{Precisão + Sensibilidade} \quad (10)$$

Para selecionar o melhor modelo CNN do conjunto de validação foi utilizada a métrica acurácia global.

#### 4.4 Segmentação de Regiões Homogêneas

Neste trabalho, as CNNs avaliadas foram treinadas utilizando imagens mosaico compostas de retalhos de 40 x 40 *pixels*. Observamos que as redes treinadas com esses retalhos, tendem a gerar, na saída, imagens recortadas com retalhos do mesmo tamanho. Para neutralizar esta tendência e melhorar a classificação/segmentação, foi utilizado um processo de fusão de dados que associou a saída da CNN e um algoritmo de segmentação de regiões homogêneas de multi-resolução.

A segmentação de regiões homogêneas consiste em criar uma imagem com regiões homogêneas, utilizando como base o valor do pixel da imagem. Neste trabalho, foi utilizado o algoritmo conhecido como “segmentação multiresolução”, proposto por Baatz e Schape (2000). Esse algoritmo realiza uma segmentação geral baseada em definições de homogeneidade, combinada com técnicas de otimização local e global. Um parâmetro de escala é usado para controlar o tamanho médio do objeto de imagem. Segundo os autores, a segmentação multiresolução é uma técnica de fusão de regiões começando com objetos de um pixel. Em

várias etapas subsequentes, objetos de imagem menores são mesclados em objetos maiores. Ao longo desse processo de agrupamento por pares, o procedimento de otimização subjacente minimiza a heterogeneidade dos objetos segmentados na imagem resultante. Em cada etapa, esse par de objetos adjacentes semelhantes são mesclados, o que resulta no menor crescimento da heterogeneidade definida. Se o menor crescimento exceder o limite definido pelo parâmetro de escala, o processo para. Dessa maneira, podemos afirmar que a segmentação multiresolução é um procedimento de otimização local. Neste trabalho, o uso deste algoritmo foi realizado através do *Software e-Cognition*. O fator de escala utilizado foi 200, ajustado experimentalmente.

#### 4.5 Fusão de Dados

A fusão de dados consiste em compor informações de segmentação provenientes do algoritmo de segmentação multiresolução com informações de segmentação provenientes da segmentação semântica realizada pela rede CNN. A fusão de dados utilizada neste trabalho é ilustrada por meio do fluxograma apresentado na Figura 37. Para cada região gerada pelo algoritmo de segmentação multiresolução, avaliamos a classificação dos pixels das mesmas realizada pela CNN. A classe de pixels mais frequente em uma região, resultante da classificação realizada pela CNN, é considerada a classe daquela região. A Figura 38 ilustra este processo: A Figura 38(a) mostra uma imagem original; a Figura 38(b) mostra os resultados do algoritmo de região homogênea aplicado na Figura 38(a). Na Figura 38(b), dentro de um retângulo preto, são selecionadas três pequenas regiões vizinhas, que são ampliadas no lado direito. Estas regiões selecionadas são preenchidas com uma classificação hipotética de pixels pela CNN. A região superior tem mais *pixels* brancos, portanto, é classificada como floresta. A região central tem mais *pixels* pretos, portanto é classificada como agricultura, e a região inferior tem mais *pixels* cinzas, portanto é classificada como pasto.

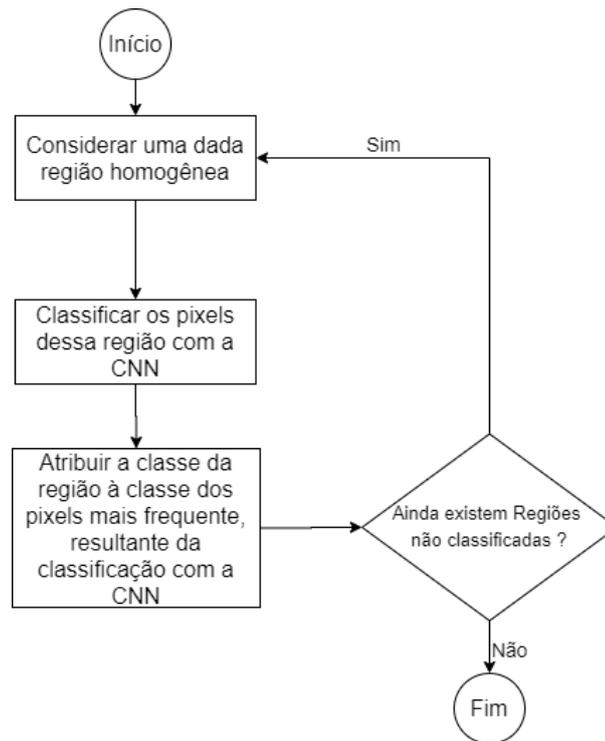


Figura 37 – Fluxograma do processo de refinamento da classificação da imagem.

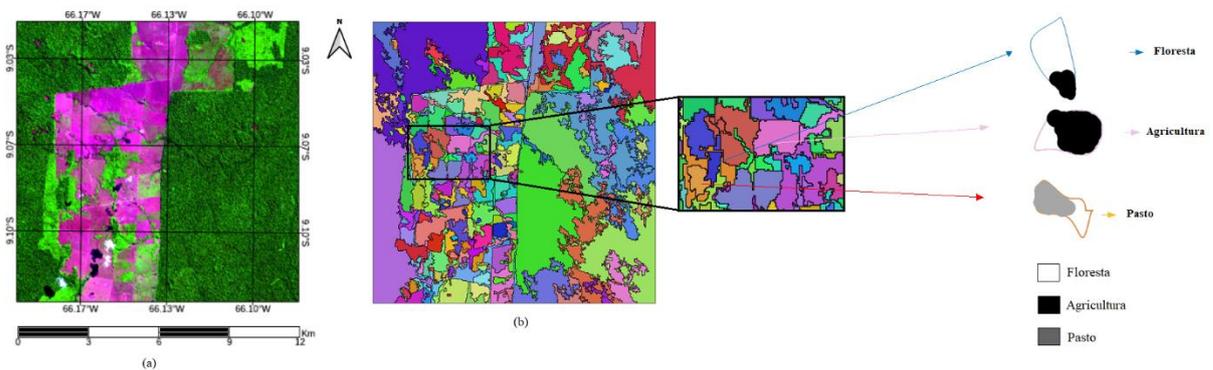


Figura 38 – Ilustração do processo de fusão: (a) imagem original; (b) imagem resultante do algoritmo de segmentação de região homogênea. Dentro de pequenos retângulos, três pequenas regiões são selecionadas e ampliadas. A região inferior tem mais pixels cinza, por isso é classificada como pasto. A região central tem mais pixels pretos, portanto é classificada como agricultura, e como há mais pixels brancos na região superior, é classificada como floresta.

#### 4.6 Ambiente de Desenvolvimento

Neste trabalho foram utilizados os seguintes *softwares*: MATLAB ® 2019<sup>a</sup>, ENVI 5.5 e e-Cognition.

O *software* ENVI foi utilizado para gerar as imagens de referência para o classificador. O *software* MATLAB foi empregado no desenvolvimento do classificador da rede. O *software* e-Cognition foi empregado para utilizar o algoritmo de segmentação multi-resolução proposto por Baatz e Schäpe (2000).

A infraestrutura de *hardware* utilizada para o desenvolvimento deste trabalho foi um computador *desktop* com características descritas na Tabela 10. Destacamos, nessa configuração, a quantidade de memória da GPU, 32GB, que possibilitou realizar as 72 simulações em um tempo aceitável. Em média, cada simulação demorou em torno de 80 horas.

**Tabela 10 – Características do computador utilizado**

<b>Sistema Operacional</b>	<i>Windows 10</i>
<b>Processador</b>	Intel ® Xeon® W-2145 3,70 GHz
<b>GPU</b>	Quadro GV100 32GB e 5120 núcleos CUDA
<b>Memória RAM</b>	64,0 GB
<b>Tipo de Sistema</b>	Sistema Operacional de 64 Bits

## CONSIDERAÇÕES FINAIS

Esse capítulo definiu com muita clareza a metodologia proposta nesse trabalho para classificação do uso do solo a partir de um conjunto de bandas obtidas a partir de imagens do satélite LANDSAT-8/OLI. A não ser pela Seção 2, que apresenta dados específicos para aplicações de classificação do uso do solo, o conteúdo das outras seções é genérico e pode ser utilizado em aplicações de detecção de regiões desmatadas. Particularmente, as etapas mostradas no diagrama em blocos da Figura 30 foram utilizadas no próximo capítulo para detecção de regiões desmatadas. O capítulo que ora finda caracteriza-se por propor uma metodologia robusta para a avaliação do método de fusão de dados em sensoriamento remoto, contribuição maior dessa tese. A robustez é conferida pelo método de validação cruzada, pela proposta de vários tamanhos de retalho de imagens, pelas inúmeras métricas a serem utilizadas para avaliação do desempenho, pelos diversos métodos de otimização e para melhoria da generalização das CNNs e pelas diversas arquiteturas de CNN propostas. No capítulo que se segue ficará claro que só as redes convolucionais não foram suficientes para garantir desempenhos que consideramos satisfatórios e, portanto, a fusão de dados foi imprescindível para se obter um desempenho no estado da arte.

---

## 5 RESULTADOS E DISCUSSÕES

Esse Capítulo pode ser dividido em duas partes. Na primeira parte, é realizada a caracterização da metodologia proposta utilizando os conjuntos de imagens mosaico de três e seis bandas, com as classes floresta, agricultura e pasto. Essa primeira parte abrange as Seções 5.1, 5.2 e 5.3. Na segunda parte deste capítulo, é realizada uma avaliação da metodologia proposta, utilizando imagens reais de regiões da Amazônia Legal Brasileira. Essa segunda parte abrange as Seções 5.4 e 5.5.

Na Seção 5.1 são apresentados os resultados das 36 simulações realizadas utilizando as imagens mosaico com seis bandas, as três CNNs descritas em 4.3.1, com os três métodos de otimização e com os dois métodos para melhoria da generalização. Ao final, são apresentados os resultados do modelo com melhor desempenho utilizando uma metodologia de validação cruzada de 5-pastas.

Na Seção 5.2 são apresentadas outras 36 simulações que utilizaram como dado de entrada imagens-mosaico de três bandas. Ao final, são apresentados os resultados do modelo com melhor desempenho utilizando uma metodologia de validação cruzada de 5-pastas.

Na Seção 5.3 são apresentados os resultados para a classificação das imagens-mosaico utilizando redes pré-treinadas por meio da técnica de transferência de conhecimento.

Na Seção 5.4 são apresentados os resultados da segmentação semântica de cinco imagens que correspondem a regiões da Amazônia Legal Brasileira. Inicialmente são apresentados os resultados utilizando o processo de fusão de dados. Ao final, é apresentada uma comparação entre a segmentação semântica utilizando o processo de fusão de dados e utilizando apenas modelos CNNs treinados com imagens mosaico geradas a partir de retalhos de imagens de diferentes tamanhos, 10x10 e 20x20 *pixels*.

Na Seção 5.5 são apresentados os resultados da aplicação da metodologia nas mesmas regiões avaliadas por Adarme *et al.* (2020) e Bem *et al.* (2020). Esses trabalhos não tiveram como objetivo realizar a classificação de uso de solo na Amazônia Legal, mas sim a detecção de áreas desmatadas. Nesse caso, a metodologia proposta no capítulo anterior foi adaptada, nos seguintes aspectos: foram utilizadas as classes de desmatamento e não desmatamento, as mesmas empregadas por Adarme *et al.* (2020) e Bem *et al.* (2020) e foram realizados novos treinamentos das redes CNNs, porém as arquiteturas utilizadas foram as mesmas. Quanto ao

algoritmo de regiões homogêneas, não houve modificações em relação àquele utilizado para a caracterização do uso do solo.

### 5.1 Resultados das Simulações (Imagens com seis bandas)

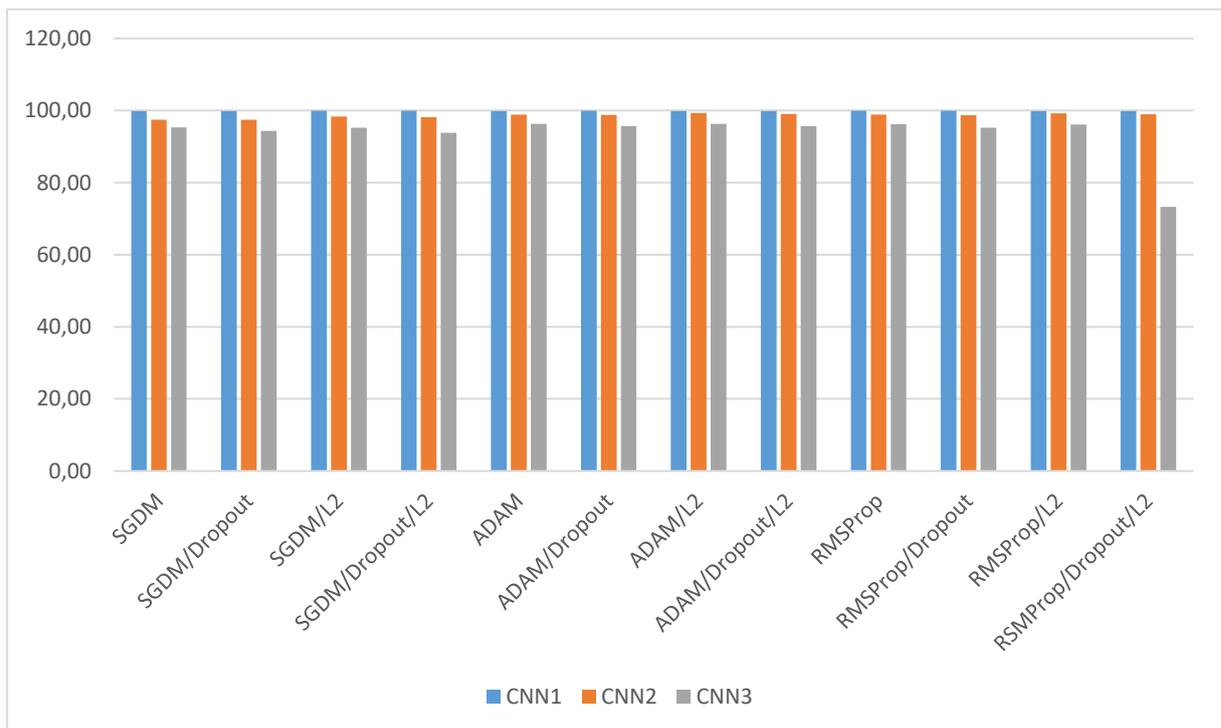
Esta seção apresenta os valores das métricas de desempenho das três arquiteturas CNNs propostas, combinadas com cada um dos métodos de otimização e com cada um dos métodos para melhoria da generalização. Como entrada das redes, utilizou-se imagens-mosaico de seis bandas. A Tabela 11 apresenta os desempenhos no conjunto de validação obtidos utilizando as redes CNN1, CNN2 e CNN3. A Figura 39 apresenta um gráfico com o desempenho de todos os experimentos realizados utilizando as três CNNs, os métodos de otimização e generalização. Os modelos de cada CNN com melhor desempenho no conjunto de validação são resumidos na Tabela 12.

**Tabela 11 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, com imagens de 6 bandas.**

	<b>Experimento</b>	<b>ACCG<sup>(1)</sup> (%)</b>	<b>ACCM<sup>(2)</sup> (%)</b>	<b>J<sup>(3)</sup> (%)</b>	<b>JP<sup>(4)</sup> (%)</b>	<b>F1<sup>(5)</sup> (%)</b>
1	CNN1/SGDM	99,83	99,83	99,65	99,65	99,45
2	CNN1/SGDM/Dropout	99,87	99,87	99,75	99,75	99,64
3	CNN1/SGDM/L2	99,92	99,92	99,85	99,85	99,82
4	CNN1/SGDM/Dropout/L2	99,94	99,94	99,88	99,88	99,85
5	CNN1/ADAM	99,87	99,87	99,75	99,75	99,70
6	CNN1/ADAM/Dropout	99,94	99,94	99,89	99,89	99,91
7	CNN1/ADAM/L2	99,84	99,84	99,68	99,68	99,76
8	CNN1/ADAM/Dropout/L2	99,84	99,84	99,68	99,68	99,76
9	CNN1/RMSProp	99,98	99,98	99,96	99,96	99,95
10	CNN1/RMSProp/Dropout	99,96	99,96	99,93	99,93	99,91
11	CNN1/RMSProp/L2	99,89	99,89	99,78	99,78	99,84
12	CNN1/RMSProp/Dropout/L2	99,90	99,90	99,81	99,81	99,85
13	CNN2/SGDM	97,52	97,53	95,21	95,20	91,30
14	CNN2/SGDM/Dropout	97,48	97,49	95,13	95,12	92,08
15	CNN2/SGDM/L2	98,33	98,33	96,73	96,72	93,29
16	CNN2/SGDM/Dropout/L2	98,20	98,20	96,49	96,48	93,60
17	CNN2/ADAM	98,91	98,92	97,86	97,86	94,64
18	CNN2/ADAM/Dropout	98,79	98,79	97,63	97,62	94,63
19	CNN2/ADAM/L2	99,36	99,36	98,72	98,72	96,47
20	CNN2/ADAM/Dropout/L2	99,11	99,11	98,24	98,24	95,76
21	CNN2/RMSProp	98,90	98,90	97,83	97,83	94,50
22	CNN2/RMSProp/Dropout	98,75	98,75	97,54	97,54	94,50
23	CNN2/RMSProp/L2	99,23	99,23	98,47	98,47	95,92
24	CNN2/RMSProp/Dropout/L2	99,02	99,02	98,06	98,06	95,39

	Experimento	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
25	CNN3/SGDM	95,38	95,40	91,33	91,31	86,01
26	CNN3/SGDM/Dropout	94,40	94,41	89,63	89,60	86,44
27	CNN3/SGDM/L2	95,29	95,31	91,17	91,15	86,00
28	CNN3/SGDM/Dropout/L2	93,88	93,88	88,73	88,70	82,67
29	CNN3/ADAM	96,32	96,34	93,02	93,00	86,73
30	CNN3/ADAM/Dropout	95,70	95,71	91,88	91,86	87,98
31	CNN3/ADAM/L2	96,30	96,31	92,97	92,95	87,27
32	CNN3/ADAM/Dropout/L2	95,73	95,75	91,96	91,94	88,13
33	CNN3/RMSProp	96,20	96,21	92,80	92,78	86,59
34	CNN3/RMSProp/Dropout	95,30	95,31	91,19	91,17	87,19
35	CNN3/RMSProp/L2	96,17	96,18	92,74	92,72	86,51
36	CNN3/RMSProp/Dropout/L2	73,32	73,24	58,55	58,54	50,24

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado, <sup>(5)</sup> F1 = Score F1



**Figura 39 – Gráfico com o desempenho (acurácia) de cada modelo CNN proposto neste trabalho no conjunto de validação, com imagens de 6 bandas.**

**Tabela 12 – Desempenho dos melhores modelos de cada CNN no conjunto de validação, com imagens de 6 bandas.**

	Experimento	Acurácia (%)
1	CNN1/RMSProp	99,98
2	CNN2/ADAM/L2	99,36
3	CNN3/ADAM	96,32

Conforme mostrado na Tabela 12, o modelo com melhor desempenho foi aquele no qual foi utilizada a CNN1, utilizando o método otimização RMSProp, sem utilizar métodos de regularização e métodos para melhoria da generalização. Esse modelo foi avaliado através do método de validação cruzada com 5-pastas. O conjunto de dados utilizado para esse experimento foi o conjunto de treinamento e teste. Dessa maneira, foram utilizados 3400 imagens-mosaico de seis bandas. Essas imagens foram divididas aleatoriamente em 5 conjuntos com 680 imagens. A Tabela 13 apresenta os desempenhos obtidos para esse experimento.

**Tabela 13 – Desempenho do melhor modelo CNN1/RMSProp utilizando a metodologia de validação cruzada com 5-pastas e imagens de 6 bandas.**

Pasta	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
1	100,00	100,00	100,00	100,00	99,99
2	99,99	99,99	99,97	99,97	99,97
3	99,89	99,89	99,78	99,79	99,78
4	99,99	99,99	99,99	99,99	99,99
5	99,99	99,99	99,97	99,97	99,97
<b>Média</b>	<b>99,97</b>	<b>99,97</b>	<b>99,94</b>	<b>99,94</b>	<b>99,94</b>

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado, <sup>(5)</sup> F1 = Score F1

## 5.2 Resultados das Simulações (Imagens com três bandas)

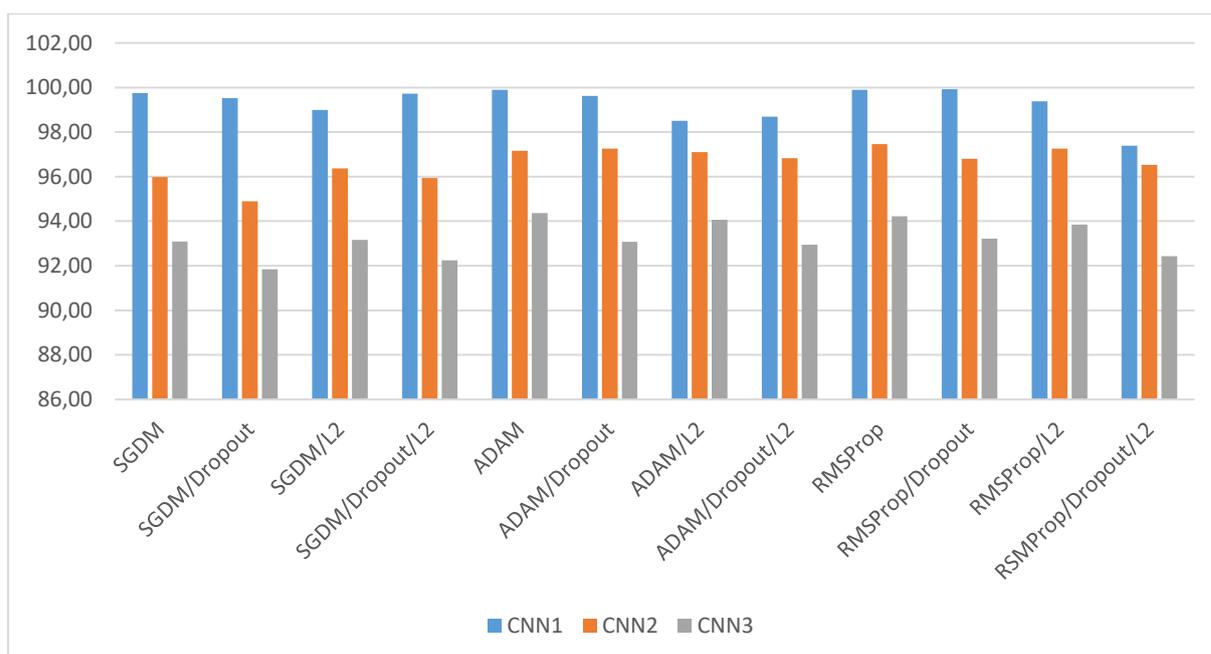
Esta seção apresenta as métricas de desempenho das três arquiteturas CNNs propostas, combinadas com cada um dos métodos de otimização e com cada um dos métodos para melhoria da generalização, utilizando como entrada imagens-mosaico de três bandas (B4, B5 e B6). A Tabela 14 apresenta os desempenhos obtidos utilizando as redes CNN1, CNN2 e CNN3. A Figura 40 apresenta um gráfico com o desempenho de todos os experimentos realizados utilizando as três CNNs, com os métodos de otimização e com os métodos para melhoria da generalização. Os modelos de cada CNN com melhor desempenho no conjunto de validação são resumidos na Tabela 15.

**Tabela 14 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, com imagens de 3 bandas.**

	Experimento	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
1	CNN1/SGDM	99,75	99,75	99,50	99,50	99,30
2	CNN1/SGDM/Dropout	99,52	99,53	99,06	99,05	98,84
3	CNN1/SGDM/L2	99,00	99,00	98,02	98,02	97,95
4	CNN1/SGDM/Dropout/L2	99,72	99,73	99,45	99,45	99,46
5	CNN1/ADAM	99,89	99,89	99,78	99,78	99,77
6	CNN1/ADAM/Dropout	99,63	99,63	99,26	99,26	99,40
7	CNN1/ADAM/L2	98,51	98,52	97,08	97,08	97,56

	<b>Experimento</b>	<b>ACCG<sup>(1)</sup> (%)</b>	<b>ACCM<sup>(2)</sup> (%)</b>	<b>J<sup>(3)</sup> (%)</b>	<b>JP<sup>(4)</sup> (%)</b>	<b>F1<sup>(5)</sup> (%)</b>
8	CNN1/ADAM/Dropout/L2	98,70	98,70	97,44	97,43	97,92
9	CNN1/RMSProp	99,89	99,89	99,79	99,79	99,78
10	CNN1/RMSProp/Dropout	99,92	99,92	99,84	99,84	99,77
11	CNN1/RMSProp/L2	99,38	99,38	98,76	98,76	98,74
12	CNN1/RMSProp/Dropout/L2	97,38	97,40	94,93	94,92	95,93
13	CNN2/SGDM	95,98	95,99	92,41	92,39	86,85
14	CNN2/SGDM/Dropout	94,90	94,92	90,49	90,46	87,11
15	CNN2/SGDM/L2	96,37	96,38	93,09	93,07	87,98
16	CNN2/SGDM/Dropout/L2	95,95	95,96	92,33	92,31	89,08
17	CNN2/ADAM	97,16	97,17	94,54	94,52	89,43
18	CNN2/ADAM/Dropout	97,27	97,27	94,74	94,73	89,83
19	CNN2/ADAM/L2	97,09	97,10	94,42	94,41	89,29
20	CNN2/ADAM/Dropout/L2	96,82	96,84	93,91	93,89	90,23
21	CNN2/RMSProp	97,47	97,48	95,11	95,10	89,52
22	CNN2/RMSProp/Dropout	96,81	96,81	93,88	93,87	88,31
23	CNN2/RMSProp/L2	97,26	97,27	94,73	94,72	90,25
24	CNN2/RMSProp/Dropout/L2	96,53	96,55	93,39	93,37	90,27
25	CNN3/SGDM	93,09	93,12	87,41	87,38	81,95
26	CNN3/SGDM/Dropout	91,84	91,86	85,35	85,32	82,74
27	CNN3/SGDM/L2	93,16	93,18	87,52	87,49	82,07
28	CNN3/SGDM/Dropout/L2	92,25	92,28	86,01	85,97	83,94
29	CNN3/ADAM	94,37	94,38	89,57	89,54	81,95
30	CNN3/ADAM/Dropout	93,08	93,10	87,37	87,34	83,36
31	CNN3/ADAM/L2	94,06	94,08	89,03	89,00	82,92
32	CNN3/ADAM/Dropout/L2	92,94	92,97	87,15	87,12	84,47
33	CNN3/RMSProp	94,22	94,24	89,32	89,30	81,63
34	CNN3/RMSProp/Dropout	93,22	93,24	87,61	87,58	83,72
35	CNN3/RMSProp/L2	93,85	93,87	88,66	88,64	83,21
36	CNN3/RMSProp/Dropout/L2	92,43	92,45	86,32	86,29	81,60

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado, <sup>(5)</sup> F1 = Score F1



**Figura 40 – Gráfico com o desempenho (acurácia) de cada modelo CNN proposto neste trabalho no conjunto de validação, com imagens de 3 bandas**

**Tabela 15 – Desempenho dos melhores modelos de cada CNN, no conjunto de validação, com imagens de 3 bandas.**

	Experimento	Acurácia (%)
1	CNN1/RMSProp/Dropout	99,92
2	CNN2/RMSProp	97,47
3	CNN3/ADAM	94,37

Conforme mostrado na Tabela 15, o modelo com melhor desempenho foi aquele no qual foi utilizada a CNN1, utilizando o método otimização RMSProp e *dropout* como método de melhoria da generalização. Esse modelo foi avaliado através do método de validação cruzada com 5-pastas. O conjunto de dados utilizado para esse experimento foi o conjunto de treinamento e teste. Dessa maneira, foram utilizados 3400 imagens-mosaico de três bandas (B4, B5 e B6). Essas imagens foram divididas aleatoriamente em 5 conjuntos com 680 imagens. A Tabela 16 apresenta os desempenhos obtidos para esse experimento.

**Tabela 16 – Desempenho do modelo CNN1/RMSProp/Dropout utilizando uma metodologia de validação cruzada com 5-pastas, com imagens de três bandas.**

Pasta	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
1	99,86	99,86	99,72	99,72	99,65
2	99,87	99,87	99,74	99,74	99,71
3	99,76	99,76	99,52	99,52	99,59
4	99,87	99,87	99,74	99,74	99,72
5	99,94	99,94	99,89	99,89	99,87
<b>Média</b>	<b>99,86</b>	<b>99,86</b>	<b>99,72</b>	<b>99,72</b>	<b>99,71</b>

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup>JP = Coeficiente Jaccard Ponderado,

<sup>(5)</sup> F1 = Score F1

### 5.3 Resultados das Simulações utilizando transferência de conhecimento

Os resultados obtidos utilizando as arquiteturas CNN1, CNN2 e CNN3 foram comparados com os resultados obtidos com redes treinadas através da metodologia de transferência de conhecimento, conforme detalhada na Seção 4.3.2. Para o treinamento com transferência de conhecimento foram utilizados os métodos de otimização SGDM, ADAM e RMSProp. Os dados de entrada foram imagens-mosaico de seis bandas. Nesse caso, um total de doze experimentos foram realizados. A Tabela 17 apresenta os resultados obtidos utilizando o conjunto de validação. O melhor desempenho foi obtido utilizando a rede Resnet50 e o método ADAM. Para este modelo, foi obtida uma ACCG de 94,50%. Este desempenho foi inferior ao do modelo CNN1 com o método de otimização RMSProp, que obteve uma ACCG de 99,98%.

**Tabela 17 – Acurácia utilizando transferência de conhecimento para o conjunto de validação.**

Experimento	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
ResNet50/SGDM	92,26	92,28	86,03	86,00	86,31
ResNet50/ADAM	94,50	94,51	89,77	89,75	87,73
ResNet50/RMProp	94,17	94,19	89,21	89,18	88,70
InceptionResnetv2/SGDM	89,77	89,79	81,84	81,81	80,44
InceptionResnetv2/ADAM	87,62	87,66	78,14	78,11	77,75
InceptionResnetv2/RMSProp	88,23	88,25	79,38	79,35	79,91
MobileNetv2/SGDM	77,87	77,93	64,00	63,95	57,17
MobileNetv2/ADAM	77,41	77,44	63,80	63,76	56,27
MobileNetv2/RMSProp	77,83	77,86	64,27	64,23	58,92
Xception/SGDM	92,25	92,27	85,96	85,93	82,67
Xception/ADAM	91,34	91,36	84,41	84,37	83,02
Xception/RMSProp	91,18	91,20	84,21	84,17	82,21

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup>JP = Coeficiente Jaccard Ponderado,

<sup>(5)</sup> F1 = Score F1

### 5.4 Classificação/Segmentação de imagens da Região Amazônica

Essa seção apresenta os resultados de classificação utilizando imagens de algumas regiões da Amazônia. Essa seção é dividida em duas partes. Inicialmente apresenta-se os

resultados da segmentação das imagens reais de 400 x 400 *pixels* (e não imagens mosaico) utilizando o processo de fusão de dados. Posteriormente, apresenta-se uma comparação entre a segmentação semântica utilizando o processo de fusão de dados e a segmentação semântica utilizando apenas modelos CNNs treinados com imagens mosaico geradas a partir de retalhos de imagens de tamanho 10x10 e 20x20 *pixels*.

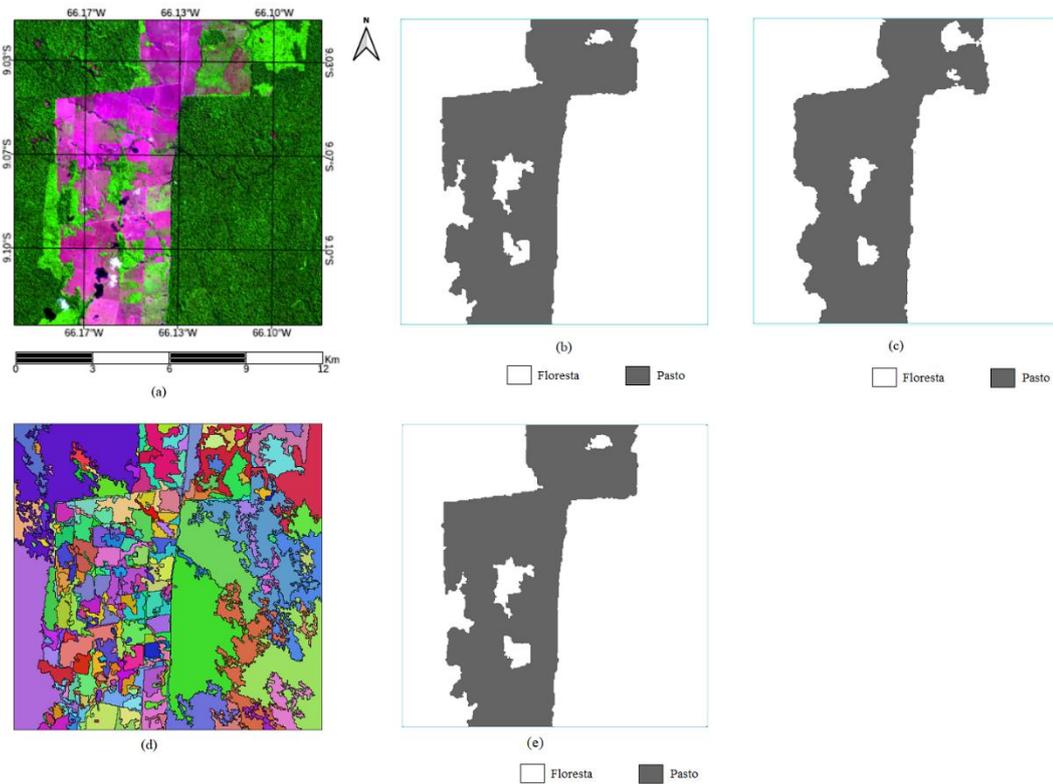
#### **5.4.1 Segmentação de imagens reais (e não imagens mosaico) da Região Amazônica utilizando o método de Fusão de Dados**

Nesse experimento trabalhou-se com imagens de 6 bandas. Foi avaliado o desempenho dos três melhores modelos de CNN1, CNN2 e CNN3 mostrados na Tabela 12. O modelo que apresentou os melhores resultados foi o CNN2/ADAM/L<sub>2</sub>. Em seguida foi realizado um refinamento do resultado obtido através do método de fusão de dados, apresentado na Seção 4.5.

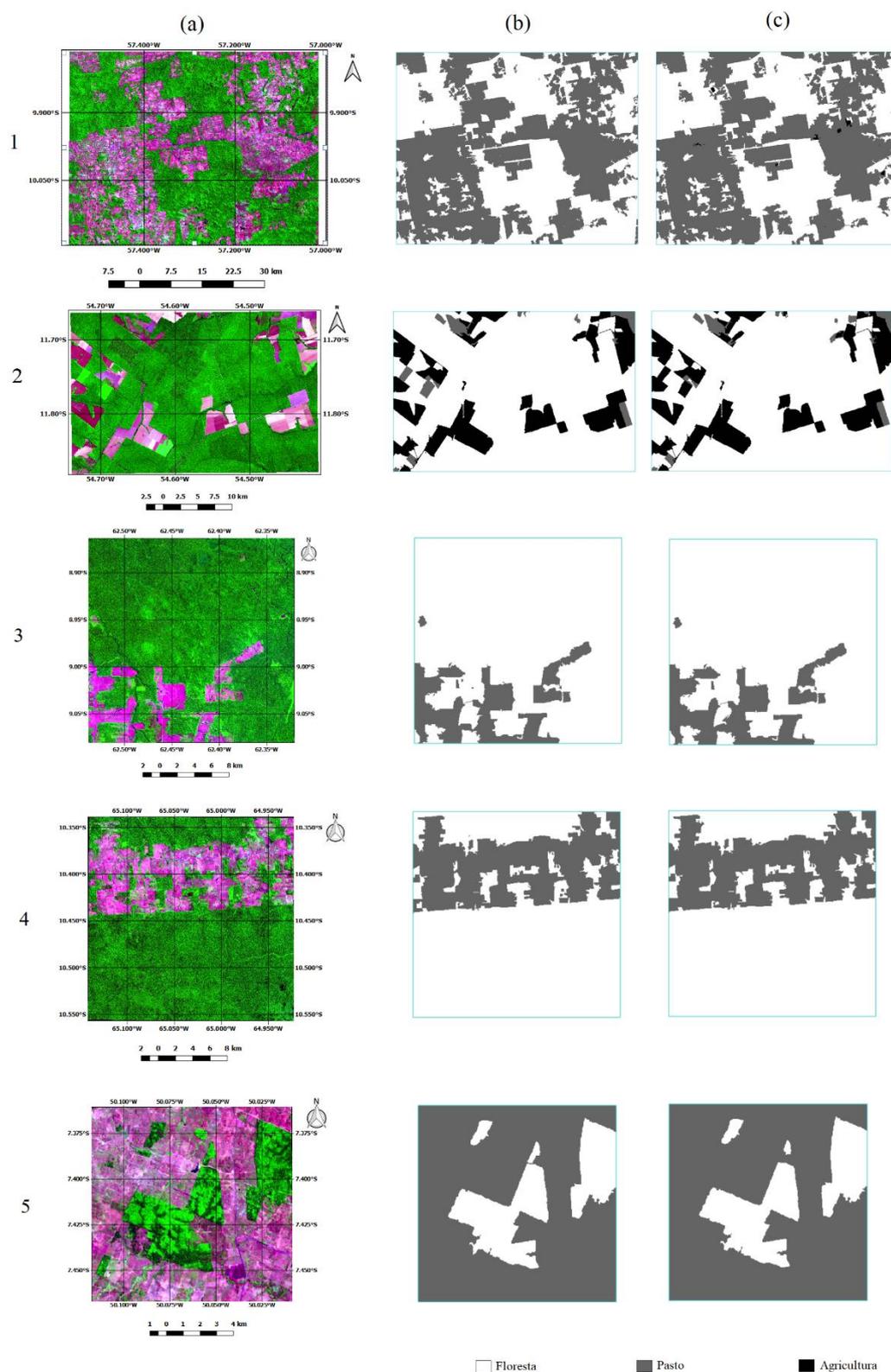
Na Figura 41, apresenta-se os resultados da classificação de uma imagem LANDSAT-8/OLI obtidos através do modelo que utiliza a CNN2/ADAM/L<sub>2</sub>, bem como o resultado utilizando a fusão de dados (modelo CNN2/ADAM/L<sub>2</sub> e segmentação de regiões homogêneas). Como mostrado na Figura 41(c), no resultado da classificação pela CNN2, há algumas regiões classificadas incorretamente como floresta. Na Figura 41(e), que mostra o resultado através do processo de fusão, estas classificações errôneas são minimizadas. A acurácia obtida utilizando apenas a segmentação semântica com CNN2/ADAM/L<sub>2</sub> foi de 96,84%. Após o processo de fusão dos dados, a acurácia aumentou para 99,81%.

Pode-se observar que a metodologia proposta apresentou desempenhos muito próximos ao desempenho máximo de 100% de acurácia, o que pode ser considerado no estado da arte para a área. O processo de fusão de dados aumentou a acurácia da classificação da imagem.

Na Figura 42 mostra-se cinco imagens LANDSAT-8/OLI de regiões da Amazônia Legal Brasileira com seus respectivos padrões ouro e as imagens classificadas usando o processo de fusão de dados (modelo CNN2/ADAM/L<sub>2</sub> e segmentação de regiões homogêneas). A Tabela 18 apresenta os resultados obtidos para a segmentação semântica dessas imagens utilizando somente o modelo CNN2/ADAM/L<sub>2</sub> e o processo de fusão de dados. Pode-se observar que o processo de fusão de dados melhorou os valores de acurácia obtidos para essas regiões, que ficaram acima de 99% em todos os casos.



**Figura 41 – LANDSAT-8/OLI imagens das regiões amazônicas. (a) imagem original exibida em composição colorida das bandas B6-B5- B4; (b) padrão ouro; (c) classificação semântica feita pela CNN2, com uma acurácia de 96,84%. (d) imagem de saída resultante da segmentação multiresolução; (e) resultado da classificação da imagem após processo de fusão de dados, com uma acurácia de 99,80%.**



**Figura 42 – LANDSAT-8/OLI imagens da região Amazônica. (a) imagens originais exibidas em composição colorida B6-B5-B4; (b) imagem padrão em ouro; (c) imagens classificadas pelo modelo CNN2/ADAM/L<sub>2</sub> e processo de fusão. As acurácias obtidas para as imagens 1, 2, 3, 4 e 5 foram de 99,63%, 99,08%, 99,66%, 99,83% e 99,85% respectivamente.**

**Tabela 18 – Resultados obtidos para a segmentação semântica das imagens da Figura 42 utilizando os modelos CNN2/ADAM/L<sub>2</sub> e Fusão de Dados**

Figura	Região/Cena	Data	Método de Segmentação	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
41(1)	228/67	18/08/2014	CNN2/ADAM/L <sub>2</sub>	95,35	95,33	60,99	91,49	94,86
			Fusão de Dados	99,63	99,63	99,33	99,49	99,55
41(2)	228/68	18/08/2014	CNN2/ADAM/L <sub>2</sub>	95,65	84,16	73,96	92,56	83,41
			Fusão de Dados	99,08	99,02	97,62	97,82	93,15
41(3)	231/66	23/08/2014	CNN2/ADAM/L <sub>2</sub>	97,95	96,65	61,53	96,10	94,41
			Fusão de Dados	99,66	99,73	98,64	99,32	97,57
41(4)	233/67	21/08/2014	CNN2/ADAM/L <sub>2</sub>	97,43	97,45	62,69	95,03	93,67
			Fusão de Dados	99,83	99,84	99,59	99,66	98,75
41(5)	224/65	22/08/2014	CNN2/ADAM/L <sub>2</sub>	97,30	94,95	92,18	94,76	87,69
			Fusão de Dados	99,85	99,65	99,55	99,7	99,28

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado,

<sup>(5)</sup> F1 = Score F1

#### 5.4.2 Avaliação da metodologia utilizando retalhos de imagens 20x20 e 10x10 pixels

Esta subsecção apresenta os resultados da segmentação semântica utilizando apenas modelos CNNs treinados com imagens mosaico geradas a partir de retalhos de imagens de tamanho 10x10 e 20x20 pixels. Nesse caso, o objetivo é verificar se o desempenho da segmentação semântica utilizando apenas redes CNNs treinadas com retalhos menores é equivalente ao desempenho utilizando o processo de fusão de dados. Os modelos CNNs escolhidos foram os melhores modelos para o conjunto de validação apresentados na Tabela 12.

A partir das imagens LANDSAT-8/OLI apresentadas na Tabela 5 foram extraídos retalhos de tamanho 20x20 e 10x10 pixels referentes as classes floresta, agricultura e pasto. A Tabela 19 e

Tabela 20 apresentam a distribuição dos retalhos de imagens de tamanho 20x20 e 10x10 pixels, respectivamente.

**Tabela 19 – Divisão dos retalhos de imagens 20x20 pixels**

Classe	Treinamento	Validação
Floresta	260000	260000
Pasto	54000	54000
Agricultura	20000	20000

**Tabela 20 – Divisão dos retalhos de imagens 10x10 pixels**

Classe	Treinamento	Validação
Floresta	637500	637500
Pasto	379500	379500
Agricultura	138000	138000

Utilizando os conjuntos de retalhos de imagens de treinamento e validação descritos anteriormente, foram gerados dois bancos de imagens mosaico com 4.000 imagens cada um, conforme apresentado na Tabela 21. Foram destinadas 3.000 imagens para treinamento e 1.000 para validação. As Tabela 22 e 31 apresentam o desempenho no conjunto de validação com retalhos de 20x20 pixels e 10x10 pixels, respectivamente.

**Tabela 21 – Banco de Imagens-mosaicos gerados com a quantidade para os conjuntos de treinamento e validação**

Versão	Bandas	Total de imagens-mosaico	Treinamento	Validação
20x20	B2, B3, B4, B5, B6, B7	4000	3000	1000
10x10	B2, B3, B4, B5, B6, B7	4000	3000	1000

**Tabela 22 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, treinadas com imagens de 6 bandas e imagens mosaico com retalhos 20x20 pixels**

	Experimento	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
1	CNN1/RMSProp	95,85	95,84	92,13	92,14	95,41
2	CNN2/ADAM/L2	94,55	94,55	89,76	89,77	92,97
3	CNN3/ADAM	94,38	94,37	89,52	89,53	90,85

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado, <sup>(5)</sup> F1 = Score F1

**Tabela 23 – Desempenho das redes CNN1, CNN2 e CNN3 no conjunto de validação, treinadas com imagens de 6 bandas e imagens mosaico com retalhos 10x10 pixels**

	Experimento	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
1	CNN1/RMSProp	93,59	93,59	88,16	88,17	93,17
2	CNN2/ADAM/L2	94,55	94,55	89,76	89,77	92,97
3	CNN3/ADAM	94,38	94,37	89,52	89,53	90,85

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado, <sup>(5)</sup> F1 = Score F1

Utilizando os modelos CNNs apresentados na Tabela 22 e Tabela 23, foi realizada a segmentação semântica das imagens apresentadas na Figura 41 e 41. A Tabela 24 e a Tabela 25 apresentam o desempenho para a classificação dessas imagens utilizando os modelos CNNs treinados com retalhos de tamanho 20x20 e 10x10 pixels respectivamente.

**Tabela 24 – Resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelos CNNs treinados com imagens mosaico compostas de retalhos 20x20 pixels**

Modelo	Figura	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
CNN1/RMSProp	Figura 41 (a)	43,80	43,43	27,20	41,36	47,69
CNN2/ADAM/L <sub>2</sub>	Figura 41 (a)	97,04	97,26	93,57	94,29	84,66
CNN3/ADAM	Figura 41 (a)	<b>97,18</b>	96,78	93,80	94,53	86,04
CNN1/RMSProp	Figura 42 (1)	26,52	26,56	16,24	24,36	86,35
CNN2/ADAM/L <sub>2</sub>	Figura 42 (1)	94,70	94,69	60,28	90,42	93,80
CNN3/ADAM	Figura 42 (1)	<b>94,91</b>	94,96	60,71	91,05	90,95
CNN1/RMSProp	Figura 42 (2)	28,40	43,85	16,58	14,57	49,00
CNN2/ADAM/L <sub>2</sub>	Figura 42 (2)	95,28	85,17	73,82	91,95	82,87
CNN3/ADAM	Figura 42 (2)	<b>95,81</b>	81,95	73,05	92,81	81,33
CNN1/RMSProp	Figura 42 (3)	23,07	29,53	16,86	21,34	38,83
CNN2/ADAM/L <sub>2</sub>	Figura 42 (3)	97,74	97,27	91,71	95,74	94,24
CNN3/ADAM	Figura 42 (3)	<b>98,00</b>	97,23	92,54	96,20	93,43
CNN1/RMSProp	Figura 42 (4)	28,16	30,34	18,92	26,47	64,17
CNN2/ADAM/L <sub>2</sub>	Figura 42 (4)	97,04	97,29	62,15	94,32	93,27
CNN3/ADAM	Figura 42 (4)	<b>97,87</b>	97,63	63,35	95,89	94,17
CNN1/RMSProp	Figura 42 (5)	30,51	37,81	21,83	28,06	43,65
CNN2/ADAM/L <sub>2</sub>	Figura 42 (5)	96,89	94,01	60,68	93,99	83,50
CNN3/ADAM	Figura 42 (5)	<b>97,44</b>	96,13	61,80	95,07	89,83

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado,

<sup>(5)</sup> F1 = Score F1

**Tabela 25 – Resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelos CNNs treinados com imagens mosaico compostas de retalhos 10x10 pixels**

Modelo	Figura	ACCG (%)	ACCM (%)	J (%)	JP (%)	F1 (%)
CNN1/RMSProp	Figura 41 (a)	71,94	77,82	48,60	68,59	42,15
CNN2/ADAM/L <sub>2</sub>	Figura 41 (a)	95,71	96,52	60,64	91,91	77,49
CNN3/ADAM	Figura 41 (a)	<b>97,72</b>	97,84	94,99	95,57	89,54
CNN1/RMSProp	Figura 42 (1)	77,75	77,38	49,63	74,81	68,99
CNN2/ADAM/L <sub>2</sub>	Figura 42 (1)	94,30	94,41	59,79	89,68	88,67
CNN3/ADAM	Figura 42 (1)	<b>94,37</b>	94,46	60,20	90,27	88,80
CNN1/RMSProp	Figura 42 (2)	58,76	64,34	32,19	50,47	22,35
CNN2/ADAM/L <sub>2</sub>	Figura 42 (2)	94,00	78,98	67,24	90,39	73,15
CNN3/ADAM	Figura 42 (2)	<b>96,10</b>	79,11	72,90	92,99	81,80
CNN1/RMSProp	Figura 42 (3)	51,96	65,81	39,97	49,96	35,90
CNN2/ADAM/L <sub>2</sub>	Figura 42 (3)	97,86	94,33	61,15	95,99	89,55
CNN3/ADAM	Figura 42 (3)	<b>98,33</b>	96,28	62,34	96,76	92,56
CNN1/RMSProp	Figura 42 (4)	68,69	75,75	48,17	66,38	45,39
CNN2/ADAM/L <sub>2</sub>	Figura 42 (4)	97,66	97,46	63,03	95,47	92,06
CNN3/ADAM	Figura 42 (4)	<b>97,97</b>	97,55	63,50	96,08	93,74
CNN1/RMSProp	Figura 42 (5)	90,37	81,69	53,13	88,78	65,55
CNN2/ADAM/L <sub>2</sub>	Figura 42 (5)	97,46	97,02	61,92	95,14	88,41
CNN3/ADAM	Figura 42 (5)	<b>97,47</b>	96,84	61,94	95,17	89,30

Na Tabela 26 são apresentados os resultados para a segmentação semântica das imagens das Figura 41 e Figura 42, utilizando o processo de fusão de dados e a segmentação feita utilizando somente os modelos CNNs treinados com retalhos de imagens de tamanho 20x20 e 10x10 *pixels*. Pode-se notar que o método de fusão de dados obteve melhores resultados, com uma acurácia superior a 99% em todos os casos.

**Tabela 26 – Comparação dos resultados obtidos para a classificação das imagens das Figura 41 e Figura 42 utilizando modelo CNNs treinados com imagens mosaico compostas de retalhos e utilizando o processo de fusão de dados**

Modelo	Banco de Imagem do Classificador CNN	Figura	ACCG (%)
CNN3/ADAM	Imagens mosaico Retalhos 10x10 <i>pixels</i>	Figura 41 (a)	97,72
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 41 (a)	<b>99,81</b>
CNN3/ADAM	Imagens mosaico Retalhos 20x20 <i>pixels</i>	Figura 42 (1)	94,91
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 42 (1)	<b>99,63</b>
CNN3/ADAM	Imagens mosaico Retalhos 10x10 <i>pixels</i>	Figura 42 (2)	96,10
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 42 (2)	<b>99,09</b>
CNN3/ADAM	Imagens mosaico Retalhos 10x10 <i>pixels</i>	Figura 42 (3)	98,33
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 42 (3)	<b>99,66</b>
CNN3/ADAM	Imagens mosaico Retalhos 20x20 <i>pixels</i>	Figura 42 (4)	97,87
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 42 (4)	<b>99,83</b>
CNN3/ADAM	Imagens mosaico Retalhos 10x10 <i>pixels</i>	Figura 42 (5)	97,47
Fusão de Dados (CNN2/ADAM/L <sub>2</sub> + Imagem Homogênea)	Imagens Mosaico Retalhos 40x40 <i>pixels</i>	Figura 42 (5)	<b>99,85</b>

### 5.5 Avaliação da Metodologia em Regiões usadas por Adarme et al. (2020) e Bem et al. (2020)

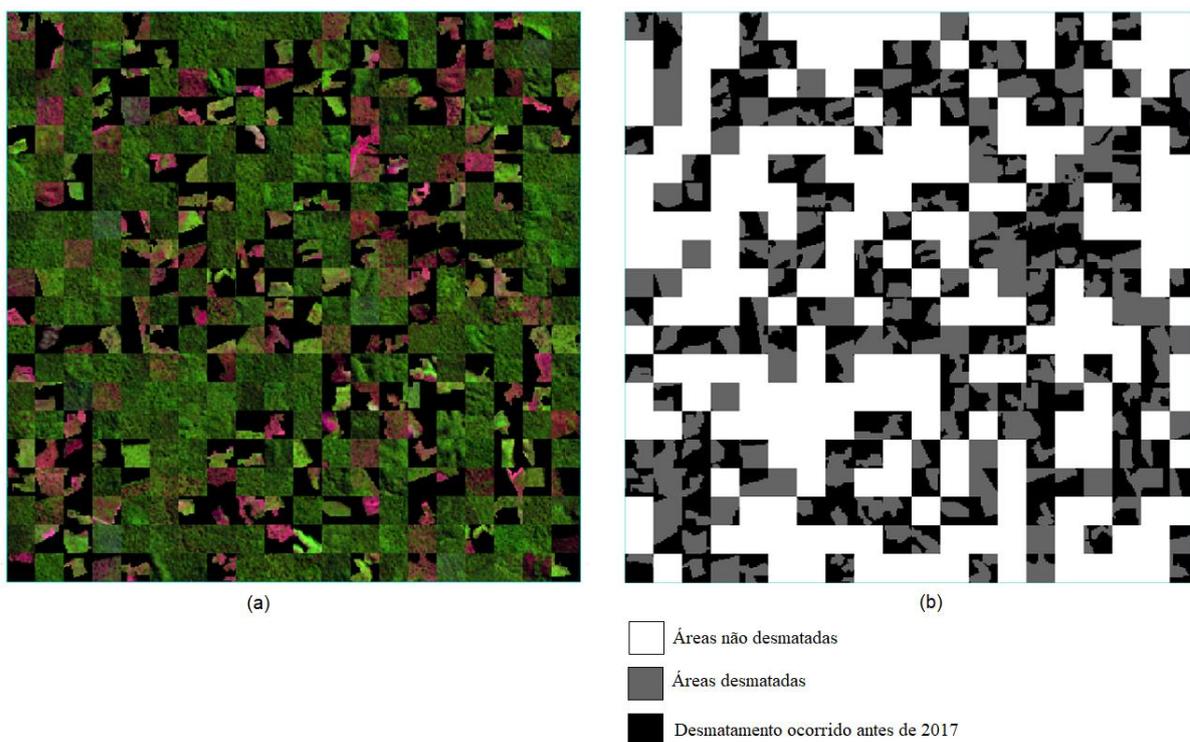
Conforme já salientado anteriormente, a metodologia apresentada neste trabalho também foi aplicada nas mesmas regiões avaliadas por Bem *et al.* (2020) e Adarme *et al.* (2020), para classificação de áreas desmatadas. Bem *et al.*, (2020) mapearam o desmatamento ocorrido entre 2017 e 2019. Adarme *et al.*, (2020) mapearam o desmatamento entre 2016 e 2017. Nesse caso, as saídas das arquiteturas CNNs foram ajustadas para realizar a classificação das mesmas classes utilizadas por Bem *et al.* (2020) e Adarme *et al.* (2020): desmatamento e não desmatamento.

Foram montados dois conjuntos de dados com 2.000 imagens-mosaico cada um utilizando as imagens avaliadas por Bem *et al.* (2020) e Adarme *et al.* (2020), respectivamente. A Tabela 27 mostra as imagens utilizadas para a construção de cada conjunto de dados e a divisão de cada conjunto em treinamento e validação. Para esse experimento, o tamanho de cada imagem mosaico foi de 400x400 *pixels*. Devido ao tamanho das áreas de desmatamento serem muito pequenos, em geral áreas menores que 40x40 *pixels*, para esse caso o tamanho dos retalhos utilizados foi 20x20 *pixels*.

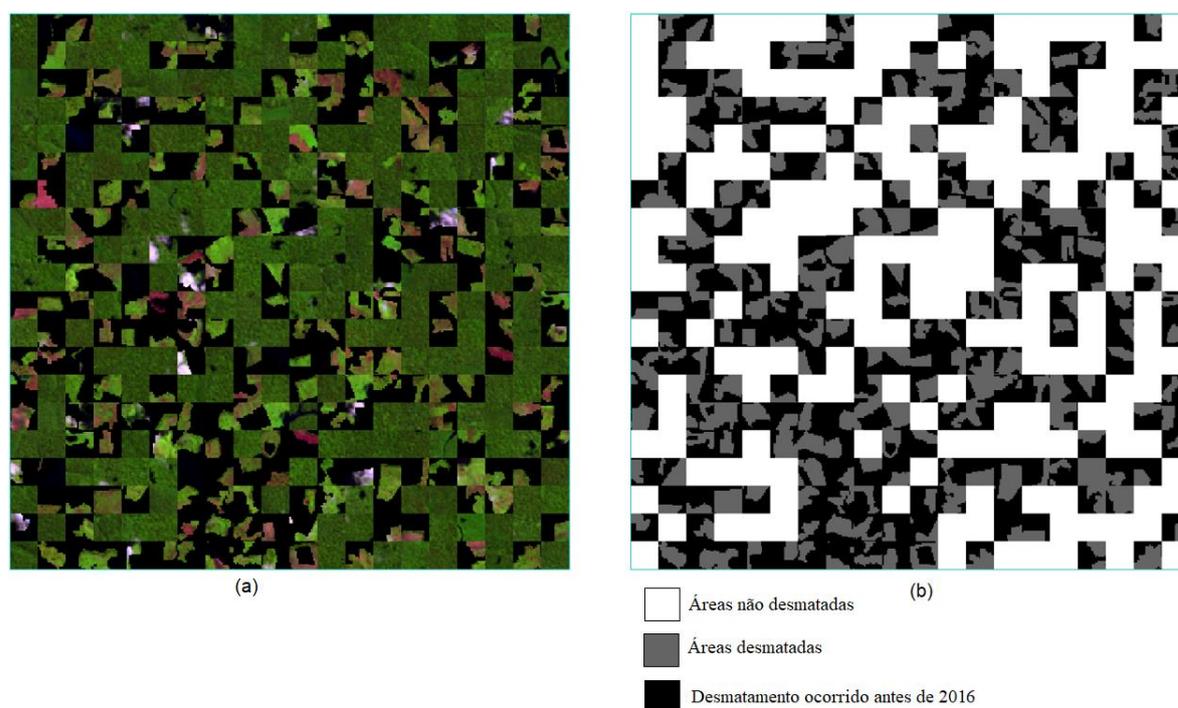
**Tabela 27 – Bancos de imagens montados utilizando as imagens avaliadas por Bem *et al.* (2020) e Adarme *et al.* (2020)**

Banco de Imagem	Orbita/Ponto	Data	Treinamento	Validação
1 (Bem <i>et al.</i> (2020))	227/63	08/07/2017		
	227/63	21/07/2017	1400	600
	227/65	18/08/2017		
	230/65	21/07/2018		
2 (Adarme <i>et al.</i> (2020))	230/65	21/06/2017	1400	600
	230/65	24/06/2018		

As Figuras Figura 43 e Figura 44 mostram exemplos de imagens mosaico e os correspondentes padrões ouro para os conjuntos de dados 1 e 2, respectivamente. Cada retalho de imagem corresponde a uma região de desmatamento ou não-desmatamento. A Tabela 28 apresenta o desempenho das redes CNNs para cada um desses conjuntos de dados. Foi utilizado o melhor modelo de cada CNN apresentado na Tabela 12.



**Figura 43 – Exemplo de imagem-mosaico para a região utilizada por Bem *et al.* (2020). Em (a) é a imagem-mosaico mostrada em uma composição colorida B6-B5-B4 e (b) o padrão ouro correspondente.**



**Figura 44** – Exemplo de imagem-mosaico para a região utilizada por Adarme *et al.* (2020). Em (a) é a imagem-mosaico mostrada em uma composição colorida B6-B5-B4 e (b) o padrão ouro correspondente.

**Tabela 28** – Desempenho de três melhores modelos de cada CNN usando as mesmas regiões utilizadas por Bem *et al.* (2020) e Adarme *et al.* (2020) para os respectivos conjuntos de validação

Banco de Imagem	Experimento	ACCG (%)	ACCM (%)	J (%)	JP (%)	F1 (%)
1	CNN1/RMSProp	98,48	97,98	96,26	97,03	99,68
	CNN2/ADAM/L <sub>2</sub>	98,44	97,94	96,89	96,93	98,68
	CNN3/RMSProp/L <sub>2</sub>	97,96	97,30	95,91	95,99	98,14
2	CNN1/RMSProp	98,59	98,10	96,26	97,25	99,90
	CNN2/ADAM/L <sub>2</sub>	99,55	99,32	98,95	99,10	99,60
	CNN3/RMSProp/L <sub>2</sub>	99,52	99,33	98,85	99,05	99,55

O método de fusão de dados foi empregado para classificar as imagens apresentadas nas Figura 45 e Figura 46. Para esses casos o modelo utilizado foi CNN2/ADAM/L<sub>2</sub>. A Figura 45 corresponde a uma região avaliada por Bem *et al.* (2020). Para esta imagem, a acurácia obtida foi de 99,91% e o valor de F1-Score foi de 99,94%. O valor da acurácia foi muito semelhante ao valor obtido por Bem *et al.* (2020). O F1-Score foi maior que o valor obtido pelos autores, que foi de 94,65%. A Figura 46 corresponde a uma região avaliada por Adarme *et al.* (2020). Para esta imagem, a acurácia obtida foi de 99,94%, e o valor do F1-Score foi de 99,92%. O valor da acurácia foi superior ao melhor resultado obtido por Adarme *et al.* (2020), que foi de 95%. Em relação ao valor do F1-Score, o valor foi superior ao obtido pelos autores, que foi de 63%.

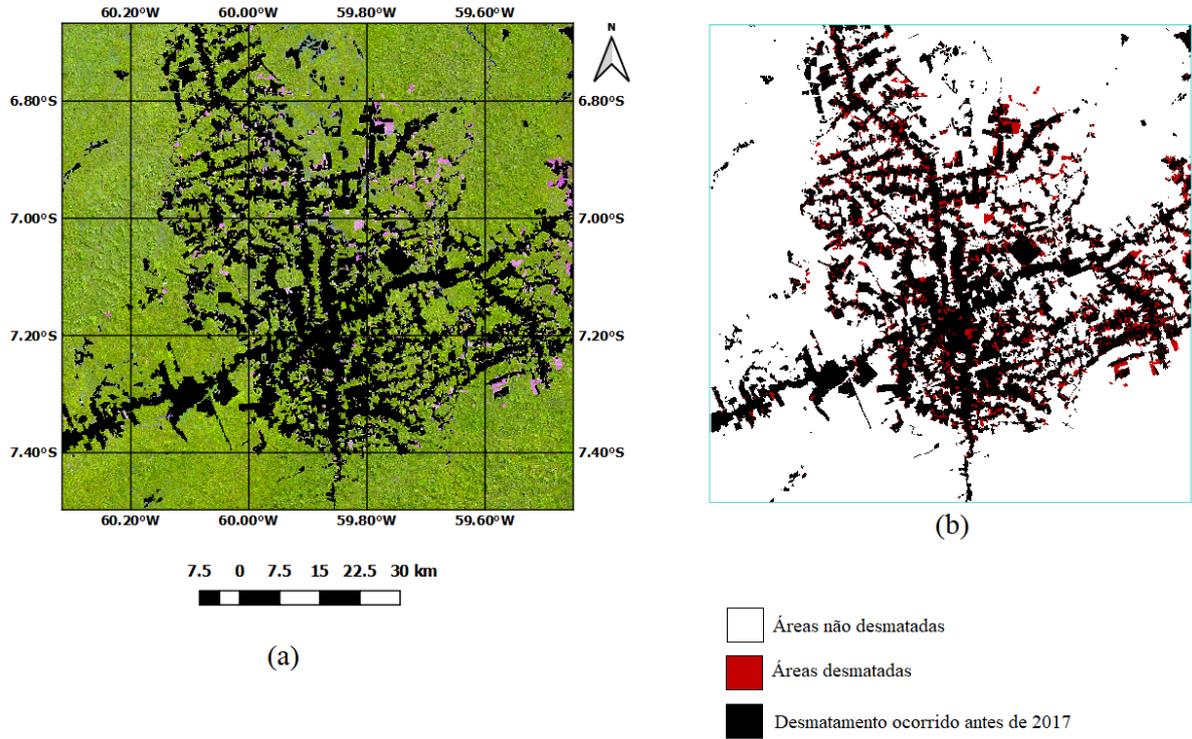


Figura 45 – Imagem LANDSAT-8/OLI da região amazônica. (a) Imagem de uma região (3200x3200pixels) da cena LANDSAT-8/OLI 230/65 avaliada por Bem *et al.* (2020) exibida em composição colorida B6-B5-B4. (b) mostra imagem classificada, com acurácia de 99,91%.

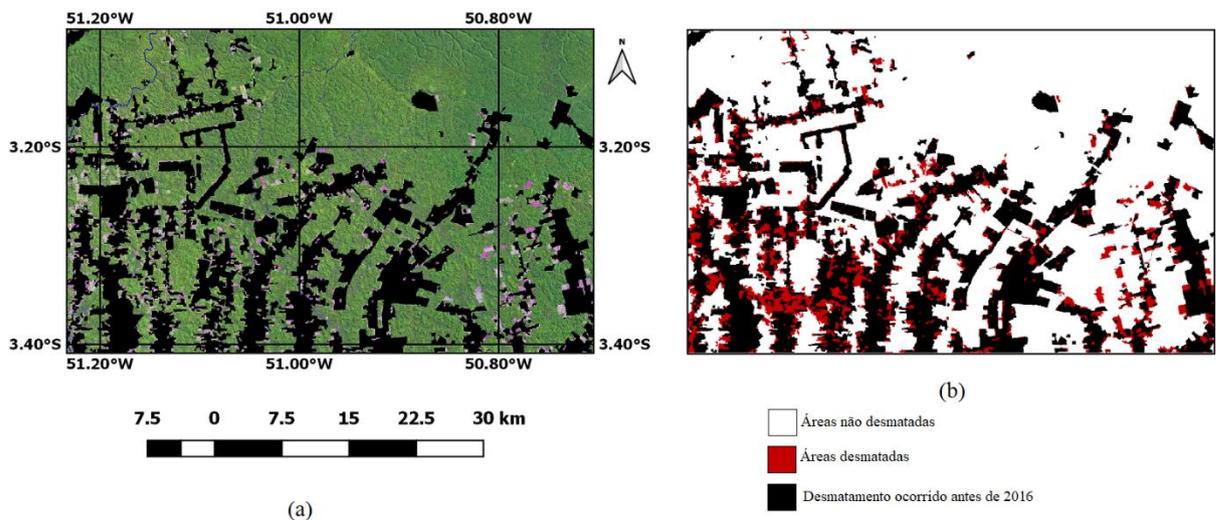


Figura 46 – Imagem LANDSAT-8/OLI da região amazônica. Em (a), imagem de uma região (2000x1200pixels) da cena LANDSAT-8/OLI 225/62 avaliada por Adarme *et al.* (2020) exibida em composição colorida B6-B5-B4. Em (b), a imagem é classificada com acurácia de 99,94%.

Na Tabela 29 são apresentados os resultados obtidos para a segmentação semântica das imagens da Figura 45 e da Figura 46 utilizando o modelo CNN2/ADAM/L<sub>2</sub> e o processo de fusão de dados. Ao observar esses resultados, nota-se que o processo de fusão de dados aumentou a acurácia da segmentação das imagens. Para a imagem da Figura 45, a acurácia aumentou de 95,80% para 99,91. Em relação a imagem da Figura 46, a acurácia aumentou de 93,49% para 99,94%. Isso mostra a eficácia da metodologia proposta classificação de uso do solo também para a segmentação de imagens em regiões de desmatamento e não desmatamento.

**Tabela 29 – Resultados obtidos para a segmentação semântica das imagens da Figura 45 e Figura 46 utilizando os modelos CNN2/ADAM/L<sub>2</sub> e Fusão de Dados**

Figura	Região/Cena	Data	Método de Segmentação	ACCG <sup>(1)</sup> (%)	ACCM <sup>(2)</sup> (%)	J <sup>(3)</sup> (%)	JP <sup>(4)</sup> (%)	F1 <sup>(5)</sup> (%)
44	230/65 Bem et al. 2020	21/07/2018	CNN2/ADAM/L <sub>2</sub>	95,80	92,73	78,84	93,68	97,74
			Fusão de Dados	99,91	99,84	99,14	99,82	99,94
45	225/62 Adarme et al. 2020	24/06/2018	CNN2/ADAM/L <sub>2</sub>	93,49	91,12	79,12	89,83	95,33
			Fusão de Dados	99,94	99,97	99,71	99,89	99,92

<sup>(1)</sup> ACCG = Acurácia Global, <sup>(2)</sup> ACCM = Acurácia Média, <sup>(3)</sup> J = Coeficiente Jaccard, <sup>(4)</sup> JP = Coeficiente Jaccard Ponderado,

<sup>(5)</sup> F1 = Score F1

## 6 CONCLUSÕES

Registra-se inicialmente o amplo escopo desse trabalho, que pode ser considerado através da quantidade de objetivos específicos propostos. Procurou-se não só utilizar redes convolucionais para uma aplicação específica de sensoriamento remoto, mas a duas: caracterização do uso de solo e identificação de regiões desmatadas. Os resultados obtidos em ambas as aplicações, com valores de acurácias superiores a 99%, atestam a excelência do método proposto. Essa excelência é atestada não só pelos altos valores das métricas, mas também pela comparação com outros trabalhos da literatura que se debruçaram com o problema de desmatamento na região amazônica.

A amplitude dos dados utilizados, abrangendo regiões de vários estados, confere uma credibilidade adicional a metodologia proposta, que não se ressentiu, nesse aspecto, da escassez de dados, nem na circunscrição dos mesmos a uma área específica da Amazônia legal. Procurou-se abranger uma grande região em que o uso da terra já se encontra diversificado em função dos sucessivos desmatamentos ocorridos ao longo dos últimos anos.

Diante do problema encontrado na prática do desbalanceamento das diversas classes de uso do solo, propôs-se utilizar uma técnica inovadora na área de sensoriamento remoto, a técnica de imagens mosaico, que utiliza retalhos de imagens para constituírem uma imagem maior. Tal técnica, no entanto, tende a produzir regiões de uso do solo que se parecem com a forma dos retalhos utilizados, diminuindo a precisão do método de classificação.

Diante dessa limitação, embora os resultados de acurácia já fossem aceitáveis, quando comparados a outros trabalhos da literatura, não houve uma acomodação. Foi-se além, ao propor um método de fusão de dados que agregou à saída das redes convolucionais dados do método de regiões homogêneas, através de um processo original de fusão de dados. Foi esse método de fusão que alçou a metodologia proposta para classificação do uso do solo ao estado da arte na área.

Procurou-se, em outro viés da pesquisa, utilizar-se tamanhos de retalhos menores. No entanto, a prática mostrou que os resultados aí obtidos não foram satisfatórios. Destarte, o método de fusão de dados com redes CNN treinadas com retalhos de 40x40 *pixels* ficou consagrado como o método de excelência.

A caracterização das redes convolucionais foi feita à exaustão, com diferentes métodos de otimização, diferentes técnicas para melhoria da generalização, tamanhos diferentes de retalhos e diferentes tamanhos de arquitetura. Pelos resultados obtidos com as três arquiteturas

no treinamento e na validação, pode-se afirmar que o *overfitting* não esteve presente em nenhuma delas.

Com o objetivo de avaliar uma tendência da área de aprendizado profundo de se trabalhar com transferência de conhecimento, sobretudo na área de processamento de imagens, procurou-se explorá-la, através da utilização de diversas arquiteturas pré-treinadas. Verificou-se, no entanto, que as arquiteturas propostas nesse trabalho que foram treinadas a partir de parâmetros inicializados randomicamente, superaram o desempenho dessas redes treinadas com transferência de conhecimento.

Um trabalho bem conduzido tem que, necessariamente mostrar que atingiu os objetivos a que se propôs. Nesse sentido, no texto a seguir são listados todos os objetivos específicos e mostrado que os mesmos foram atingidos na sua integralidade.

### **6.1 Avaliar o desempenho de arquiteturas de redes convolucionais para segmentação semântica de imagens de sensoriamento remoto em classes de uso do solo nas seguintes regiões da Amazônia Legal Brasileira: Amazonas, Rondônia, Mato Grosso e Pará**

Das três CNNs avaliadas neste trabalho, nas regiões acima mencionadas, a que apresentou melhores resultados foi a rede de tamanho médio, CNN2. Essa CNN, treinada com o método de otimização ADAM e com regularização  $L_2$  apresentou os melhores resultados para a segmentação semântica dos usos de solo em imagens de sensoriamento da Amazônia. O desempenho desse modelo no conjunto de validação obteve uma acurácia de 99,36%. Quando aplicado para segmentação de regiões da Amazônia, esse método obteve valores de acurácia superiores à 99% em todas as imagens avaliadas, conforme pode ser observado na Tabela 18.

### **6.2 Caracterizar o desempenho dessas arquiteturas em função de métodos utilizados para melhoria da generalização, como: regularização $L_2$ e *dropout***

Os métodos regularização  $L_2$  e *dropout* foram aplicados para cada CNN avaliada. Em geral, os resultados dos modelos CNNs empregando esses métodos de generalização foram muito próximos dos resultados dos modelos CNNs sem a utilização dos mesmos. Por exemplo, a partir da Tabela 11, pode-se observar que o desempenho do modelo CNN1, utilizando o método de otimização SGDM foi de 99,83%. Quando aplicado os métodos para melhoria da generalização, o melhor desempenho obtido foi para o modelo CNN1 utilizando *dropout* e regularização  $L_2$ , que nesse caso apresentou uma acurácia de 98,94%. Uma diferença de desempenho de apenas 0,11%.

### **6.3 Comparar o desempenho da arquitetura proposta com o desempenho de arquiteturas pré-treinadas, por meio da técnica de transferência de conhecimento**

Utilizando a técnica de transferência de conhecimento, o melhor desempenho foi obtido com a rede pré-treinada Resnet50 e o método de otimização ADAM. Para este modelo, foi obtida uma ACCG de 94,50%. Este desempenho foi bem inferior ao do modelo CNN1 treinado com o método de otimização RMSProp, onde obteve-se uma ACCG de 99,98%. Portanto, a comparação demonstrou a eficácia do método proposto em relação a arquiteturas pré-treinadas.

### **6.4 Utilizar o método desenvolvido para segmentação semântica de uso do solo em problemas de detecção de áreas desmatadas e não desmatadas**

A metodologia proposta neste trabalho para caracterização do uso do solo foi adaptada para realizar a classificação de áreas desmatadas e não desmatadas. Essa adaptação foi necessária para realizar uma avaliação da mesma nas regiões utilizadas por Bem *et al.* (2020) e Adarme *et al.* (2020). Esses autores tiveram como objetivo realizar a classificação de áreas desmatadas em regiões da Amazônia Legal Brasileira. O desempenho obtido na tarefa de classificação de áreas desmatadas foi semelhante àquele obtido na caracterização do uso do solo.

### **6.5 Comparar o desempenho da metodologia proposta com o desempenho de outros métodos publicados na literatura**

Na mesma região utilizada por Bem *et al.* (2020), a acurácia e o F1-score obtidas foram de 99,91% e 99,94%, respectivamente. A acurácia obtida por Bem *et al.* (2020) foi de 99,93%. Em relação a métrica F1-score, o valor obtido foi superior àquele obtido pelo referido autor, que obteve um valor 94,65%. Na mesma região utilizada por Adarme *et al.* (2020), a acurácia e o F1-Score obtidos pelo método proposto foram de 99,82% e 99,63%, respectivamente. Os resultados obtidos foram superiores àqueles obtidos pelo referido autor, que foi de 95% e 63%, respectivamente.

### **6.6 Propor a melhoria do desempenho das redes convolucionais na classificação do uso do solo e na detecção do desmatamento, através da fusão de dados entre a segmentação semântica das redes convolucionais com o método de segmentação de regiões homogêneas**

O método de fusão de dados proposto neste trabalho consistiu em utilizar a saída da CNN e o método de segmentação de regiões homogêneas para realizar um refinamento na segmentação da imagem. Considerando a Figura 41, a segmentação utilizando apenas o modelo CNN obteve uma acurácia de 96,84%. Quando aplicado o método de fusão de dados, a acurácia passou para 99,80%. Ao considerar os resultados apresentados na Tabela 18, observa-se também que o desempenho da segmentação semântica foi melhor utilizando o processo de fusão de dados. Portanto, o método de fusão de dados conseguiu melhorar o desempenho da segmentação semântica com CNN de imagens da Região Amazônica quando aplicado na tarefa de classificação de uso de solo.

### **6.7 Propor a técnica de imagem mosaico para o pré-processamento dos dados de entrada da CNN**

O treinamento dos modelos CNNs avaliados neste trabalho foi realizado por meio de um conjunto de imagens mosaico gerados a partir de retalhos de imagens. Esses retalhos correspondem a amostras de cada classe avaliada neste estudo. Com isso, foi possível criar um conjunto de dados balanceado, onde cada imagem possuía praticamente a mesma quantidade de amostras das três classes. A partir da Tabela 11, que apresenta os resultados dos modelos CNNs para o conjunto de validação, observa-se que foram obtidos bons desempenhos, principalmente para os modelos CNN1 e CNN2, com acurácia média de 99,90% e 98,63%, respectivamente.

### **6.8 Avaliar o melhor conjunto de bandas (3 ou 6) para classificação de imagens de sensoriamento para a usos do solo na região Amazônica**

Em geral, os valores de acurácia obtidos com os modelos CNNs treinados com o conjunto de seis bandas, foram superiores aos modelos CNNs treinados com o conjunto de três bandas. No entanto, para os modelos CNN1 e CNN3, essas diferenças de desempenho foram de apenas 0,60% e 0,46%, em média. Para o modelo CNN2, a diferença de desempenho foi superior, neste caso, a superioridade de desempenho na acurácia foi de 1,97%. A acurácia obtida com o modelo CNN2 treinados com o conjunto de seis e três bandas foi de 98,63% e 96,66%, respectivamente.

## CONSIDERAÇÕES FINAIS

O banco de imagens utilizado neste trabalho está disponível mediante solicitação. Para receber uma cópia do banco de dados de imagens, o pesquisador é primeiramente solicitado a entrar em contato no e-mail joelparente@gmail.com solicitando acesso a base de dados. O pesquisador receberá um formulário e deverá enviá-lo preenchido e assinado. Finalmente, será permitido o acesso ao banco de dados de imagens.

Em vista dos resultados apresentados, conclui-se que a metodologia proposta neste trabalho mostrou-se eficiente na realização da tarefa de segmentação/classificação de imagens de sensoriamento remoto para regiões da Amazônia Legal Brasileira, com aplicações voltadas tanto para caracterização do uso do solo como classificação de regiões desmatadas e não desmatadas. Isso mostra a capacidade de adaptação do modelo proposto.

Como trabalhos futuros, propõe-se expandir a área geográfica de estudos, selecionando mais regiões, e utilizando-se classes de uso do solo mais variadas, além de utilizar outros tipos de imagens de sensoriamento remoto, por exemplo, imagens de radar de abertura sintética (do inglês *Synthetic Aperture Radar* - SAR). Isso possibilitaria, por exemplo, realizar análises de imagens da região da Amazônica geradas durante o período conhecido como estação chuvosa, período no qual existe uma grande quantidade de nuvens na região, pois o tipo de sensor utilizado na geração de imagens SAR independe das condições meteorológicas existentes na região em análise.

O trabalho limitou-se a observar apenas um instante de tempo, não se detendo em avaliar uma janela temporal. Acredita-se que trabalhos futuros possam vir a ser feitos, na medida em que a metodologia foi extensamente caracterizada na presente proposta de tese.

Outra possibilidade que anotamos é utilizar fusão de dados de redes pré-treinadas, uma técnica que tem mostrado resultados promissores em outras áreas de processamento digital de imagens.

## 7 REFERÊNCIAS

- ADARME, Mabel Ortega *et al.* Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery. *Remote Sensing*, v. 12, n. 6, 2020.
- AGGARWAL, C C. *Neural Networks and Deep Learning*. New York: Springer, 2018.
- ALBAWI, Saad; MOHAMMED, Tareq Abed Mohammed; ALZAWI, Saad. Understanding of a Convolutional Neural Network. 2017, Antalya, Turkey: 2017 International Conference on Engineering and Technology (ICET), 2017. p. 1–6. Disponível em: <<https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network>>.
- AUGUSTO-SILVA, P. B. *et al.* Análise de classificadores para mapeamento de uso e cobertura do solo. 2013, Foz do Iguaçu, PR: [s.n.], 2013. p. 6917–6922.
- BAATZ, M; SCHAPE, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. 2000, Wichmann Verlag, Karlsruhe: [s.n.], 2000. p. 12–23.
- BARBOSA, D P *et al.* Um Estudo Experimental com Classificadores baseados em Regiões e Perfis EVI. 2015, João Pessoa, PB, Brazil: [s.n.], 2015. p. 6381–6388.
- BEM, Pablo Pozzobon *et al.* Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. *Remote Sensing*, v. 12, n. 6, 2020.
- BENGIO, Yoshua. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, v. 2, n. 1, p. 1–127, 2009.
- BENGIO, Yoshua; LECUN, Yann. Scaling Learning Algorithms toward AI. *Large-Scale Kernel Machines*, n. 1, p. 1–41, 2007.
- BHARATKAR, Pravada S.; PATEL, Rahila. Evaluation of RSI classification methods for effective land use mapping. *Proceedings - 2013 International Conference on Communication Systems and Network Technologies, CSNT 2013*, p. 109–113, 2013.
- BULLOCK, Eric L.; WOODCOCK, Curtis E.; OLOFSSON, Pontus. Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. *Remote Sensing of Environment*, v. 238, n. November, p. 0–1, 2018. Disponível em: <<https://doi.org/10.1016/j.rse.2018.11.011>>.
- CASSOL, Henrique Luis Godinho *et al.* Maximum fraction images derived from year-based project for on-board autonomy-vegetation (PROBA-V) data for the rapid assessment of land use and land cover areas in Mato Grosso state, Brazil. *Land*, v. 9, n. 5, 2020.
- CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, v. 2017- Janua, p. 1800–1807, 2017.
- DENG, Li. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, v. 3, n. January 2014, 2014.
- FLORENZANO, T. G. *Iniciação em Sensoriamento Remoto*. 3. ed. São Paulo: Oficina de

textos, 2011.

GOODFELLOW, I; BENGIO, Y; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Disponível em: <[http://www.deeplearningbook.org/front\\_matter.pdf](http://www.deeplearningbook.org/front_matter.pdf)>.

GUO, Yanming *et al.* Deep learning for visual understanding: A review. *Neurocomputing*, v. 187, p. 27–48, 2016.

HE, Kaiming *et al.* Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 2016- Decem, p. 770–778, 2016.

HUBEL, D. H.; WIESEL, T. N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, v. 124, p. 574–591, 1959.

INPE. *Instituto Nacional de Pesquisas Especiais. Projeto TerraClass*. Disponível em: <<https://www.terraclass.gov.br/>>. Acesso em: 14 ago. 2020.

INPE. *National Institute of Space Research - projects and research - terraclass*. Disponível em: <[http://www.inpe.br/cra/projetos%5C\\_pesquisas/dados%5C\\_terraclass.php](http://www.inpe.br/cra/projetos%5C_pesquisas/dados%5C_terraclass.php)>. Acesso em: 17 ago. 2019a.

INPE. *Satellite Monitoring of Deforestation of the Brazilian Amazon Forest*. Disponível em: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>>. Acesso em: 11 set. 2020b.

INPE. *The consolidated rate of deforestation by shallow cut for the nine states of Legal Amazon (AC, AM, AP, MA, MT, PA, RO, RR e TO) in 2019*. Disponível em: <[http://www.inpe.br/noticias/noticia.php?Cod%5C\\_Noticia=5465](http://www.inpe.br/noticias/noticia.php?Cod%5C_Noticia=5465)>. Acesso em: 17 ago. 2020c.

JACCARD, P. The distribution of the flora in the alpine zone. *New Phytologist*, v. 11, n. 2, p. 37–50, 1912.

JENSEN, J R. *Remote sensing of the environment: an earth resource perspective second edition*. second ed. Harlow, England: Pearson Education Limited, 2014. v. 1.

KHOOSHABI, Ghazaleh Safarzadeh. *Segmentation Validation Framework*. 2013. Linköping, 2013.

KINGMA, Diederik P.; BA, Jimmy Lei. Adam: A method for stochastic optimization. 2015, San Diego: [s.n.], 2015. p. 1–15.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015.

LEE, Jacky; CARDILLE, Jeffrey A.; COE, Michael T. Agricultural expansion in mato grosso from 1986-2000: A bayesian time series approach to tracking past land cover change. *Remote Sensing*, v. 12, n. 4, p. 1–18, 2020.

LESEUX, Vitor *et al.* Análise entre métodos para construção de mapa temático de cobertura do solo utilizando sentinel-2 (sensor msi). 2019, Santos-SP, Brazil: INPE, 2019. p. 720–723.

LI, Ying *et al.* Deep learning for remote sensing image classification: A survey. *Wiley*

*Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 6, p. 1–17, 2018.

LIMA, Thaís Almeida *et al.* Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian Amazon. *Remote Sensing*, v. 11, n. 8, p. 961, 2019.

MARETTO, Raian V. *et al.* Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geoscience and Remote Sensing Letters*, p. 1–5, 2020.

MATHWORKS. *Evaluate semantic segmentation data set against ground truth*. Disponível em: <<https://in.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html>>. Acesso em: 10 ago. 2019.

MATHWORKS. *Practical Deep Learning Examples with MATLAB*. Disponível em: <<https://in.mathworks.com/campaigns/offers/deep-learning-examples-with-matlab.html>>. Acesso em: 10 maio 2020.

MENESES, Paulo Roberto; ALMEIDA, Tati De. *Introdução ao processamento de imagem de sensoriamento remoto*. Brasília: CNPq, 2012.

MUNDOGEO. *Sensoriamento Remoto Hiperespectral*. Disponível em: <<https://mundogeo.com/2004/08/23/sensoriamento-remoto-hiperespectral/>>.

NASCIMENTO, Filipe Silveira *et al.* Land cover changes in open-cast mining complexes based on high-resolution remote sensing data. *Remote Sensing*, v. 12, n. 4, p. 1–21, 2020.

NOVO, E M L M. *Sensoriamento Remoto: Princípios e Aplicações*. 3. ed. São Paulo: Edgard Blücher Ltda, 2008.

OVIEDO, Antonio; LIMA, William Pereira; AUGUSTO, Cicero. *O arco do desmatamento e suas flechas*. Disponível em: <[https://www.socioambiental.org/sites/blog.socioambiental.org/files/nsa/arquivos/nova\\_geografia\\_do\\_arco\\_do\\_desmatamento\\_isa.pdf](https://www.socioambiental.org/sites/blog.socioambiental.org/files/nsa/arquivos/nova_geografia_do_arco_do_desmatamento_isa.pdf)>.

PAN, Sinno Jialin; YANG, Qiang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, 2010.

RUSSAKOVSKY, Olga *et al.* ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, v. 115, n. 3, p. 211–252, 2015. Disponível em: <<http://dx.doi.org/10.1007/s11263-015-0816-y>>.

SANDLER, Mark *et al.* MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 4510–4520, 2018.

SCHERER, Dominik; MULLER, Andreas; BEHNKE, Sven. Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction. *BMC Bioinformatics*, v. 19, p. 92–101, 2010.

SERRÃO, M. K. M. *Detecção automática de bacilos em baciloscopia de campo claro usando aprendizado profundo e técnica de imagem mosaico*. 2020. Universidade Federal do Amazonas, 2020. Disponível em: <[https://tede.ufam.edu.br/bitstream/tede/7938/7/Dissertação\\_MikaelaSerrão\\_PPGEE.pdf](https://tede.ufam.edu.br/bitstream/tede/7938/7/Dissertação_MikaelaSerrão_PPGEE.pdf)>.

SERRE, Thomas *et al.* A quantitative theory of immediate visual recognition. *Progress in Brain Research*, v. 165, p. 33–56, 2007.

SEVO, Igor; AVRAMOVIC, Aleksej. Convolutional neural network based automatic object detection on aerial images. *IEEE Geoscience and Remote Sensing Letters*, v. 13, n. 5, p. 740–744, 2016.

SHIMABUKURO, Yosio Edemir *et al.* Monitoring deforestation and forest degradation using multi-temporal fraction images derived from Landsat sensor data in the Brazilian Amazon. *International Journal of Remote Sensing*, v. 40, n. 14, p. 5475–5496, 2019. Disponível em: <<https://doi.org/10.1080/01431161.2019.1579943>>.

SRIVASTAVA, Nitish *et al.* Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Physics Letters B*, v. 15, p. 1929–1958, 2014.

SZEGEDY, Christian *et al.* Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, p. 4278–4284, 2017.

TIKHONOV, A. N.; ARSENIN, A. *Solutions of Ill Posed Problems*. [S.l.]: Vh Winston, 1977.

U.S. GEOLOGICAL SURVEY. *Landsat 8 Data Users Handbook*. [S.l.: s.n.], 2019. v. 8. Disponível em: <<https://landsat.usgs.gov/documents/Landsat8DataUsersHandbook.pdf>>.

USGS. *Earth Explorer*. Disponível em: <<http://earthexplorer.usgs.gov/>>. Acesso em: 19 maio 2019a.

USGS. *What are the best spectral bands to use for my study?* Disponível em: <<https://landsat.usgs.gov/what-are-best-spectral-bands-use-my-study>>. Acesso em: 19 maio 2019b.

YAO, Chuchu *et al.* A review on image classification of remote sensing using deep learning. *2017 3rd IEEE International Conference on Computer and Communications, ICC3 2017*, p. 1947–1955, 2017.

YU, Zhiqi *et al.* Selection of landsat 8 OLI band combinations for land use and land cover classification. *2019 8th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2019*, n. July, p. 1–5, 2019.

**APÊNDICE 1**  
**ARTIGO PUBLICADO NO CONGRESSO GEOINFO 2020**

Proceedings XXI GEOINFO, November 30 - December 03, 2020, São José dos Campos, SP, Brazil. p 198-203

---

**Segmentação Semântica de Tipos de Uso de Solo na Amazônia  
Utilizando Aprendizado Profundo**

**Joel P. de Oliveira<sup>1</sup>, Marly G. F. Costa<sup>2</sup> e Cícero F. F. Costa Filho<sup>2</sup>**

<sup>1</sup>Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM)  
Manaus – AM – Brasil

<sup>2</sup>Universidade Federal do Amazonas (UFAM)  
Manaus – AM – Brasil

joelparente@gmail.com, marlygfcosta@gmail.com, cffccfilho@gmail.com

**Abstract.** *This work proposes a deep learning methodology to perform the semantic segmentation of LANDSAT-8 images of the following types of uses and land cover: forest, pasture and agriculture. The field of study is the Amazon region. The reference data were extracted from the TerraClass project in 2014. A CNN architecture was evaluated along with three optimization methods: SGDM, ADAM and RMSProp and the dropout and  $L_2$  regularization methods, as methods for generalization improvement. The best results were obtained using the RMSProp optimization method. The accuracy values obtained for the evaluated images were above 93%.*

**Resumo.** *Este trabalho propõe uma metodologia de aprendizado profundo para realizar a segmentação semântica de imagens LANDSAT-8 dos seguintes usos e cobertura de solos: floresta, pasto e agricultura. O campo de estudo é a região Amazônica. Os dados de referência foram extraídos do projeto TerraClass do ano de 2014. Foi avaliada uma arquitetura CNN juntamente com três métodos de otimização: SGDM, ADAM e RMSProp e os métodos dropout e regularização  $L_2$ , como métodos para a melhoria de generalização. Os melhores resultados foram obtidos utilizando o método de otimização RMSProp. Os valores de acurácia obtidos para as imagens avaliadas ficaram acima de 93%.*

## **1. Introdução**

Sensoriamento remoto é a utilização de diversas tecnologias com o objetivo de estudar os fenômenos que ocorrem na superfície da Terra. Essas tecnologias compreendem sensores, equipamentos instalados a bordo de aeronaves, espaçonaves e outras plataformas. Os dados gerados a partir de sistemas de sensoriamento remoto são de grande utilidade para várias aplicações, dentre as quais podemos citar: planejamento urbano, agrícolas, geológicas, monitoramento de desmatamento [Novo 2008].

Com respeito ao monitoramento do desmatamento na Amazônia através de sensoriamento remoto, o Instituto Nacional de Pesquisas Espaciais (INPE) é uma referência mundial. Dentre os vários projetos desenvolvidos pelo INPE para esse monitoramento, destacam-se o Programa de Monitoramento do Desflorestamento na Amazônia Legal (PRODES) e o TerraClass. O PRODES fornece dados por meio de mapas anuais de desmatamento na região Amazônica. O projeto TerraClass utiliza os dados gerados pelo

PRODES para realizar uma classificação de uso e cobertura de solo nas seguintes classes: floresta, pasto, agricultura, áreas urbanas, mineração e outros. Esse tipo de informação ajuda os órgãos do governo a desenvolverem políticas públicas de prevenção para conter o avanço do desmatamento [Noma et al. 2013]. Embora o PRODES e o TerraClass sejam grandes projetos e forneçam dados bastante confiáveis, eles ainda contam com uma significativa parcela do trabalho realizada com intervenção humana.

Para tornar mais eficiente o trabalho que envolva a análise de imagens de sensoriamento remoto, trabalhos têm sido propostos na área de aprendizado de máquina. [Bem et al. 2020] utilizaram aprendizado profundo para mapear o desmatamento entre 2017 e 2018 e entre 2018 e 2019. Foram utilizadas três cenas do LANDSAT-8 de regiões do estado do Pará e do Amazonas. Os dados de referência utilizados foram os dados PRODES dos anos de 2017 a 2019. Os melhores resultados obtidos com a CNN ResUnet foram acurácia e F1-Score de 99,93% e 94,65%, respectivamente. [Adarme et al. 2020] utilizaram aprendizado profundo para detecção automática de desmatamento por meio de duas imagens LANDSAT-8. As áreas de estudo foram duas regiões com diferentes padrões de desmatamento: os biomas Amazônia e Cerrado no Brasil. Os dados de referências utilizados foram os dados PRODES dos anos de 2017 e 2018. Os autores obtiveram como melhores resultados uma acurácia e um F1-Score de 95% e 63%, respectivamente, na Amazônia, e de 97% e 78%, respectivamente, no Cerrado.

Segundo [Aggarwal 2018], a partir da primeira década desse século, a rede neural renasceu sob o novo rótulo chamado de aprendizado profundo. Aprendizado profundo aborda o uso de modelos computacionais com arquiteturas hierárquicas compostas por múltiplas camadas de processamento a fim de "aprender" determinadas representações de dados nos mais diferentes formatos: áudio, imagens e texto [Lecun et al. 2015]. Quatro modelos principais compõem o aprendizado profundo: Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN), Máquinas de Boltzmann, *Autoencoders* e *Sparse Coding*. Segundo [Aggarwal 2018], a grande quantidade de dados disponíveis nos últimos anos, juntamente com o aumento do poder computacional, permitiram a utilização de arquiteturas mais profundas em relação ao que era possível anteriormente.

Os trabalhos de [Bem et al. 2020] e [Adarme et al. 2020] utilizaram aprendizado profundo para detectar áreas desmatadas em regiões da Amazônia brasileira. Os resultados alcançados por esses autores foram bastante satisfatórios. No entanto, os autores avaliaram uma região específica da Amazônia. Outra crítica que fazemos aos trabalhos previamente publicados é que os mesmos não se preocupam em disponibilizar o conjunto de dados utilizado para *benchmark*.

Um grande desafio para se treinar CNNs com imagens de sensoriamento remoto é que, normalmente, as classes são desbalanceadas. Em outras palavras, em uma região capturada pela imagem, existe um grande desbalanceamento em termos de área dos diversos tipos de solos. Esse problema pode levar, no treinamento da CNN, aos métodos de otimização terem um melhor desempenho nas classes mais frequentes. Para contornar esse problema, nesse trabalho utiliza-se a técnica de imagem-mosaico. Essa técnica consiste de extrair, a partir da imagem de satélite, pequenas retalhos retangulares uniformes, ou seja, com apenas um tipo de solo. A partir desses retalhos monta-se uma imagem maior, denominada de imagem-mosaico.

Portanto, considerando a ausência de um modelo de rede em aprendizado profundo adequado para classificação dos diversos tipos de solo para a região amazônica, e a dificuldade de treinamento de uma CNN com bom desempenho em todas as classes, esse trabalho pretende, através da técnica de imagem mosaico, utilizar uma CNN para segmentação e classificação de diversas regiões da Amazônia nos seguintes tipos de uso e cobertura de solo: agricultura, pasto e floresta.

## 2. Material e métodos

### 2.1. Conjunto de dados utilizados

Neste trabalho foram utilizadas imagens LANDSAT-8/OLI. As imagens estão disponíveis gratuitamente no site: <http://earthexplorer.usgs.gov/>. Foram utilizadas as seguintes bandas: Vermelho (B4); Infravermelho próximo (B5); Infravermelho de ondas curtas 1 (B6). Segundo [Yu et al. 2019], B4, B5 e B6 é a melhor combinação de três bandas para aplicações de sensoriamento remoto cujo o objetivo é realizar a classificação do solo. As imagens utilizadas abrangem os estados do Amazonas, Mato Grosso, Pará e Rondônia. Para a geração do padrão-ouro do classificador, foram utilizados os resultados do projeto TerraClass, do ano de 2014. Os dados foram adquiridos gratuitamente no site do INPE [Inpe 2019]. Foram extraídas informações sobre áreas de floresta, pasto e agricultura, correspondendo a um problema de reconhecimento de três classes.

Utilizando os dados de referências do projeto TerraClass, foram gerados retalhos de imagens de tamanho 40x40 *pixels* de cada uma das classes de solo. Foram gerados 4.000, 225.000 e 6.000 retalhos de agricultura, floresta e pasto, respectivamente. Utilizando os conjuntos de retalhos de imagens, foram geradas imagens que definiremos como imagens-mosaico. Cada uma dessas imagens-mosaico tem dimensão 400x400 *pixels* e são geradas selecionando-se aleatoriamente retalhos de agricultura, floresta ou pasto. Para cada imagem-mosaico gerada, gera-se também uma imagem-mosaico correspondente ao padrão-ouro da mesma. Para a construção dessas imagens que constituíram o padrão-ouro, os *pixels* que correspondem a região de floresta, pasto e agricultura foram marcados com o valor 255, 100 e 1, respectivamente. O conjunto de dados gerado possui um total de 3.600 imagens-mosaico, sendo que 3000 foram destinadas para o conjunto de treinamento e 600 para o conjunto de validação. Na Figura 1 mostra-se um exemplo de imagem-mosaico e o padrão-ouro correspondente. Pode ser observado que, dos 100 retalhos, 34 são de agricultura, 35 de floresta e 31 de pasto.

### 2.2. Métodos

Neste trabalho, foi avaliada uma arquitetura de CNN, três métodos de otimização e três métodos para a melhoria da generalização. Os métodos de otimização avaliados foram o Gradiente Descendente Estocástico com Momento (SGDM), Propagação da Raiz Média Quadrática (RMSProp) e Estimativa de Dinâmica Adaptativa (ADAM). Para cada um desses métodos foram empregados as seguintes técnicas para a melhoria de generalização: nenhuma técnica, camada de *dropout*, regularização L2 e camada de *dropout* com regularização L2. Os dados de entrada utilizados foram as imagens-mosaico. Foram realizadas 12 simulações (1 arquitetura x 3 métodos de otimização x 4 métodos para melhoria de generalização). Em seguida, o modelo com o melhor desempenho no conjunto de validação foi selecionado para classificar algumas imagens da Amazônia.

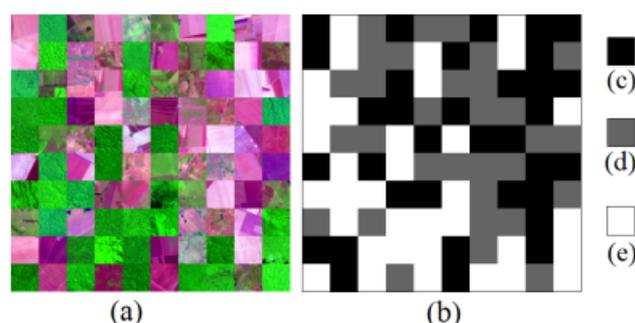


Figura 1. Exemplo de imagem-mosaico. Em (a) mostra-se uma imagem mosaico composta por retalhos de agricultura, floresta e pasto, em (b) a imagem padrão-ouro, e em (c), (d) e (e) temos o padrão-ouro para cada retalho de imagem, correspondente a área de agricultura, pasto e floresta, respectivamente.

### 2.2.1. Arquitetura CNN, Métricas e Parâmetros de Treinamento

A arquitetura CNN utilizada neste trabalho foi baseada nas arquiteturas propostas no trabalho de [Miyagawa et al. 2018]. Neste trabalho, os autores realizaram a segmentação do lúmen em imagens de tomografia por coerência ótica intravascular (IVOCT). Os melhores resultados para acurácia, valor de Dice e de Jaccard ficaram acima de 99%, 98% e 97%, respectivamente. No trabalho ora apresentado, a CNN utilizada possui duas etapas de subamostragem (*maxpooling*) e duas de sobreamostragem. Antes de cada subamostragem, existem três sequências de camadas convolucionais 3 x 3, camada *batch normalization* e ReLU.

Neste trabalho foram calculadas a Acurácia Global (ACCG), Acurácia Média (ACCM), coeficiente de similaridade de Jaccard (J), coeficiente de similaridade Jaccard ponderado (JP) e Score F1 (F1) [MathWorks 2017]. Uma estação de trabalho com *Windows 10*, *Matlab 2019a* e com *NVIDIA Quadro GV100 32GB* e 5120 núcleos CUDA foi utilizado nos experimentos. Em relação aos parâmetros de treinamento da CNN, foi utilizada a taxa de aprendizado inicial = 0,001, fator de queda de taxa de aprendizado = 0,5, número de épocas = 200, tamanho do lote = 2, parâmetro da camada *dropout* = 0,3, fator de regularização  $L_2=0,001$ . Esses valores foram ajustados de maneira experimental.

## 3. Resultados

A Tabela 1 apresenta o desempenho obtido para a CNN quando combinada com os método de otimização e com os métodos para melhoria da generalização. A Tabela 2 apresenta a matriz de confusão para o modelo com o melhor desempenho no conjunto de validação, que neste caso foi aquele no qual foi empregado o método RMS-Prop como método de otimização e sem utilizar nenhuma técnica para melhoria da generalização. Esse modelo foi utilizado para segmentar/classificar algumas imagens da região Amazônica. Na Figura 2 mostramos três imagens LANDSAT-8 de regiões da Amazônia com seus respectivos padrão-ouro, e a imagem classificada pelo modelo CNN. As imagens 1 e 2 correspondem regiões da cena 001/66, enquanto que a imagem 3, a uma área da cena 224/68. Os valores de acurácia obtidos para as imagens 1, 2 e 3 foram 98,91%, 96,72% e 93,10% respectivamente.

	<b>Experimento</b>	<b>ACCG (%)</b>	<b>ACCM (%)</b>	<b>J (%)</b>	<b>JP (%)</b>	<b>F1 (%)</b>
1	SGDM	95,984	95,995	92,406	92,389	86,854
2	SGDM/Dropout	94,900	94,919	90,487	90,463	87,111
3	SGDM/L2	96,370	96,381	93,091	93,075	87,981
4	SGDM/Dropout/L2	95,947	95,960	92,328	92,311	89,076
5	ADAM	97,158	97,171	94,537	94,524	89,433
6	ADAM/Dropout	97,265	97,272	94,740	94,727	89,827
7	ADAM/L2	97,095	97,102	94,420	94,408	89,292
8	ADAM/Dropout/L2	96,824	96,838	93,906	93,892	90,229
9	RMSProp	97,467	97,477	95,109	95,097	89,517
10	RMSProp/Dropout	96,807	96,810	93,877	93,865	88,308
11	RMSProp/L2	97,260	97,268	94,729	94,717	90,253
12	RMSProp/Dropout/L2	96,533	96,546	93,387	93,371	90,273

Tabela 1. Desempenho da CNN.

		<b>Classes Preditas</b>		
		<b>Agricultura</b>	<b>Pasto</b>	<b>Floresta</b>
<b>Classes Reais</b>	<b>Agricultura</b>	0,9799	1,96E-02	4,11E-04
	<b>Pasto</b>	4,87E-02	0,9466	4,67E-03
	<b>Floresta</b>	2,56E-04	1,99E-03	0,9978

Tabela 2. Matriz de Confusão para o modelo CNN/RMSProp.

#### 4. Discussão

A partir da Tabela 1, com respeito aos métodos otimização, conclui-se que os resultados para a acurácia obtidos utilizando-se os métodos ADAM e RMSProp ficaram muito próximos, em média 97,08% e 97,017%, respectivamente. Por outro lado, os resultados obtidos com o método SGDM, foram inferiores, com média de 95,80%. A partir da tabela de confusão, verificou-se um maior erro de classificação em regiões em que o padrão-ouro apontava áreas como sendo de pasto mas que foram classificadas como agricultura e vice-versa. O modelo com melhor desempenho foi o modelo CNN com o método RMSProp, sem utilizar nenhuma técnica para melhoria da generalização. Esse modelo foi empregado para avaliar a classificação de algumas regiões na Amazônia apresentadas na Figura 2. A Acurácia obtida para essas três regiões variaram significativamente, entre 98,91% e 93,10%.

#### 5. Conclusões

Este trabalho propôs uma metodologia para segmentar o uso do solo para a região amazônica para as classes de pasto, agricultura e floresta. A metodologia consistiu em avaliar uma arquitetura de CNN e treiná-la utilizando um banco de dados de imagens-mosaico. Foram utilizadas imagens óticas LANDSAT-8 da região amazônica. Diante dos resultados apresentados, pode-se concluir que a metodologia proposta neste trabalho mostrou-se promissora para realizar a tarefa de segmentação/classificação de imagens de sensoriamento para regiões da Amazônia. Com o objetivo de melhorar a acurácia da segmentação das imagens, pretende-se, em trabalhos futuros, avaliar arquiteturas de CNNs mais profundas, utilizar transferência de conhecimento e utilizar mais bandas da imagem LANDSAT-8. Por fim, pretende-se também avaliar a metodologia proposta com um maior número de regiões da Amazônia.

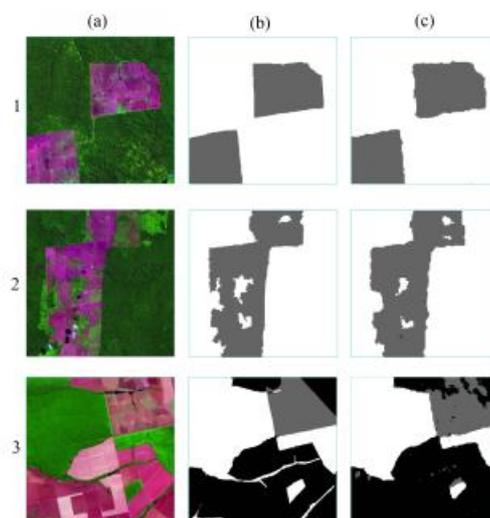


Figura 2. Imagens LANDSAT-8 de regiões da Amazônia. Na coluna (a) temos as imagens originais. Em (b) temos o padrão-ouro. Em (c) temos as imagens classificadas pelo modelo CNN/RMSProp. A acurácia obtida para as imagens 1,2 e 3 foi de 98,91%, 96,72% e 93,10% respectivamente.

## Referências

- Adarme, M. O., Feitosa, R. Q., Happ, P. N., A., A. C., and 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*, 12:910.
- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer, New York.
- Bem, P. P., Carvalho Junior, O. A., Guimarães, R. F., and Gomes, R. A. T. (2020). Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. *Remote Sensing*, 12:901.
- Inpe (2019). TerraClass. [http://www.inpe.br/cra/projetos\\_pesquisas/dados\\_terraClass.php](http://www.inpe.br/cra/projetos_pesquisas/dados_terraClass.php).
- Lecun, Y., Bengio, Y., and Hilton, G. (2015). Deep learning. *Nature*, 521:436–444.
- MathWorks (2017). Evaluate semantic segmentation data set against ground truth. <https://in.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html>.
- Miyagawa, M., Costa, M. G. F., Gutierrez, M. A., Costa, J. P. G. F., and Costa Filho, C. F. F. (2018). Lumen segmentation in optical coherence tomography images using convolutional neural network. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1:1–4.
- Noma, A., Körting, T. S., and Fonseca, L. M. G. (2013). Uma comparação entre classificadores usando regiões e perfis evi para agricultura. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto*, 1:2250–2257.
- Novo, E. M. L. M. (2008). *Sensoriamento Remoto: Princípios e Aplicações*. Edgard Blücher Ltda, São Paulo.
- Yu, Z., L. D., Yang, R., and Tang, J. (2019). Selection of landsat 8 oli band combinations for land use and land cover classification. *8th International Conference on Agro-Geoinformatics*, 1:1–5.

## APÊNDICE 2

### ARTIGO PUBLICADO NO PERIÓDICO IEEE ACCESS



Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2020, accepted October 13, 2020, date of publication October 15, 2020, date of current version October 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031533

# Methodology of Data Fusion Using Deep Learning for Semantic Segmentation of Land Types in the Amazon

JOEL PARENTE DE OLIVEIRA<sup>1</sup>, MARLY GUIMARÃES FERNANDES COSTA<sup>2</sup>, (Member, IEEE),  
AND CÍCERO FERREIRA FERNANDES COSTA FILHO<sup>2</sup>, (Member, IEEE)

<sup>1</sup>Operations and Management Center of the Amazon Protection System (CENSIPAM), Manaus 69049-630, Brazil

<sup>2</sup>Federal University of Amazonas, Manaus 69080-900, Brazil

Corresponding author: Joel Parente De Oliveira (joelparente@gmail.com)

This work was supported in part by the Samsung Electronics of Amazonia Ltd., through the terms of Federal Law no. 8.387/1991, by the agreement no. 004, assigned by the Center for Research and Development in Electronics and Information from the Federal University of Amazonas—CETEL/UFAM, in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Funding Code 001, and in part by the Academic English Solutions (<https://www.academicenglishsolutions.com>) revised this paper.

**ABSTRACT** This study proposes a methodology using deep learning and a multi-resolution segmentation algorithm to perform the semantic segmentation of remote sensing images. Initially the image is segmented using a CNN, and then an image with homogeneous regions is generated using a multi-resolution segmentation algorithm. Finally, a data fusion process is performed with these two images, generating the final classified image. The field of study was the Brazilian Amazon region. The proposed methodology classifies images in the following classes: forest, pasture and agriculture. The input data used were LANDSAT-8/OLI images. The reference data were extracted from the results of the TerraClass project in 2014. Two datasets were evaluated: the first with six bands and the second with three bands. Three CNN architectures were evaluated together with three optimization methods: SGDM, ADAM, and RMSProp and the dropout and  $L_2$  regularization methods as methods for generalization improvement. The best model, CNN + optimization method + technique for generalization improvement, evaluated in the validation set, was submitted to a 5-fold cross validation methodology, and the results were compared with pre-trained networks using the learning transfer methodology; in this case the networks used for comparison were ResNet50, InceptionResNetv2, MobileNetv2 and Xception. The proposed methodology was evaluated through image segmentation of some regions of the Amazon. Finally, the proposed methodology was evaluated in regions used by other authors. The accuracy values obtained for the images evaluated were over 99%.

**INDEX TERMS** Deep learning, convolutional neural networks, remote sensing, image segmentation.

## I. INTRODUCTION

Remote sensing consists of collecting information about an object or geographic area without direct contact. This information was compiled through various technologies such as sensors installed on-board aircraft, satellites, or other platforms [1]. The data generated from remote sensing systems are very useful for several applications, for example: urban - demographic interventions, government registrations, urban planning, support for the real estate sector; agricultural - monitoring crop conditions, crop forecasting, soil erosion; geological - minerals, oil and natural gas

exploration; ecological - monitoring flooded regions, soils, oceans and continental waters; forest - lumber production, biomass estimation, monitoring of deforestation) [2].

With regard to monitoring deforestation in the Amazon through remote sensing, the National Institute for Space Research (INPE) is a world reference. Among the various projects developed by INPE for this monitoring, the Program for Monitoring Deforestation in the Legal Amazon (PRODES) and TerraClass stand out [3], [4]. PRODES provides data through annual maps of deforestation in the Amazon region. The TerraClass project uses the data generated by PRODES to perform a land use and land cover classification in the following classes: forest, pasture, agriculture, urban areas, mining and others. This information helps government

The associate editor coordinating the review of this manuscript and approving it for publication was Aneel Rahim.

agencies develop public prevention policies to contain deforestation [5]. In 2018 and 2019, deforestation in the Amazon affected 7,535 km<sup>2</sup> and 10,129 km<sup>2</sup>, respectively. Therefore, the deforested area showed an increase of 34% from 2018 to 2019. In 2019, 84.56% of all deforestation observed in the Amazon occurred in the states of Pará, Mato Grosso, Amazonas, and Rondônia [6]. Although PRODES and TerraClass are large projects and provide very reliable data, they still rely on a significant portion of work done by human intervention. In order to speed up and make the process more efficient, several studies have been done using machine learning tools [7]–[18].

In [11], [12], [15], the authors performed image segmentation using threshold classification. The results obtained by these authors had an accuracy above 90%. Some works classified images through selecting random forests, as seen in [8], [13], [16]. In [13], the authors obtained an accuracy above 98%. In [8] and [16] the accuracy was 85% and 88%, respectively.

The works of [7] and [17] employed unsupervised classification to segment remote sensing images. In [17], the authors used Bayesian Updating of Land Cover: Unsupervised (BULC-U) and obtained an accuracy of 80%. In [7], the authors used K-Nearest Neighbors (KNN) and obtained an accuracy of 75.95%.

According to [19], from the first decade of this century, the neural network was reborn under the new label known as deep learning. Deep learning makes use of computational models with hierarchical architectures composed of multiple processing layers, in order to “learn” representations of data in very different formats: audio, images, and text [20]. According to [19], the large amount of data available in recent years, together with the increase in computational power, allowed the use of deeper architectures that were not previously possible.

The works of [9], [10], [18] used deep learning to detect deforested areas in regions of the Brazilian Amazon. The results obtained by these authors were quite satisfactory. However, the authors evaluated a specific region of the Amazon. Another criticism we make of the previously published studies is that they have not made available the dataset used for benchmarking.

A major challenge in training CNNs with remote sensing images is that the classes are unbalanced. In other words, in a region captured by a satellite image, there is a great imbalance in terms of the area of various types of soils. This problem can lead, in CNN training, to optimization methods performing better in the most frequent classes. This work proposes a method to solve the difficulties in the segmentation of land use and coverage in different regions of the Amazon using CNNs. To circumvent the imbalance problem, the mosaic image technique is proposed in this work for training CNN in remote sensing applications. In this technique, small rectangular samples of agriculture, forest and pasture are extracted from satellite images. From these samples, a large image is

created, with almost the same number of small rectangles of forest, agriculture, and pasture.

Another contribution of this study is that it boosts the results obtained with CNN in soil classification by fusing its output with the output of a segmentation algorithm of homogeneous regions. The fusion results outperform the results obtained only with CNN.

Finally, we compare the results obtained with customized CNN architectures trained from scratch with the results obtained with transfer learning (TL), using shallow training of the following architectures: ResNet50 (RN50), InceptionResnetv2 (IR2), MobileNetv2 (MN2), and Xception (XC) [21]–[24].

Therefore, the main contributions of this paper are the following:

- 1) Proposing customized CNN architectures for semantic segmentation of different types of use and land cover, in different regions of the Amazon.
- 2) Proposing a training, validation and testing methodology of CNN networks, for the remote sensing area, using the mosaic image concept.
- 3) Comparing the results of the proposed CNN architectures, trained from scratch, with the results of transfer learning architectures.
- 4) Proposing a fusion technique using CNN data and data from a segmentation algorithm of homogeneous regions for improving the soil classification in remote sensing applications.
- 5) Availability of a dataset of types of land use/cover (forest, agriculture and pasture) in the Amazon region, which may be used as a base for another research in the Amazon region.

## II. LITERATURE REVIEW

In [7], [16], [17], the authors carried out the classification of images in different regions of the state of Mato Grosso, Brazil. The classes used by the authors were agriculture, forest, and pasture. In [7], the authors used LANDATA-5/TM images as input dataset. The training and testing dataset were built from information from the TerraClass2008 project. The authors used the following classifiers: Decision Tree (DT), Naive Bayes (NB), KNN, Support Vector Machine (SVM), and Optimal Path Forest (OPF). The best results were obtained using the SVM classifier in the region corresponding to image 227/68 (LANDSAT-5/TM image). For this region, an accuracy of 86.04% was obtained. In [16], the authors used LANDSAT-8/OLI and PROBA-V C1 images. The classifier used by the authors was the random forest, and the accuracy obtained was 88%. In [17], the region of the state of Mato Grosso evaluated by the authors included the municipalities of Alto Boa Vista, Querência, Ribeirão Cascalheira, and Canarana. The images used were LANDSAT-5, and the data from the GlobCover project of 2009 were used as the gold standard. The authors used the unsupervised classifier called BULC-U. Using this classifier, the accuracy obtained was 80%.

In [8], the authors used random forests to classify regions of the state of Rondônia in the following classes: deforestation, degradation, and forest. The authors used LANDSAT-5/TM and LANDSAT-8/OLI images. The accuracy obtained was 85%.

In [11], [12], [15], the authors employed a threshold segmentation for remote sensing image classification. In [11], the authors evaluated a region of the state of Mato Grosso with the objective of classifying deforestation and environmental degradation in the region. The region assessed corresponds to an area of scene 226/68 of LANDSAT. LANDSAT-5/TM images from 2005 to 2011 and LANDSAT-8/OLI images from 2013 to 2017 were used. The accuracy obtained by the authors was 94.93%. In [12], the authors used LANDSAT-8/OLI and Sentinel-2 images to classify selective cut in a region of southern Amazonas State, near the village of Santo Antonio do Matupi. The accuracy obtained by the authors was 95.7% for the LANDSAT-8/OLI images and 96.7% for the Sentinel-2 images. In [15], the authors carried out the classification of images in the region of the Carajás mines, located in the south of the state of Pará. The classes defined by the authors were cerrado, forest, water and mining area. The authors used iknos, geoeye and world view images, and the gold standard was generated by the authors themselves through visual interpretation. The accuracy obtained by the authors was 91%.

In [18], the authors used CNN to perform the classification of deforested areas in a region of southern Pará state. The authors used Landsat-8/OLI images and PRODES data as the gold standard. Using CNN U-NET, the authors obtained an accuracy of about 95%.

In [9], the authors mapped deforestation between images one year apart from the other, between 2017 and 2018 and between 2018 and 2019, using a deep learning methodology. The images used were scenes 227/63, 227/65, and 230/65 from LANDSAT-8/OLI, which correspond to regions of the state of Pará and Amazonas, Brazil. The deforestation was extracted from the PRODES project. The authors used three CNN architectures available in the literature: SharpMask, U-Net and ResUnet, and two classical machine learning algorithms: random forest (RF) and multi-layer perceptron (MLP). The best results obtained with the CNN ResUnet had an accuracy and an F1-Score of 99.93% and 94.65%, respectively.

In [10], the authors evaluated deep learning strategies for automatic deforestation detection. The experiments used two Landsat-8/OLI images, acquired at different dates. The first study area corresponds to a small region of the Amazon biome, located, more specifically, in the state of Pará, Brazil. The state of Pará comprises 26% of the Brazilian Amazon, and most of it is covered by dense tropical forest. The reference data used in this experiment refer to the deforestation that occurred between August 2016 and July 2017. The second study area is a small region of the Brazilian Cerrado biome, located in the state of Maranhão, Brazil. For this case, the deforestation took place in the years 2017

and 2018. The authors obtained an accuracy and F1-Score of 95% and 63%, respectively, in the Amazon, and 97% and 78%, respectively, in the Cerrado. The last results, obtained in the Cerrado reached higher percentages than those of the Amazon. According to the authors, this is due to the pattern of deforestation in the Cerrado biome that removes all vegetation, in contrast with the Amazon, where vegetation is not completely removed.

The papers presented in this review showed that machine learning techniques have great potential to classify remote sensing images and to perform automatic segmentation in the most varied types of vegetation cover in the Amazon region. In [9], [10], [18], the authors used deep learning techniques to classify deforestation regions in the Amazon. The results achieved were quite satisfactory. However, in each work, a specific region of the Amazon was used. Another criticism we make of the previously published works is that they have not made available the dataset used for the benchmark. This limits the diversification of solutions to the real world. No papers were found that presented an adequate model in deep learning for land use/cover classification for the Amazon region, encompassing the many varieties of existing environments.

In this work we make available the dataset used, which covers different areas of the Amazon. These areas cover the Brazilian states of Amazonas, Mato Grosso, Pará, and Rondônia.

### III. MATERIALS AND METHODS

This work aims proposing a method to solve the difficulties in the segmentation of land use and coverage of different regions of the Amazon region. Figure 1 shows a flowchart of the methodology used for segmentation of land use and coverage. As shown, there are two branches: The left one corresponds to CNN processing and the right one to segmentation with homogeneous regions. In the first block of the left branch, the band's selection is performed. After, in the semantic segmentation block, the CNN performs the identification of the different soil coverage. In the right branch, we perform the segmentation of homogeneous regions. At the end, uniting the two branches, we have the fusion block. In the following sections, each block of this flowchart will be explained in detail.

#### A. LANDSAT IMAGE

In this work, LANDSAT-8/OLI images from the Amazon region, from 2014, were used. These images are available for free at [25]. As shown in Table 1, the images used cover the Brazilian states of Amazonas, Mato Grosso, Pará and Rondônia.

For generating the gold-standard images for the classifier, the results of the TerraClass project, from 2014, were used. The data were acquired free of charge on the INPE website [4]. The data generated in the TerraClass project delimit the regions of the Brazilian Amazon in the following classes: forest, agriculture, pasture, unobserved area, urban area, mining, others, non-forest and hydrography.

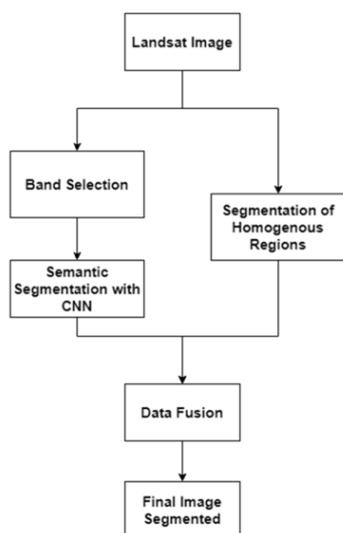


FIGURE 1. Flowchart of methodology used.

TABLE 1. List of images used.

O/P*	Date	O/P*	Date	O/P*	Date
001/65	08/12/2014	229/66	08/09/2014	231/66	08/23/2014
001/66	08/12/2014	230/64	08/16/2014	232/64	08/30/2014
001/67	08/12/2014	230/65	08/16/2014	232/65	08/30/2014
002/65	08/03/2014	230/66	07/15/2014	233/65	08/21/2014
002/66	08/19/2014	231/64	09/08/2014	233/66	08/21/2014
229/65	08/09/2014	231/65	08/23/2014	233/67	08/21/2014
223/68	08/15/2014	226/67	08/04/2014	227/71	08/11/2014
224/67	08/06/2014	226/68	08/04/2014	228/67	08/18/2014
224/68	07/21/2014	226/69	08/04/2014	228/68	08/18/2014
224/69	08/06/2014	226/70	08/04/2014	228/69	07/01/2014
224/70	08/22/2014	226/71	08/20/2014	228/70	08/18/2014
225/67	08/13/2014	227/67	07/26/2014	228/71	08/18/2014
225/69	08/13/2014	227/68	08/11/2014	229/68	08/09/2014
225/70	08/29/2014	227/69	08/11/2014	229/70	08/09/2014
225/71	08/13/2014	227/70	08/11/2014	229/71	08/09/2014
222/62	10/27/2014	224/63	07/05/2014	225/62	10/27/2014
223/62	09/16/2014	224/64	07/05/2014	225/65	10/27/2014
223/66	07/30/2014	224/65	08/22/2014	226/62	08/20/2014
223/67	08/15/2014	224/66	08/06/2014	227/65	08/11/2014
230/68	08/16/2014	231/67	08/23/2014	232/66	08/14/2014
230/69	07/31/2014	231/68	08/23/2014	232/67	08/14/2014

\*Orbit/Point

For this work, information on forest, pasture and agriculture areas were used corresponding to a problem with three classes. The classes of unobserved area, urban area, mining and others were not considered because they represent a very small percentage area, and it would not be possible to extract a significant amount of samples to be used in the process of semantic classification with CNNs. The nonforest and hydrography areas were not used because they are part of an INPE exclusion mask. Thus, these regions are not considered in the systematic mapping.

### B. BAND SELECTION

According to [26] the bands Blue (B2), Green (B3), Red (B4), Near Infrared (B5), Shortwave Infrared 1 (B6), and Shortwave Infrared 2 (B7) are the most suitable for vegetation analysis. According to [27], B4, B5, and B6 is the best



FIGURE 2. Image patches in color composition of bands B6-B5-B4 corresponding to: (a) forest (b) pasture (c) agriculture.

TABLE 2. Splitting image patches.

Class	Training	Validation	Test
Agriculture	2,000	1,000	1,000
Forest	75,000	75,000	75,000
Pasture	3,000	1,500	1,500

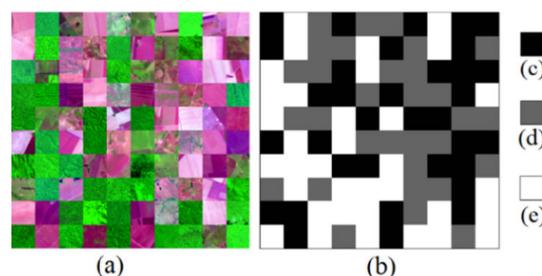


FIGURE 3. Example of mosaic image. In (a) is shown a mosaic image in color composition of bands B6-B5-B4, composed of agricultural, forest and pasture patches, in (b) the gold standard image, and in (c), (d) and (e) we have the gold standard for each image patch, corresponding to agriculture, pasture and forest area, respectively.

combination of three bands for remote sensing applications whose purpose is to perform soil classification. Thus, two versions of image datasets were assembled: the first with the six bands (B2, B3, B4, B5, B6, and B7) and the second with three bands (B4, B5, B6).

Using the reference data from the TerraClass project for the year 2014,  $40 \times 40$  pixel-sized patches were generated for each class. Figure 2 presents three examples of these patches, corresponding to forest, pasture, and agriculture. Using the LANDSAT-8/OLI images, 4,000 agricultural patches, 225,000 forest patches, and 6,000 pasture patches were generated. These patches were divided into three sets: training, validation and testing, as shown in Table 2.

Using the patches previously described, large images, defined as mosaic images, are generated. Each mosaic image has dimensions of  $400 \times 400$  pixels and is formed by 100 patches. Each unit of the mosaic image is randomly generated selecting a patch of agriculture, forest, or pasture. Each mosaic image has a gold standard image. For building the gold standard image, the pixels corresponding to forest regions were marked with the value 255, the pixels corresponding to pastures region were marked with the value 100, and the pixels corresponding to agriculture regions were marked with the value 1. Figure 3 shows an example of a mosaic image and the corresponding gold standard. It can be observed that, out of the 100 patches, 34 are from agriculture, 35 from forest, and 31 from pasture regions.

Two versions of image datasets were generated with a total of 4,000 mosaic images each. For each version, 70% of the images (2800) were destined for the training set, 15% (600),

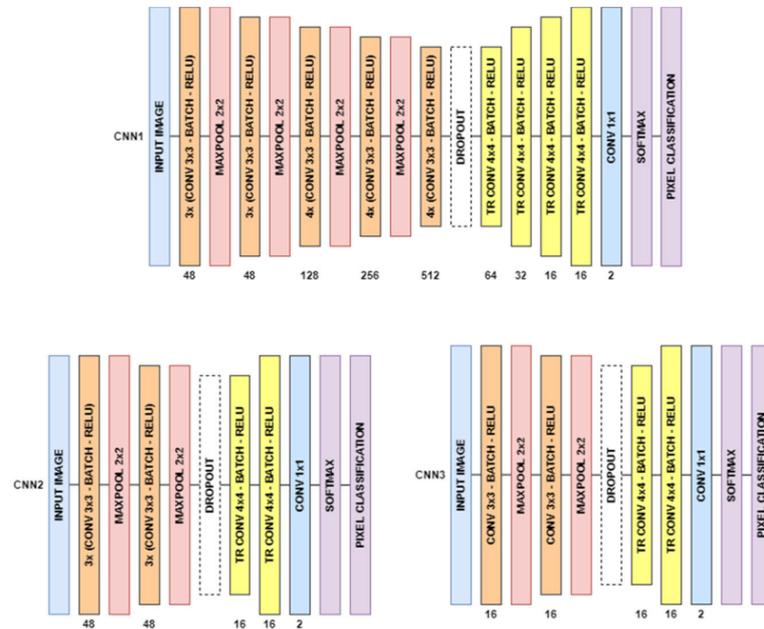


FIGURE 4. CNN architectures used in this work.

TABLE 3. Mosaic image dataset generated with the quantity for training, validation, and testing sets.

Class	Bands	Total	Training	Validation	Test
1	B2, B3, B4, B5, B6, B7	4,000	2,800	600	600
2	B4, B5, B6	4,000	2,800	600	600

for the validation set and 15% (600), for the test set. The first version consists of 6-band images: B2, B3, B4, B5, B6 and B7. The second version is composed of three-band images: B4, B5, and B6. Table 3 presents the image datasets generated used in this work.

C. CNN SEMANTIC SEGMENTATION

Three CNN architectures (Figure 4) used in this work were based on the architecture proposed in the work of [28]. In this work, the authors performed the lumen segmentation in intravascular optical coherence tomography (IVOCT) images.

As shown in Figure 4, CNN1 has four sub-sampling steps (maxpooling). Before and after the first subsampling step, there are three sequences of  $3 \times 3$  convolutive layers, batch normalization layer, and ReLU. After the second, third and fourth subsampling steps, there are four sequences of  $3 \times 3$  convolutive layers. In addition, CNN1 is composed of four oversampling steps. CNN2 has two subsampling steps (maxpooling) and two oversampling steps. Before each subsampling, there are three sequences of  $3 \times 3$  convolutive layers, a batch normalization layer, and ReLU. Finally, CNN3, like CNN2, also has two subsampling steps (maxpooling) and two oversampling steps. However, before each subsampling there is only one sequence of  $3 \times 3$  convolutive layer, batch normalization layer, and ReLU.

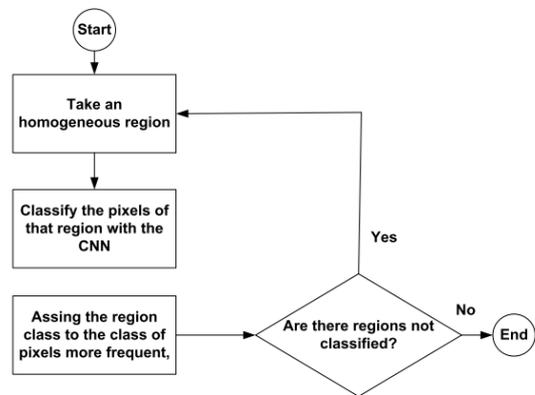
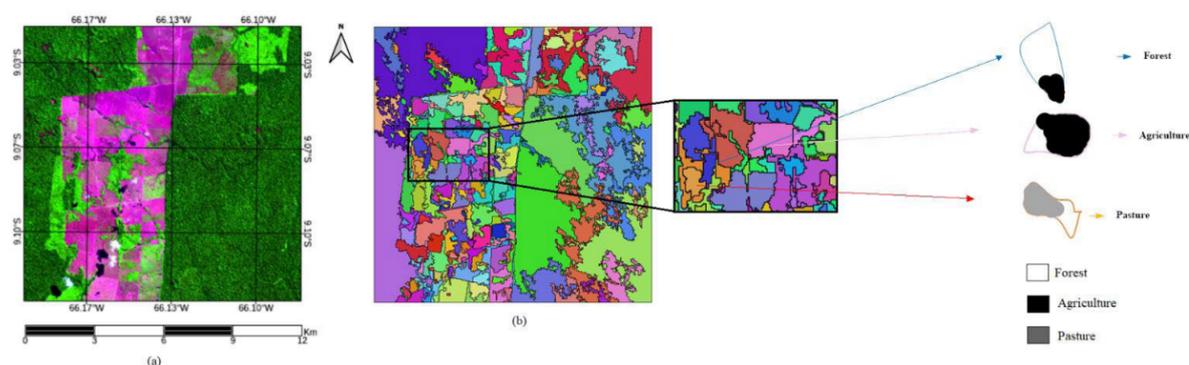


FIGURE 5. Flowchart of the fusion process.

The optimization methods evaluated were SGDM, ADAM and RMSProp. For each of these methods the following methods were used to improve generalization: dropout, L<sub>2</sub> regularization and dropout with L<sub>2</sub> regularization. Mosaic images were used as the input dataset. Two groups of simulations were done. Both groups are comprised of 36 simulations (3 architectures  $\times$  3 optimization methods  $\times$  4 improving generalization methods). In the first group, six-band mosaic images were used as CNN inputs, while in the second group, three-band mosaic images were used as CNN inputs.

For each dataset version, the model (architecture + optimization method + improving generalization method) with best performance, evaluated in terms of accuracy in the validation set was selected.



**FIGURE 6.** Illustration of fusion process: (a) original image; (b) result image of homogeneous region segmentation algorithm. Inside the small rectangle, three small regions are selected and magnified. The upper region has more gray pixels so it is classified as pasture. The central region has more black pixels so it is classified as agriculture, and because there are more white pixels in the lower region, it is classified as forest.

The CNN performances were compared with pre-trained network performances, using transfer learning, with shallow training. The pre-trained networks used in this work were ResNet50, InceptionResnetv2, MobileNetv2, and Xception combined with SGDM, ADAM, and RMSProp optimization methods, six-band mosaic images as input. A total of twelve experiments were performed.

A Dell Precision 5820 Workstation with Intel Xeon W-2102 2.9GHz, 64GB RAM, Windows 10, Matlab 2019a and NVIDIA Quadro GV100 32GB was used in the experiments. The following parameters were used for the CNN training: initial learning rate = 0.001, learning rate dropout factor = 0.5, number of epochs = 200, batch size = 2, dropout layer parameter = 0.3,  $L_2$  smoothing factor = 0.001.

To evaluate the results, the Global Accuracy (GA), Average Accuracy (AA), Jaccard Similarity Coefficient (J), Weighted Jaccard Similarity Coefficient (WJ) and F1 Score (F1) were calculated [29].

In this work, CNN was trained using mosaic images consisting of  $40 \times 40$ -pixel patches that tend to generate images cut out with patches of the same size. To neutralize this trend and improve classification/segmentation, a fusion process using the CNN output and a multi-resolution homogeneous region segmentation algorithm was used. This algorithm and the fusion process are described in the sequences.

#### D. SEGMENTATION OF HOMOGENOUS REGIONS

The segmentation of homogeneous regions consists of creating an image with regions of the same homogeneity based on the value of the pixel of the image. In this work, the algorithm called a multiresolution segmentation proposed by [30] was used, which performs a general segmentation based on homogeneity definitions combined with local and global optimization techniques. A scale parameter is used to control the average size of the image object. According to the authors, multi-resolution segmentation is a technique for merging regions starting with one-pixel objects. In several subsequent steps, smaller image objects are merged into larger objects.

Throughout this pairwise grouping process, the underlying optimization procedure minimizes the heterogeneity of the resulting image objects. At each step, this pair of adjacent image objects is merged, resulting in the smallest growth of the defined heterogeneity. If the smallest growth exceeds the limit defined by the scale parameter, the process stops. In this way, we can say that multiresolution segmentation is a local optimization procedure. In this work, the use of this algorithm was performed through the e-Cognition Software. The scale factor used was 200.

#### E. DATA FUSION

The data fusion consists of using the image resulting from the multi-resolution algorithm and the image resulting from the semantic classification made by the CNN model. The data fusion used in this work is illustrated by means of the flowchart presented in Figure 5. For each region generated by the homogeneous region segmentation algorithm, we evaluated the pixel classification by the CNN. The most frequent class of pixels in a region is considered the class of that region. Figure 6 illustrates this process: Figure 6(a) shows an original image; Figure 6(b) shows the results of the homogeneous region algorithm applied in Figure 6(a). In Figure 6(b), inside a black rectangle, three small neighbor regions, which are magnified in the right side, are selected. These selected regions are filled with a hypothetical pixel classification by the CNN. The upper region has more white pixels, so it is classified as forest. The central region has more black pixels, so it is classified as agriculture, and the lower region has more gray pixels, so it is classified as pasture.

### IV. RESULTS AND DISCUSSION

#### A. CNN SIMULATION

Tables 4 and 7 present the performance metrics of the three CNN architectures proposed in this work using the six-band and three-band mosaic image dataset, respectively. For the six-band dataset, the model with the best performance was the one in which CNN1 was used, together with the RMSProp optimization method and without using methods for

**TABLE 4. Performance of CNN1 with six-band dataset.**

Experiment	GA (%)	AA (%)	J (%)	WJ (%)	F1 (%)	Train Time (h)
CNN1/SGDM	99.83	99.83	99.65	99.65	99.45	94.14
CNN1/SGDM/Dropout	99.87	99.87	99.75	99.75	99.64	58.99
CNN1/SGDM/L <sub>2</sub>	99.93	99.93	99.85	99.85	99.82	99.50
CNN1/SGDM/Dropout/L <sub>2</sub>	99.94	99.94	99.88	99.88	99.85	96.09
CNN1/ADAM	99.87	99.87	99.75	99.75	99.70	142.30
CNN1/ADAM/Dropout	99.95	99.95	99.89	99.89	99.91	130.46
CNN1/ADAM/L <sub>2</sub>	99.84	99.84	99.68	99.68	99.76	139.95
CNN1/ADAM/Dropout/L <sub>2</sub>	99.84	99.84	99.68	99.68	99.76	138.28
CNN1/RMSProp	99.98	99.98	99.96	99.96	99.95	123.41
CNN1/RMSProp/Dropout	99.96	99.96	99.93	99.93	99.91	133.00
CNN1/RMSProp/L <sub>2</sub>	99.89	99.89	99.78	99.78	99.84	134.26
CNN1/RMSProp/Dropout/L <sub>2</sub>	99.91	99.91	99.81	99.81	99.85	134.34
CNN2/SGDM	97.52	97.53	95.21	95.20	91.30	52.89
CNN2/SGDM/Dropout	97.48	97.49	95.13	95.12	92.08	50.01
CNN2/SGDM/L <sub>2</sub>	98.33	98.33	96.73	96.72	93.29	40.22
CNN2/SGDM/Dropout/L <sub>2</sub>	98.20	98.20	96.49	96.48	93.60	74.84
CNN2/ADAM	98.92	98.92	97.86	97.86	94.64	95.02
CNN2/ADAM/Dropout	98.79	98.80	97.63	97.62	94.63	52.68
CNN2/ADAM/L <sub>2</sub>	99.36	99.36	98.73	98.72	96.47	94.95
CNN2/ADAM/Dropout/L <sub>2</sub>	99.11	99.11	98.24	98.24	95.76	91.50
CNN2/RMSProp	98.90	98.90	97.83	97.83	94.50	92.61
CNN2/RMSProp/Dropout	98.75	98.75	97.54	97.54	94.50	51.62
CNN2/RMSProp/L <sub>2</sub>	99.23	99.23	98.47	98.47	95.92	105.30
CNN2/RMSProp/Dropout/L <sub>2</sub>	99.02	99.02	98.06	98.06	95.39	81.47
CNN3/SGDM	95.38	95.40	91.33	91.31	86.01	45.21
CNN3/SGDM/Dropout	94.40	94.41	89.63	89.60	86.44	48.49
CNN3/SGDM/L <sub>2</sub>	95.29	95.31	91.17	91.15	86.00	32.75
CNN3/SGDM/Dropout/L <sub>2</sub>	93.88	93.88	88.73	88.70	82.67	42.50
CNN3/ADAM	96.32	96.34	93.02	93.00	86.73	67.73
CNN3/ADAM/Dropout	95.70	95.71	91.88	91.86	87.98	49.92
CNN3/ADAM/L <sub>2</sub>	96.30	96.31	92.97	92.95	87.27	57.11
CNN3/ADAM/Dropout/L <sub>2</sub>	95.73	95.75	91.96	91.94	88.13	68.47
CNN3/RMSProp	96.20	96.21	92.80	92.78	86.59	67.39
CNN3/RMSProp/Dropout	95.30	95.31	91.19	91.17	87.19	49.50
CNN3/RMSProp/L <sub>2</sub>	96.17	96.18	92.74	92.72	86.51	57.40
CNN3/RMSProp/Dropout/L <sub>2</sub>	73.32	73.24	58.55	58.54	50.24	68.50

improving generalization. This model obtained a GA of 99.98%. For the three-band dataset, the model with the best performance was the one in which CNN1 is used together with the RMSProp optimization method and dropout method to improve generalization. This model obtained a GA of 99.92%. Tables 5 and 8 present the confusion matrix for the best CNN model for the six-band and three-band dataset, respectively. From Tables 5 and 8, we observe that these models presented a greater classification error in regions where the gold-standard identified areas as forest but were classified as pasture. These CNN models were evaluated in a 5-fold cross-validation methodology. The dataset used for these experiments was the training and test dataset. Tables 6 and 9 present the performance obtained for this experiment using the six-band and three-band dataset, respectively. For the 5-fold cross-validation methodology, GA of 99.97% and 99.86% were obtained for the six- and three-band dataset, respectively.

### B. SIMULATION RESULTS USING TRANSFER LEARNING

Table 10 presents the results obtained using pre-trained nets. The best performance was achieved using the Resnet50 network and the ADAM method. For this model, a GA of 94.50% was obtained. However, this performance was inferior to the CNN1 model with RMSProp optimization method. This model presented a GA of 99.98%.

**TABLE 5. Confusion matrix for the best CNN model (CNN1/RMSProp) using six-band dataset.**

	Agriculture	Pasture	Forest
Agriculture	1	2.19E-07	0
Pasture	9.93E-05	0.99	1.12E-04
Forest	3.14E-08	4.07E-04	0.99

**TABLE 6. CNN1/RMSProp model performance using a 5-fold cross-validation methodology and six-band dataset.**

Fold	GA(%)	AA(%)	J(%)	WJ(%)	F1(%)
1	99,99	99,99	99,99	99,99	99,99
2	99,99	99,99	99,97	99,97	99,97
3	99,89	99,89	99,79	99,79	99,78
4	99,99	99,99	99,99	99,99	99,99
5	99,99	99,99	99,97	99,97	99,98
Average	99,97	99,97	99,97	99,97	99,94

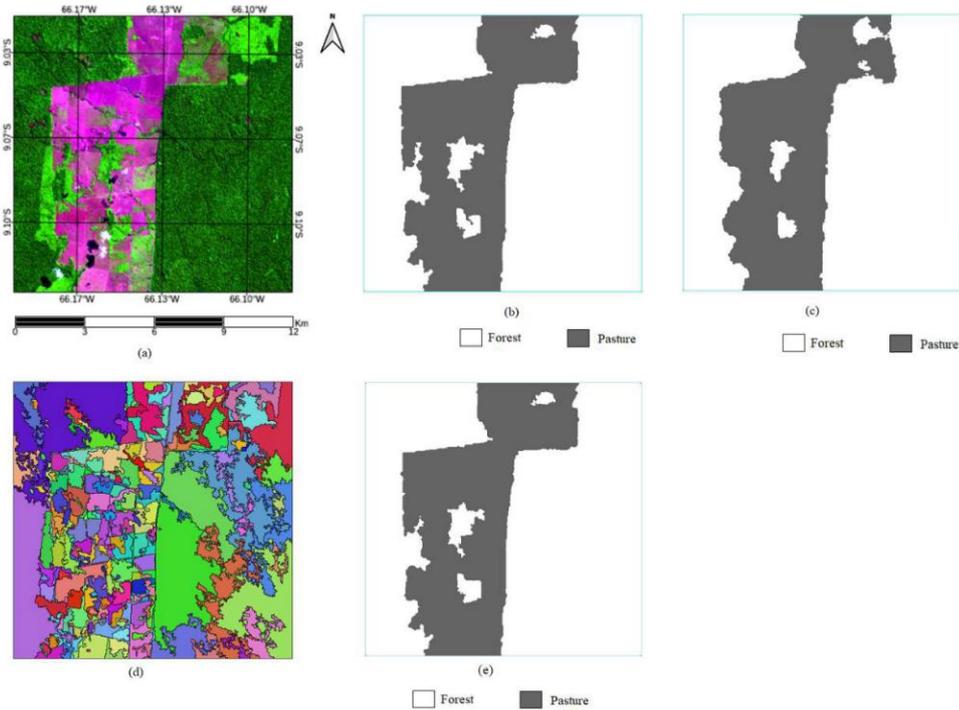
**TABLE 7. Performance of CNN1 with three band dataset.**

Experiment	GA(%)	AA(%)	J(%)	WJ(%)	F1(%)
CNN1/SGDM	99.75	99.75	99.50	99.50	99.30
CNN1/SGDM/Dropout	99.52	99.53	99.06	99.05	98.84
CNN1/SGDM/L <sub>2</sub>	99.00	99.00	98.02	98.02	97.95
CNN1/SGDM/Dropout/L <sub>2</sub>	99.73	99.73	99.45	99.45	99.46
CNN1/ADAM	99.89	99.89	99.78	99.78	99.77
CNN1/ADAM/Dropout	99.63	99.63	99.26	99.26	99.40
CNN1/ADAM/L <sub>2</sub>	98.51	98.52	97.08	97.08	97.56
CNN1/ADAM/Dropout/L <sub>2</sub>	98.70	98.70	97.44	97.43	97.92
CNN1/RMSProp	99.89	99.89	99.79	99.79	99.78
CNN1/RMSProp/Dropout	99.92	99.92	99.84	99.84	99.77
CNN1/RMSProp/L <sub>2</sub>	99.38	99.38	98.76	98.76	98.74
CNN1/RMSProp/Dropout/L <sub>2</sub>	97.38	97.40	94.93	94.92	95.93
CNN2/SGDM	95.98	96.00	92.41	92.39	86.85
CNN2/SGDM/Dropout	94.90	94.92	90.49	90.46	87.11
CNN2/SGDM/L <sub>2</sub>	96.37	96.38	93.09	93.08	87.98
CNN2/SGDM/Dropout/L <sub>2</sub>	95.95	95.96	92.33	92.31	89.08
CNN2/ADAM	97.16	97.17	94.54	94.52	89.43
CNN2/ADAM/Dropout	97.27	97.27	94.74	94.73	89.83
CNN2/ADAM/L <sub>2</sub>	97.10	97.10	94.42	94.41	89.29
CNN2/ADAM/Dropout/L <sub>2</sub>	96.82	96.84	93.91	93.89	90.23
CNN2/RMSProp	97.47	97.48	95.11	95.10	89.52
CNN2/RMSProp/Dropout	96.81	96.81	93.88	93.87	88.31
CNN2/RMSProp/L <sub>2</sub>	97.26	97.27	94.73	94.72	90.25
CNN2/RMSProp/Dropout/L <sub>2</sub>	96.53	96.55	93.39	93.37	90.27
CNN3/SGDM	93.09	93.12	87.41	87.38	81.95
CNN3/SGDM/Dropout	91.84	91.86	85.35	85.32	82.74
CNN3/SGDM/L <sub>2</sub>	93.16	93.18	87.52	87.49	82.07
CNN3/SGDM/Dropout/L <sub>2</sub>	92.25	92.28	86.01	85.97	83.94
CNN3/ADAM	94.37	94.39	89.57	89.54	81.95
CNN3/ADAM/Dropout	93.08	93.10	87.37	87.34	83.36
CNN3/ADAM/L <sub>2</sub>	94.07	94.08	89.03	89.00	82.92
CNN3/ADAM/Dropout/L <sub>2</sub>	92.94	92.97	87.15	87.12	84.47
CNN3/RMSProp	94.22	94.24	89.32	89.30	81.63
CNN3/RMSProp/Dropout	93.22	93.25	87.61	87.58	83.72
CNN3/RMSProp/L <sub>2</sub>	93.85	93.87	88.66	88.64	83.21
CNN3/RMSProp/Dropout/L <sub>2</sub>	92.43	92.45	86.32	86.29	81.60

### C. CLASSIFICATION/SEGMENTATION OF IMAGES FROM AMAZON REGIONS

To evaluate the methodology proposed in this work, some images from the Amazon region were selected to perform the segmentation/classification.

These images were classified using only the model CNN2 + ADAM optimization method + L<sub>2</sub> regularization and using the fusion process described in sections III-D and III-E. In Figure 7, we have an example of the classification of a LANDSAT-8/OLI image using this process. As shown



**FIGURE 7.** LANDSAT-8/OLI images of Amazon regions. (a) original image displayed in color composition of bands B6-B5- B4; (b) gold standard; (c) semantic classification made by CNN, with an accuracy of 96.84%. (d) output image resulting from the multi-resolution segmentation; (e) result of classified image after data fusion process, with an accuracy of 99.8%.

**TABLE 8.** Confusion matrix for the best CNN model (CNN1/RMSProp) using the six band dataset.

	Agriculture	Pasture	Forest
Agriculture	0.99	3.05E-04	8.82E+08
Pasture	3.03E-04	0.99	8.72E-04
Forest	2.67E-04	7.14E-04	0.99

**TABLE 9.** CNN1/RMSProp model performance using a 5-fold cross-validation methodology and six-band dataset.

Fold	GA(%)	AA(%)	J(%)	WJ(%)	F1(%)
1	99.86	99.86	99.72	99.72	99.65
2	99.87	99.87	99.74	99.74	99.71
3	99.76	99.76	99.53	99.52	99.59
4	99.87	99.87	99.74	99.74	99.72
5	99.94	99.94	99.89	99.89	99.87
Average	99.86	99.86	99.72	99.72	99.71

in Figure 7(c), in the results of CNN classification, there are some regions incorrectly classified as forest.

In Figure 7(e), which shows the result of the fusion process, these misclassifications are minimized. The accuracy obtained using only CNN semantic segmentation was 96.84%. After the data fusion process, the accuracy increased to 99.81%.

We can observe that the proposed methodology presented impressive performances. The best results were obtained using the CNN2 model, the ADAM optimization method, and  $L_2$  regularization for improving generalization. The data fusion process has increased the image classification accuracy, as shown in Figure 7. Table 11 presents a comparison of

**TABLE 10.** Performance using transfer learning.

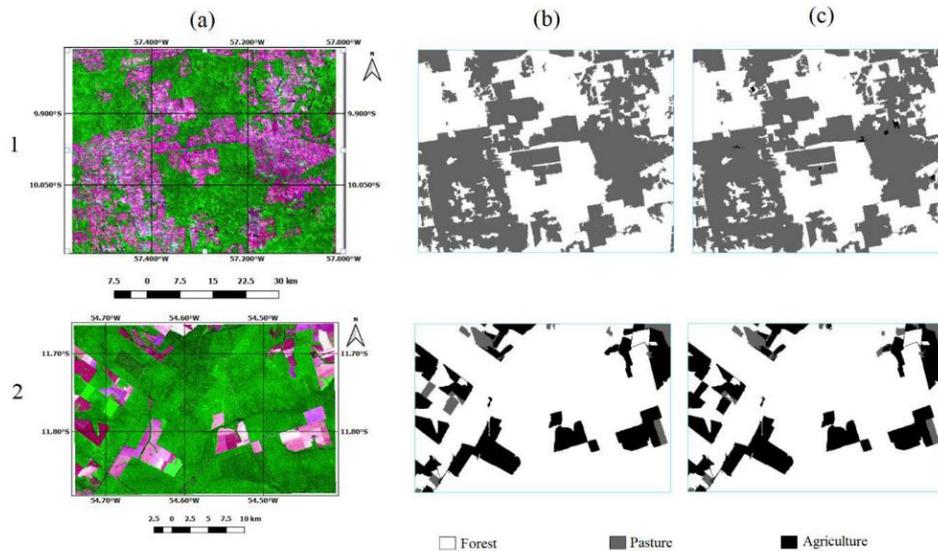
Fold	GA(%)	AA(%)	J(%)	WJ(%)	F1(%)
RN50/SGDM	92.26	92.28	86.03	86.00	86.31
RN50/ADAM	94.50	94.51	89.77	89.75	87.73
RN50/RMProp	94.17	94.19	89.21	89.18	88.70
IR2/SGDM	89.77	89.79	81.84	81.81	80.44
IR2/ADAM	87.62	87.66	78.14	78.11	77.75
IR2/RMSProp	88.23	88.25	79.38	79.35	79.91
MN2/SGDM	77.87	77.93	64.00	63.95	57.17
MN2/ADAM	77.41	77.44	63.80	63.76	56.27
MN2/RMSProp	77.83	77.86	64.27	64.23	58.92
XC/SGDM	92.25	92.27	85.96	85.93	82.67
XC/ADAM	91.34	91.36	84.41	84.37	83.02
XC/RMSProp	91.18	91.20	84.21	84.17	82.21

**TABLE 11.** Performance of image segmentation shown in Figure 7 using only CNN2/ADAM/ $L_2$  and the data fusion process.

Segmentation Method	GA (%)	AA (%)	J (%)	WJ (%)	F1 (%)
CNN2/ADAM/ $L_2$	96.84	96.42	93.09	93.90	81.08
Data Fusion Process	99.81	99.82	99.56	99.61	98.09

the image segmentation performance of Figure 7 using only CNN model and the data fusion process.

In Figure 8, we show two LANDSAT-8/OLI images of Amazon regions with their respective gold standard, and the classified image using the CNN2 + ADAM optimization method +  $L_2$  regularization and data fusion process. Image 1 corresponds to a region of scene 228/67, and image 2 corresponds to a region of scene 226/68. The accuracies



**FIGURE 8.** LANDSAT-8/OLI images of Amazon regions. (a) original images displayed in color composition B6-B5-B4; (b) gold-standard image; (c) images classified by CNN2/ADAM/L<sub>2</sub> model and fusion process. The accuracy obtained for images 1 and 2 was 99.63%, 99.08%, respectively.

**TABLE 12.** Datasets assembled using images evaluated by [9] and [10].

Dataset	Obit/Point	Date	Train	Validation
1 [9]	227/63	07/08/2017	1400	600
	227/63	07/21/2017		
	227/65	08/18/2017		
	227/65	07/21/2018		
2 [10]	230/65	06/21/2017	1400	600
	230/65	06/24/2018		

**TABLE 13.** Performance of three best models of each CNN using the same regions used by [9] and [10].

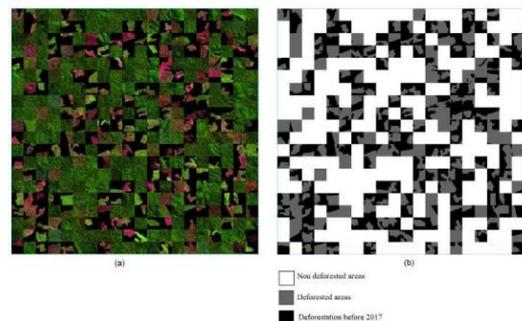
DT*	Experiment	GA(%)	AA(%)	J(%)	WJ(%)	F1(%)
1	CNN1/RMSProp	98.48	97.98	96.26	97.03	99.68
	CNN2/ADAML <sub>2</sub>	98.44	97.94	96.89	96.93	98.68
	CNN3/RMSPropL <sub>2</sub>	97.96	97.30	95.91	95.99	98.14
2	CNN1/RMSProp	98.59	98.10	96.26	97.25	99.90
	CNN2/ADAML <sub>2</sub>	99.55	99.32	98.95	99.10	99.60
	CNN3/RMSPropL <sub>2</sub>	99.52	99.33	98.85	99.05	99.55

\*Dataset

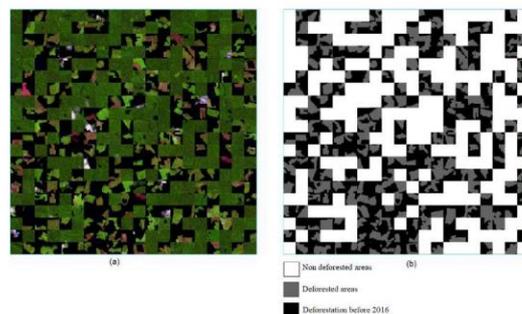
obtained for images 1 and 2 were 99.63% and 99.08%, respectively.

**D. EVALUATION OF METHODOLOGY IN REGIONS USED BY [9], [10]**

The methodology presented in this work was also evaluated in the regions evaluated by [17], [18]. Two datasets with 2000 mosaic images each were assembled using the images evaluated by [17], [18], respectively. Table 12 shows the images used for the construction of each dataset and the division of the set for training and validation. Each mosaic image is 400 × 400-pixel size and is made up of patches of 20 × 20-pixel size. Each image patch corresponds to a region of deforestation or nondeforestation. Figures 9 and 10 show examples of mosaic images and the corresponding

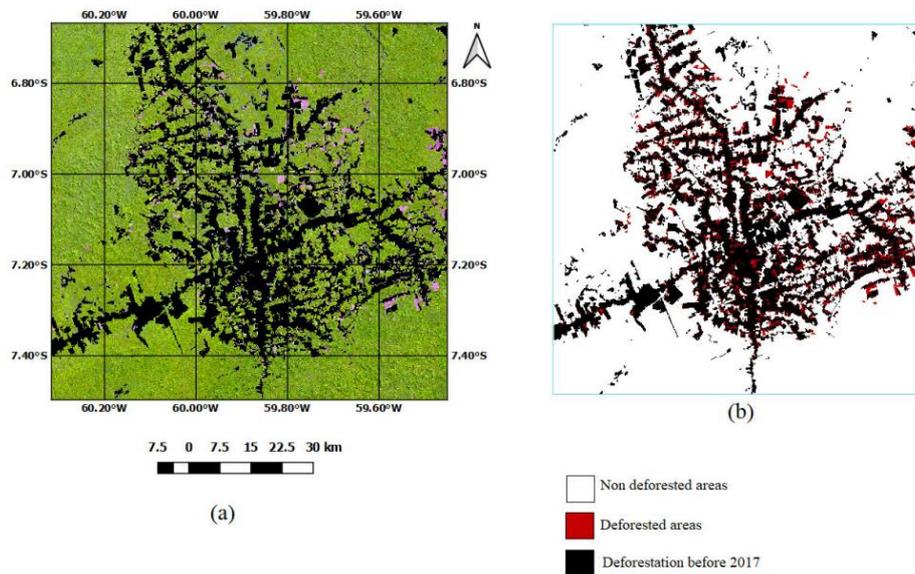


**FIGURE 9.** Example of a mosaic image for the region used by [9]. In (a), the mosaic image is displayed in color composition B6-B5-B4 and (b) the corresponding gold standard.

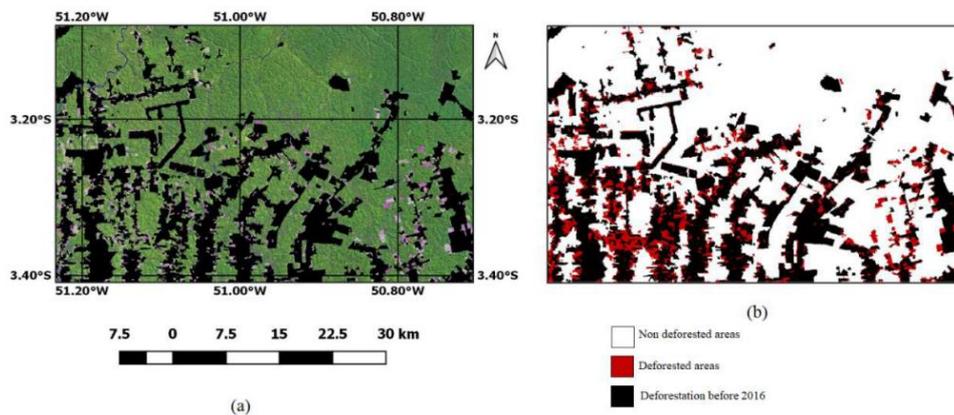


**FIGURE 10.** Example of a mosaic image for the region used by [10]. In (a), the mosaic image is displayed in color composition B6-B5-B4 and (b) corresponding gold standard.

gold standards for datasets 1 and 2, respectively. Table 13 resents the performance of the CNN models for each of



**FIGURE 11.** LANDSAT-8/OLI images of Amazon regions. (a) Image of a region (3200 × 3200pixels) of the LANDSAT 230/65 scene evaluated by [9] displayed in color composition B6-B5-B4. (b) shows classified image, with accuracy of 99.91%.



**FIGURE 12.** LANDSAT-8/OLI images of Amazon regions. In (a), image of a region (2000 × 1200pixels) of the LANDSAT 225/62 scene evaluated by [10] displayed in color composition B6-B5-B4. In (b), image is classified with accuracy of 99.94%.

these datasets. The best model of each CNN presented in Table 6 was used.

The CNN3/RMSProp/ $L_2$  model and the data fusion process were used to classify images presented in Figure 11 and 12. Figure 11 corresponds to a region evaluated by [17]. For this image, the obtained accuracy was 99.91% and the value of F1-Score was 99.94%. The accuracy value was very similar to the value obtained by [17]. The F1-Score was higher than the value obtained by the authors, which was 94.65%. Figure 12 corresponds to a region evaluated by [18]. For this image the obtained accuracy was 99.94%, and the value of the F1-Score was 99.92%. The accuracy value was higher than the best results obtained by [18], which was 95%. In regard

to the value of the F1-Score, the value was higher than the one obtained by the authors, which was 63%.

## V. CONCLUSION

This study proposed a methodology using deep learning to solve the difficulties in the segmentation of land use and land cover for the Amazon region for the pasture, agriculture and forest classes. The proposed methodology consisted of a data fusion process performed through semantic segmentation using CNN and a homogeneous region algorithm. LANDSAT-8/OLI optical images of the Brazilian Amazon region were used to validate the proposed methodology. The reference data used were the results obtained from the

TerraClass project of INPE in 2014. The CNN architectures used in this study were based on the architecture proposed in the work of [28]. Two datasets were evaluated: the first using six bands and the second with three bands. The results for the six-band dataset outperforms the one obtained with a three-band dataset.

In this study, CNN was trained using mosaic images consisting of  $40 \times 40$  pixel patches that tend to generate images cut out with patches of the same size. To neutralize this trend and improve classification/segmentation, a fusion process using the CNN output and a multi-resolution homogeneous region segmentation algorithm was used.

The proposed method was compared to some pretrained CNN architectures using transfer learning. The comparison demonstrated the effectiveness of our proposed method against these well-known CNN architectures in the task of remote sensing applications. The proposed method proved to be more appropriate for this specific application. For the same region evaluated by [9], the results were practically the same. Regarding the results obtained by [10], our method obtained better results.

The main contributions of this work to soil classification in remote sensing applications were twofold. The first one was to propose the mosaic image technique for training CNN architectures. The second one was a fusion technique to improve the soil classification, using two types of data: the CNN output image and output image of a homogeneous region segmentation algorithm. Another contribution is making the datasets used in this work available, which may be used as a base for other research in the Amazon Region. Data will be available upon request. Given the results presented, we conclude that the methodology proposed in this work proved to be quite efficient in performing the task of segmentation/classification of remote sensing images for regions of the Amazon.

For future studies, we propose using more varied classes of land use, in addition to using other types of remote sensing images, for example, synthetic aperture radar (SAR) images.

## REFERENCES

- [1] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, vol. 1, 2nd ed. 2014.
- [2] F. F. Sabins, *Remote Sensing: Principles and Applications*, 2008.
- [3] INPE. (2019). *Satellite Monitoring of Deforestation of the Brazilian Amazon Forest*. Accessed: Sep. 11, 2020. [Online]. Available: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>
- [4] INPE. (2019). *National Institute of Space Research—Projects and Research—TerraClass*. Accessed: Aug. 17, 2019. [Online]. Available: [http://www.inpe.br/cra/projetos%5C\\_pesquisas/dados%5C\\_terraclass.php](http://www.inpe.br/cra/projetos%5C_pesquisas/dados%5C_terraclass.php)
- [5] A. Noma, T. S. Korting, and L. M. G. Fonseca, "Uma comparação entre classificadores usando Regiões e perfis EVI para agricultura," in *Proc. 15th Simpósio Brasileiro de Sensoriamento Remoto. Anais*, vol. 16, Apr. 2013, pp. 2250–2257. [Online]. Available: <http://urlib.net/dpi.inpe.br/marte2/2013/05.28.23.34>.
- [6] INPE. (2019). *The Consolidated Rate of Deforestation by Shallow Cut for the Nine States of Legal Amazon (AC, AM, AP, MA, MT, PA, RO, RR e TO) in 2019*. Accessed: Aug. 17, 2020. [Online]. Available: [http://www.inpe.br/noticias/noticia.php?Cod%5C\\_Noticia=5465](http://www.inpe.br/noticias/noticia.php?Cod%5C_Noticia=5465)
- [7] D. P. Barbosa, A. Noma, T. S. Korting, and L. M. G. Fonseca, "Um estudo experimental com classificadores baseados em Regiões e Perfis EVI," in *Proc. 17th Simpósio Brasileiro de Sensoriamento Remoto*, vol. 1, 2015, pp. 6381–6388.
- [8] E. L. Bullock, C. E. Woodcock, and P. Olofsson, "Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis," *Remote Sens. Environ.*, vol. 238, Nov. 2020, Art. no. 110968, doi: [10.1016/j.rse.2018.11.011](https://doi.org/10.1016/j.rse.2018.11.011).
- [9] P. de Bem, O. de Carvalho Junior, R. F. Guimarães, and R. T. Gomes, "Change detection of deforestation in the Brazilian amazon using landsat data and convolutional neural networks," *Remote Sens.*, vol. 12, no. 6, p. 901, Mar. 2020, doi: [10.3390/rs12060901](https://doi.org/10.3390/rs12060901).
- [10] M. O. Adarme, R. Q. Feitosa, P. N. Happ, C. A. De Almeida, and A. R. Gomes, "Evaluation of deep learning techniques for deforestation detection in the Brazilian amazon and cerrado biomes from remote sensing imagery," *Remote Sens.*, vol. 12, no. 6, p. 910, Mar. 2020, doi: [10.3390/rs12060910](https://doi.org/10.3390/rs12060910).
- [11] Y. E. Shimabukuro, E. Arai, V. Duarte, A. Jorge, E. G. D. Santos, K. A. C. Gasparini, and A. C. Dutra, "Monitoring deforestation and forest degradation using multi-temporal fraction images derived from landsat sensor data in the Brazilian amazon," *Int. J. Remote Sens.*, vol. 40, no. 14, pp. 5475–5496, Jul. 2019, doi: [10.1080/01431161.2019.1579943](https://doi.org/10.1080/01431161.2019.1579943).
- [12] T. A. Lima, R. Beuchle, A. Langner, R. C. Grecchi, V. C. Griess, and F. Achard, "Comparing Sentinel-2 MSI and landsat 8 OLI imagery for monitoring selective logging in the Brazilian Amazon," *Remote Sens.*, vol. 11, no. 8, p. 961, Apr. 2019, doi: [10.3390/rs11080961](https://doi.org/10.3390/rs11080961).
- [13] C. Alves, S. Júnior, O. O. Bittencourt, F. Morelli, and R. Santos, "Classificação De Áreas Queimadas Por Machine Learning Usando Dados De Sensoriamento Remoto," in *Proc. Ann. 16th Brazilian Remote Sens. Symp. SBSR*, 2019, pp. 1784–1787.
- [14] V. Leseux et al., "Cobertura do solo utilizando Sentinel-2 (sensor Msi)," in *Proc. Ann. XIX Brazilian Remote Sensing Symp. SBSR*, 2019, pp. 720–723.
- [15] F. S. Nascimento, M. Gastauer, P. W. M. Souza-Filho, W. R. Nascimento, D. C. Santos, and M. F. Costa, "Land cover changes in open-cast mining complexes based on high-resolution remote sensing data," *Remote Sens.*, vol. 12, no. 4, pp. 1–21, 2020, doi: [10.3390/rs12040611](https://doi.org/10.3390/rs12040611).
- [16] H. L. G. Cassol, E. Arai, E. E. Sano, A. C. Dutra, T. B. Hoffmann, and Y. E. Shimabukuro, "Maximum fraction images derived from year-based project for on-board autonomy-vegetation (PROBA-V) data for the rapid assessment of land use and land cover areas in mato grosso state, Brazil," *Land*, vol. 9, no. 5, p. 139, May 2020, doi: [10.3390/LAND9050139](https://doi.org/10.3390/LAND9050139).
- [17] J. Lee, J. A. Cardille, and M. T. Coe, "Agricultural expansion in mato grosso from 1986-2000: A Bayesian time series approach to tracking past land cover change," *Remote Sens.*, vol. 12, no. 4, pp. 1–18, 2020, doi: [10.3390/rs12040688](https://doi.org/10.3390/rs12040688).
- [18] R. V. Maretto, L. M. G. Fonseca, N. Jacobs, T. S. Korting, H. N. Bendini, and L. L. Parente, "Spatio-temporal deep learning approach to map deforestation in Amazon rainforest," *IEEE Geosci. Remote Sens. Lett.*, early access, Apr. 28, 2020, doi: [10.1109/Lgrs.2020.2986407](https://doi.org/10.1109/Lgrs.2020.2986407).
- [19] C. C. Aggarwal, *Neural Networks and Deep Learning*. New York, NY, USA: Springer, 2018.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, Dec. 2016, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell. AAAI*, 2017, pp. 4278–4284.
- [23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520, doi: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [24] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807, doi: [10.1109/CVPR.2017.195](https://doi.org/10.1109/CVPR.2017.195).
- [25] USGS. (2019). *Earth Explorer*. Accessed: May 19, 2019. [Online]. Available: <http://earthexplorer.usgs.gov/>
- [26] USGS. (2017). *What are the Best Spectral Bands to Use for my Study*. Accessed: May 19, 2019. [Online]. Available: <https://landsat.usgs.gov/what-are-best-spectral-bands-use-my-study>

- [27] Z. Yu, L. Di, R. Yang, J. Tang, L. Lin, C. Zhang, M. S. Rahman, H. Zhao, J. Gaigalas, E. G. Yu, and Z. Sun, "Selection of landsat 8 OLI band combinations for land use and land cover classification," in *Proc. 8th Int. Conf. Agro-Geoinform. (Agro-Geoinform.)*, Jul. 2019, pp. 1–5, doi: 10.1109/Agro-Geoinformatics.2019.8820595.
- [28] M. Miyagawa, M. G. F. Costa, M. A. Gutierrez, J. P. G. F. Costa, and C. F. F. C. Filho, "Lumen segmentation in optical coherence tomography images using convolutional neural network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 600–603, doi: 10.1109/EMBC.2018.8512299.
- [29] MathWorks. (2017). *Evaluate Semantic Segmentation Data Set Against Ground Truth*. Accessed: Aug. 10, 2019. [Online]. Available: <https://in.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html>
- [30] M. Baatz and A. Schape, "Multiresolution Segmentation: An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informations-Verarbeitung XII*, 2000, pp. 12–23.



**JOEL PARENTE DE OLIVEIRA** received the degree in computer engineering from the State University of Amazonas (UEA) in 2010 and the master's degree in electrical engineering from the Federal University of Amazonas (UFAM), Brazil, in 2017. He is currently a Science and Technology Analyst, working in database administration, with the Center for Management and Operation of the Amazon Protection System, Operations and Management Center of the Amazon Protection System (CENSIPAM). His research interests include pattern recognition and digital image research.



**MARLY GUIMARÃES FERNANDES COSTA** (Member, IEEE) received the degree in electrical engineering from the Federal University of Amazonas (UFAM) in 1983, and the master's degree in electrical engineering, in the area of telecommunications, and the Ph.D. degree in electrical engineering, automation concentration area, from the State University of Campinas (UNICAMP), in 1988, and 1996, respectively. She is currently a Professor with the Department of Electrical Engineering, UFAM. She has experience in digital image processing, pattern recognition, and biomedical engineering, focusing on the following subjects such as image automation, the development of diagnostic aid tools, and expert systems.



**CÍCERO FERREIRA FERNANDES COSTA FILHO** (Member, IEEE) received the degree in electrical engineering from the Federal University of Pernambuco (UFPE) in 1982, and the master's degree in electrical engineering, in the area of microelectronics, and the Ph.D. degree in automation from the State University of Campinas (UNICAMP), in 1985 and 1996, respectively. He worked with the UNICAMP and the Center for Analysis, Research, and Technological Innovation Foundation-FUCAPI. Since 1990, he has been a Professor with the Federal University of Amazonas. He is currently a full professor, teaching undergraduate (electrical engineering and computer engineering) and graduate courses (master's degree in electrical engineering-PPGEE and the Ph.D. degree in computer engineering-PPGI). His research interests include artificial intelligence, pattern recognition, digital image processing, and optimization.

• • •