



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2016/01.18.22.14-TDI

## **USO DE REDES NEURAIS NEBULOSAS E FLORESTAS ALEATÓRIAS NA CLASSIFICAÇÃO DE IMAGENS EM UM PROJETO DE CIÊNCIA CIDADÃ**

Marilyn Menecucci Ibañez

Dissertação de Mestrado do  
Curso de Pós-Graduação em  
Computação Aplicada, orientada  
pelos Drs. Fernando Manuel  
Ramos, e Adenilson Roberto  
Carvalho, aprovada em 19 de  
fevereiro de 2016.

URL do documento original:

[<http://urlib.net/8JMKD3MGP3W34P/3L2D5ME>](http://urlib.net/8JMKD3MGP3W34P/3L2D5ME)

INPE  
São José dos Campos  
2016

**PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@inpe.br

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO  
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)

**Membros:**

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas (CEA)

Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)

Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial (ETE)

Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SID)

**BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

**REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

**EDITORAÇÃO ELETRÔNICA:**

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)





MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2016/01.18.22.14-TDI

## **USO DE REDES NEURAIS NEBULOSAS E FLORESTAS ALEATÓRIAS NA CLASSIFICAÇÃO DE IMAGENS EM UM PROJETO DE CIÊNCIA CIDADÃ**

Marilyn Menecucci Ibañez

Dissertação de Mestrado do  
Curso de Pós-Graduação em  
Computação Aplicada, orientada  
pelos Drs. Fernando Manuel  
Ramos, e Adenilson Roberto  
Carvalho, aprovada em 19 de  
fevereiro de 2016.

URL do documento original:

[<http://urlib.net/8JMKD3MGP3W34P/3L2D5ME>](http://urlib.net/8JMKD3MGP3W34P/3L2D5ME)

INPE  
São José dos Campos  
2016

Dados Internacionais de Catalogação na Publicação (CIP)

---

Ib1u Ibañez, Marilyn Menecucci.  
Uso de redes neurais nebulosas e florestas aleatórias na  
classificação de imagens em um projeto de ciência cidadã / Marilyn  
Menecucci Ibañez. – São José dos Campos : INPE, 2016.  
xxiv + 96 p. ; (sid.inpe.br/mtc-m21b/2016/01.18.22.14-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto  
Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

Orientadores : Drs. Fernando Manuel Ramos, e Adenilson  
Roberto Carvalho.

1. Redes neurais. 2. Florestas aleatórias. 3. Processamento de  
imagens. 4. Computação cidadã. 5. Desmatamento. I.Título.

CDU 004.036.26:630

---



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](#).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#).

## **ERRATA AGRADECIMENTOS**

Agradeço à Biblioteca do INPE pelo tempo dispensado na ajuda da padronização desta dissertação.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo número 132849/2009-7, no período de março de 2009 a dezembro de 2009.

Aluno (a): **Marilyn Menecucci Ibañez**

Título: " USO DE REDES NEURAIS NEBULOSAS E FLORESTAS ALEATÓRIAS NA CLASSIFICAÇÃO DE IMAGENS EM UM PROJETO DE CIÊNCIA CIDADÃ".

Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Mestre** em  
**Computação Aplicada**

Dr. Reinaldo Roberto Rosa

  
\_\_\_\_\_  
Presidente / INPE / SJCampos - SP

Dr. Fernando Manuel Ramos

  
\_\_\_\_\_  
Orientador(a) / INPE / SJCampos - SP

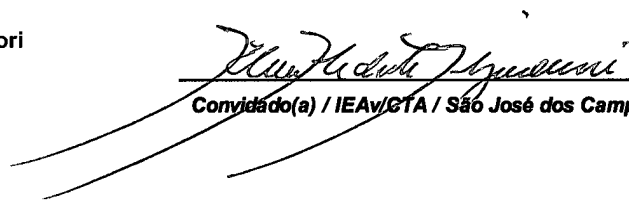
Dr. Adenilson Roberto Carvalho

  
\_\_\_\_\_  
Orientador(a) / INPE / São José dos Campos - SP

Dr. José Carlos Becceneri

  
\_\_\_\_\_  
Membro da Banca / INPE / SJCampos - SP

Dr. Elcio Hideiti Shiguemori

  
\_\_\_\_\_  
Convidado(a) / IEAv/CTA / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) maioria simples

☒ unanimidade

São José dos Campos, 19 de Fevereiro de 2016



”Foi o tempo que dedicastes a tua rosa que fez a tua rosa tão importante.”

ANTOINE DE SAINT-EXUPÉRY  
em *”O Pequeno Príncipe”*



A Deus, Nossa Senhora  
e a meus pais Aurora e Elar,  
meu irmão Delano, sobrinha Nina e cunhada Stefani.





## AGRADECIMENTOS

Agradeço a Deus e a Nossa Senhora por terem me abençoado com saúde e perseverança para sempre continuar, mesmos nos momentos mais difíceis.

Agradeço ao meu orientador, o Professor Dr. Fernando M. Ramos, pelo conhecimento passado, pelos conselhos dados, amizade gerada e por acreditar que eu poderia desenvolver e finalizar este trabalho.

Agradeço aos co-orientadores, Dr. Andenilson Carvalho e Dr. Eduardo Fávero Pacheco da Luz, do CEMADEM, pela orientação, correções, tempo dispensado e amizade gerada que foram de extrema importância para o andamento e finalização do trabalho.

Agradeço aos professores Dr. Nandamudi L. Vijaykumar, Dr. Ezzat Selim Chalhoub e Dr. Stephan Stephany pelo apoio e incentivo para que eu sempre me mantivesse no objetivo de finalizar o mestrado.

Agradeço a minha família, meus pais e irmão, pela paciência e apoio nos momentos de cansaço que me fizeram sempre permanecer concentrada para o desenvolvimento do projeto.

Agradeço ao Professor Dr. Eduardo Leal, do Instituto Federal de Educação Ciência e Tecnologia de São Paulo - IFSP, pelo apoio e conselhos para que eu pudesse estar sempre presente nas atividades do mestrado.

Agradeço aos membros da banca pelo tempo dedicado à análise do trabalho e pelo conhecimento passado.

Agradeço ao INPE/LAC/CAP pela disponibilização da infraestrutura para o desenvolvimento do conhecimento e trabalho.

Agradeço ao Instituto Federal de Educação Ciência e Tecnologia de São Paulo - IFSP pela licença concedida para o desenvolvimento dos estudos.



## RESUMO

Recentemente, um projeto de ciências cidadã chamado *ForestWatchers* (LUZ et al., 2014) foi lançado com o objetivo de envolver os cidadãos leigos no monitoramento do desmatamento. Por meio de uma interface Web, voluntários de todo o mundo são convidados a analisar imagens MODIS de regiões florestais e confirmar se atribuições automáticas de regiões de florestas desmatadas estão corretamente classificadas. Considerando a grande área em todo mundo coberta pelas florestas tropicais, torna-se fundamental o uso de um classificador rápido que atenda a um objetivo duplo: o mapeamento de pixels em duas classes ('Floresta' e 'não-Floresta') e a seleção dos pixels a serem enviados aos voluntários para a inspeção, com base em uma métrica de confiança. Nesta dissertação investiga-se o uso de dois métodos distintos — rede neural de perceptrons multicamada (*Multi-Layered Perceptron*, MLP) difusa e Floresta Aleatória (*Random Forest*, RF) — na classificação de padrões de desmatamento na Amazônia brasileira, utilizando imagens MODIS. Neste sentido, foram gerados mapas de desmatamento de diversos tamanhos, de diversas áreas do estado de Rondônia. Os resultados foram validados com os resultados de projeto PRODES, que avalia anualmente o desmatamento na Amazônia brasileira. Nestes testes, o classificador RF apresentou um desempenho amplamente superior ao das redes neurais *Multi-Layered Perceptro* e *Multi-Layered Perceptron Fuzzy*.

Palavras-chave: Redes Neurais, Florestas Aleatórias, Processamento de Imagens, Computação Cidadã, Desmatamento, Satélites.



# USE OF FUZZY NEURAL NETWORKS AND RANDOM FOREST IN IMAGE'S CLASSIFICATION OF A CITIZEN SCIENCE PROJECT

## ABSTRACT

Recently, a citizen science project called ForestWatchers (LUZ et al., 2014) was launched in order to involve the laity citizens in the monitoring of deforestation. Through a Web interface, volunteers from around the world are invited to review MODIS images of forest regions and confirm that automatic assignment of cleared forest areas are properly classified. Considering the large area worldwide covered by tropical forest, it is essential to use a fast classifier that meets a double objective: the pixel mapping into two classes ('Forest' and 'non-forest') and the selection of pixels to be sent to volunteers for inspection, based on a reliable metric. This dissertation investigates the use of two different methods - neural network multilayer perceptrons (Multi-Layered Perceptron, MLP) diffuse and Random Forest (Random Forest, RF) - the deforestation pattern classification in the Brazilian Amazon using MODIS images. In this sense, deforestation maps were generated from various sizes, from different areas of the state of Rondonia. The results were validated with the results of PRODES project, which annually evaluates deforestation in the Brazilian Amazon. In these tests, the classifier RF showed a vastly superior performance to the *Multi-Layered Perceptro* and *Multi-Layered Perceptron Fuzzy* neural networks.

Keywords: Neural Network, Image Processing, Computing Citizen, Desforestation and Satellites.



## LISTA DE FIGURAS

	<u>Pág.</u>
1.1 Taxa de desmatamento da Amazônia Legal de 1988 a 2015. . . . .	6
1.2 Exemplo de imagem disponibilizada aos voluntários para a correção da classificação da Rede Neural selecionando o botão Forest (verde) e Non- Forest (vermelho) . . . . .	7
2.1 Arquitetura do projeto ForestWatchers. . . . .	11
2.2 Tela Inicial do projeto ForestWatchers. . . . .	12
2.3 Projeção sinusoidal das imagens da NASA. . . . .	13
2.4 Transformações das imagens MODIS da NASA. . . . .	14
2.5 Taxas anuais apresentadas pelo programa PRODES . . . . .	15
2.6 Exemplo de imagem do projeto PRODES para o estado de Rondônia. . .	16
2.7 Exemplo de imagem do programa DETER. . . . .	17
2.8 (a) Neurônio Biológico e (b) Neurônio Artificial. . . . .	19
2.9 Modelo Computacional da Rede MLP. . . . .	20
2.10 Modelo de um sistema Neuro-Fuzzy. . . . .	21
2.11 Exemplo de um perceptron multicamadas difuso. . . . .	22
2.12 Árvore de decisão e seus componentes. . . . .	23
2.13 Exemplo da aplicação da Árvore de Decisão. . . . .	24
2.14 Exemplo de Random Forest. . . . .	25
2.15 Algoritmo para uma <i>Random Forest</i> . . . . .	26
2.16 Exemplo da tela inicial do <i>Weka</i> . . . . .	27
2.17 Exemplo da tela Explorer do <i>Weka</i> . . . . .	28
2.18 Exemplo da tela Experimenter do <i>Weka</i> . . . . .	29
2.19 Exemplo da tela KnowledgeFlow do <i>Weka</i> . . . . .	29
2.20 Exemplo da tela SimpleCLI do <i>Weka</i> . . . . .	30
2.21 Exemplo de Matriz de Confusão. . . . .	31
3.1 Metodologia utilizada no trabalho. . . . .	33
3.2 Janela da Ferramenta Modis Reprojection Tool - MRT. . . . .	34
3.3 Comparação de uma imagem binarizada com a imagem original PRODES. .	35
3.4 Exemplo da divisão das imagens Modis e PRODES. . . . .	35
3.5 Exemplo do arquivo de configuração da Rede. . . . .	39
3.6 Exemplo do arquivo de entrada da Random Forest. . . . .	40
3.7 Exemplo de uma Imagem Classificada pela Rede Neural MLP em Flo- resta (verde) e Não-Floresta (Vermelho). . . . .	41



4.1	Área de Floresta dos Municípios PRODES e das Imagens Classificadas e erro do cálculo das áreas. . . . .	67
4.2	Gráfico da Estimativa de tempo considerando Tempo x Número de pixels.	87
4.3	Gráfico da Estimativa de tempo considerando Tempo x Número de Árvores.	87

## LISTA DE TABELAS

	<u>Pág.</u>
3.1 Dados dos píxels em valores RGB . . . . .	37
3.2 Dados dos testes com a configuração da Rede MLP . . . . .	38
3.3 Dados dos testes com a configuração da Rede MLP . . . . .	39
3.4 Matriz de Confusão . . . . .	42
4.1 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala . . . . .	44
4.2 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala . . . . .	45
4.3 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala . . . . .	46
4.4 Resultados do Cálculo da Entropia de Kullback-Leiber para os testes de Pequena Escala . . . . .	47
4.5 Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala . . . . .	49
4.6 Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala . . . . .	50
4.7 Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala . . . . .	51
4.8 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala . . . . .	53
4.9 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala . . . . .	54
4.10 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala . . . . .	55
4.11 Resultados do Cálculo da Entropia de Kullback-Leiber para os testes de Média Escala . . . . .	56
4.12 Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala . . . . .	58
4.13 Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala . . . . .	59
4.14 Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala . . . . .	60
4.15 Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala . . . . .	61

4.16	Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala . . . . .	62
4.17	Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala . . . . .	63
4.18	Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala . . . . .	64
4.19	Resultados da Validação da Área classificada como Floresta na imagem resultante do ano de 2010 do estado de Rondônia . . . . .	66
4.20	Resultados da Validação da Área classificada como Floresta na imagem resultante do ano de 2011 do estado de Rondônia . . . . .	66
4.21	Dados de porcentagem de árvores que convergiram para uma dada classe fornecida pelo pacote <i>Weka</i> . . . . .	68
4.22	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 10 árvores . . . . .	70
4.23	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 100 árvores . . . . .	71
4.24	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 500 árvores . . . . .	72
4.25	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com com 100 árvores . . . . .	73
4.26	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 10 árvores . . . . .	74
4.27	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 100 árvores . . . . .	75
4.28	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 500 árvores . . . . .	76
4.29	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 1000 árvores . . . . .	77
4.30	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2010 com 10 árvores . . . . .	78
4.31	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2010 com 100 árvores . . . . .	79
4.32	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2010 com 500 árvores . . . . .	80

4.33	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2010 com 1000 árvores . . . . .	81
4.34	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 10 árvores . . . . .	82
4.35	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 100 árvores . . . . .	83
4.36	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 500 árvores . . . . .	84
4.37	Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 1000 árvores . . . . .	85
4.38	Tempo de processamento em segundos para a técnica de Random Forest para o número de 10, 100, 500 e 1000 árvores. . . . .	86



## LISTA DE ABREVIATURAS E SIGLAS

MOPSO	–	Multi-Objective Particle Swarm Optimization
MLC	–	Maximum Likelihood Classification
SVM	–	Support Vector Machine
GAB	–	Gentil Ada Boost
WFI	–	Wide Field Imager
PRODES	–	Programa de Monitoramento de Desmatamento da Amazônia Legal
DETER	–	Sistema de Detecção de Desmatamento em Tempo Real
GDAL	–	Geospatial Data Abstraction Library
MLP	–	Multi-Layered Perceptron
FP	–	False Positive
FN	–	False Negative
TP	–	True Positive
TN	–	True Negative
RF	–	Random Forest



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO . . . . .</b>	<b>1</b>
1.1 Revisão Bibliográfica . . . . .	2
1.2 Motivação e Objetivo . . . . .	5
<b>2 FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>9</b>
2.1 Ciência Cidadã . . . . .	9
2.2 ForestWatchers . . . . .	10
2.3 Dados Utilizados . . . . .	12
2.3.1 Imagens do sensor MODIS que está a bordo dos satélites Aqua e Terra da NASA . . . . .	12
2.3.2 PRODES . . . . .	15
2.3.3 DETER . . . . .	17
2.4 Redes Neurais . . . . .	18
2.4.1 Arquitetura de Redes . . . . .	19
2.4.2 MLP - MultiLayer Perceptron . . . . .	20
2.5 Sistemas Nebulosos . . . . .	21
2.5.1 Rede Neural Nebulosa . . . . .	21
2.6 Árvore de Decisão . . . . .	23
2.7 Árvores Aleatórias ou Random Forest . . . . .	24
2.8 Weka - Waikato Environment for Knowledge Analysis . . . . .	27
2.9 GDAL . . . . .	30
2.10 Matriz de Confusão . . . . .	30
2.11 Entropia da Imagem . . . . .	31
2.11.1 Divergência ou Entropia de Kullback-Leibler . . . . .	32
<b>3 METODOLOGIA . . . . .</b>	<b>33</b>
3.1 Aquisição e Reprojeção das Imagens . . . . .	34
3.1.1 Imagem do sensor Modis dos Satélites Aqua e Terra da NASA . . . . .	34
3.1.2 Imagem PRODES . . . . .	34
3.2 Subdivisão das Imagens . . . . .	35
3.3 Transformação dos Pixels em valores RGB . . . . .	36
3.4 Processamento dos dados RGB na Rede MLP, Rede MLP Nebulosa e Random Forest . . . . .	37



3.4.1	Processamento na Rede MLP . . . . .	37
3.4.2	Processamento da Rede MLP Nebulosa . . . . .	40
3.4.3	Processamento da Random Forest . . . . .	40
3.5	Processamento para a criação das Imagens Classificadas . . . . .	41
3.6	Validação das Imagens Classificadas . . . . .	41
3.6.1	Entropia ou Divergência de Kullback Leiber . . . . .	41
3.6.2	Matriz de Confusão . . . . .	42
<b>4</b>	<b>RESULTADOS OBTIDOS . . . . .</b>	<b>43</b>
4.1	Testes de Pequena Escala . . . . .	43
4.2	Validação dos Resultados da Classificação com os Testes da Pequena Escala . . . . .	47
4.2.1	Entropia ou divergência de Kullback-Leiber . . . . .	47
4.2.2	Matriz de Confusão . . . . .	48
4.3	Testes de Média Escala . . . . .	52
4.4	Validação dos Resultados da Classificação com os Testes da Média Escala . . . . .	56
4.4.1	Entropia ou divergência de Kullback-Leiber . . . . .	56
4.4.2	Matriz de Confusão . . . . .	57
4.5	Testes de Grande Escala . . . . .	61
4.6	Validação da Área classificada como Floresta na imagem resultante do estado de Rondônia . . . . .	65
4.6.0.1	Validação do Limiar encontrado para o cálculo da área de Floresta . . . . .	67
4.7	Cálculo do Grau de Incerteza das imagens classificadas para a diminuição dos trabalhos dos voluntários . . . . .	67
4.8	Estimativa do Tempo de Processamento da Random Forest . . . . .	86
<b>5</b>	<b>CONCLUSÕES . . . . .</b>	<b>89</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>91</b>

## 1 INTRODUÇÃO

O Brasil tem o maior e mais bem sucedido programa de monitoramento de florestas tropicais do mundo. Desde 1988, o programa PRODES (Monitoramento da Floresta da Amazônia Brasileira por Satélite) vem realizando levantamentos de desmatamento anuais detalhados da Amazônia brasileira. O PRODES utiliza um processo semi-automatizado para executar o processamento digital de imagens TM/Landsat. Mais Recentemente, o programa DETER (Sistema de Detecção de Desmatamento em Tempo Real) fornece alertas semanais de desmatamento de áreas com 25 hectares ou mais, usando imagens do sensor MODIS (Moderate Resolution Imaging Spectroradiometer) dos satélites TERRA e ACQUA da NASA. Esta estrutura de monitoramento, juntamente com uma política de dados livres e abertos (todos os softwares e os resultados estão disponíveis na Web), possibilitou maior transparência e eficácia nos esforços de conservação do Brasil, e ajudou o país a reduzir muito a sua perda anual de floresta na Amazônia. Infelizmente, ambos os projetos envolvem passos de inspeção dispendiosos e demorados do mapa de classificação automática por especialistas, o que torna difícil a sua replicação em países menos desenvolvidos.

Recentemente, um projeto de ciências cidadã chamado *ForestWatchers* ([LUZ et al., 2014](#)) foi lançado com o objetivo de envolver os cidadãos em todo o mundo no monitoramento do desmatamento. Por meio de uma interface Web, voluntários de todo o mundo são convidados a analisar imagens MODIS de regiões florestais e confirmar se atribuições automáticas de regiões de florestas desmatadas estão corretamente classificadas. Considerando a grande área em todo mundo coberta pelas florestas tropicais, torna-se fundamental o desenvolvimento de um classificador rápido que atenda a um objetivo duplo: o mapeamento de pixels em duas classes ('Floresta' e 'não-Floresta') e a seleção dos pixels a serem enviados aos voluntários para a inspeção, com base em uma métrica de confiança.

Dentro desse contexto, o objetivo desta dissertação é investigar o uso de dois métodos distintos — rede neural de perceptrons multicamada (*Multi-Layered Perceptron*) difusa e Floresta Aleatória (*Random Forest*, RF) — na classificação de padrões de desmatamento na Amazônia brasileira, utilizando imagens MODIS. Essa investigação ajudará na definição do classificador rápido para o mapeamento e seleção dos pixels a serem enviados para os voluntários do projeto *ForestWatchers*. Essencialmente, o MLP é uma rede feedforward multicamadas que utiliza um mecanismo de aprendizado supervisionado com base na configuração dos parâmetros de acordo com o erro entre as saídas calculadas e desejadas da rede. Para o processo de aprendizagem MLP, são utilizados mapas de classificação do PRODES como

saídas desejadas. Os mapas do PRODES, que originalmente mostram cinco classes de cobertura da terra, com uma resolução espacial de 60 *m*, serão redimensionadas para as duas classes de (F e N) do nosso classificador MLP. A resolução espacial será então reduzida para 250 *m*, a mesma apresentada por imagens MODIS, utilizado como dados de entrada no processo de aprendizagem. Em MLPs nebulosas, as saídas são mapeadas para a faixa de  $[0, 1]$  e referem-se com o grau de adesão ao padrão de entrada para cada uma das classes de saída. Isso permitirá usar graus normalizados de adesão como uma métrica de confiança na classificação de cada pixel, reduzindo consideravelmente a tarefa dos voluntários uma vez que apenas pixels com baixa confiança são visualmente inspecionados por eles.

Como seu nome sugere, uma RF utiliza um conjunto de árvores de decisão (AD) para executar a tarefa de classificação. Diferentemente das ADs, ao invés de crescer uma única e complexa árvore sob a supervisão de um analista, a RF agrega a saída de muitas árvores simples, que são ajustadas sem muita (ou nenhuma) supervisão externa. Nesta técnica de classificação é também possível extrair uma métrica de confiança na classificação de cada pixel, a partir do número de árvores da RF que escolheram a classe vencedora.

Com o intuito de testar e comparar as duas abordagens, produzimos mapas de desmatamento de diversos tamanhos, de diversas áreas do estado de Rondônia. Rondônia tem uma área de 237.576 *km*<sup>2</sup> (mais ou menos, o tamanho de Portugal), originalmente cobertos por Floresta ou Cerrado Amazônicos. Os resultados serão validados com o mapa de desmatamento correspondente do projeto PRODES.

## 1.1 Revisão Bibliográfica

Nas últimas duas décadas, as técnicas de Inteligência Artificial vem sendo aplicadas no processamento de imagens de satélites, gerando resultados eficientes na identificação e classificação dessas imagens. Assim, são apresentados a seguir alguns trabalhos que obtiveram resultados relevantes para o desenvolvimento desse projeto.

No ano de 1997, [Andrade e Martins \(1997\)](#) fizeram um trabalho que aborda o processamento de imagens de satélites utilizando redes neurais. O objetivo do estudo era reduzir o trabalho de análise manual de imagens dos profissionais da área de sensoriamento remoto. O estudo utilizou três modelos de redes neurais: dois modelos supervisionados (*Backpropagation* e *Learning Vector Quantization- LVQ*) e um modelo não supervisionado (*Kohonen's self-organization map - SOM*).

Em 2002, foi publicado, [Egmont-Petersen et al. \(2002\)](#), um estudo que apresentou

uma revisão sobre a utilização do processamento de imagens com redes neurais. Esse artigo fez uma análise de mais de 200 trabalhos relacionados à aplicação de redes neurais em processamento de imagens e também realizou uma discussão do papel atual e futuro das seguintes redes neurais aplicadas a imagens: *feed-forward*, *Hopfield* e *Kohonen feature maps*.

No ano de 2004, (MAS et al., 2004) publicou um estudo que analisa as imagens do LandSat para prever a distribuição do desmatamento das florestas tropicais. As imagens são classificadas para gerarem mapas de desmatamento digitais que identificam o desmatamento e as áreas com persistência de floresta. Os mapas digitais são utilizados para prever o risco de desmatamento em relação ao ambiente externo. A rede neural MultiLayer Perceptron foi utilizada para a análise das imagens.

Um artigo denominado *Detecting temporal changes in satellite imagery using ANN*, publicado por Mathur e Govil (2005), apresentou uma abordagem que usa redes neurais para analisar mudança temporal nos padrões de imagens utilizadas em sensoriamento remoto.

Em 2006, foi publicado um artigo que apresentava a implementação dos algoritmos de rede neural *K-means* e *BackPropagation* para a segmentação e classificação de imagens de satélites (SAPKAL et al., 2006).

No ano de 2007, publicou-se um estudo no qual apurava-se a capacidade de aprendizagem de operações de processamento de imagens por RNA e se esse conhecimento prévio poderia ser usado em problemas gerais de processamento de imagens e o que poderia ser aprendido sobre o problema em questão a partir de redes treinadas (RIDDER et al., 2003).

Em 2008, o artigo publicado por Ahmadi et al. (2008) apresentou uma abordagem de vetorização de imagens utilizando CAD, técnicas de processamento de imagens e redes neurais.

O artigo de Neagoe e Ropot (2005), em 2009, apresentou uma nova abordagem de redes neurais para o processamento de imagens de satélites. A abordagem é denominada *Concurrent Self-Organizing Maps* - (CSOM). Essa nova abordagem conduziu a resultados melhores do que a abordagem *Self-Organizing Maps* - (SOM) tradicional.

Também em 2009, foi publicado um estudo sobre diversas técnicas (Cadeia de Markov, Automatos Finitos, Redes Neurais) para a análise de imagens que possibilitam

prever o risco de desmatamento na região da Guatemala (BRUNO et al., 2006).

O artigo de Jiang et al. (2010) apresentou a utilização da Rede Neural com a aplicação do *BackPropagation* para melhorar a precisão da classificação de imagens de sensoriamento remoto.

Em 2011, foi publicado um estudo que apresentou uma abordagem de rede neural para a detecção de alterações nas imagens de satélites multiespectrais. As redes neurais utilizadas foram: *Multilayer Perceptron*, *Radial Basis Function Neural Network* (RBF) e *Supervised Self Organization Map* (SOM) (NEAGOE et al., 2011).

No ano de 2013, Pereira et al. (2013) elaboraram um trabalho que tinha como objetivo a criação de um algoritmo usando o ambiente MATLAB com um toolbox de Redes Neurais para a classificação de áreas cafeeiras em imagens de satélite, para identificar e mapear o uso e ocupação da terra, com ênfase na identificação de áreas cafeeiras na região de Três Pontas localizada no sul de Minas Gerais. No trabalho projetou-se como uma RNA convolutiva com um Perceptron de múltiplas camadas para reconhecer formas bidimensionais. Também utilizou-se o algoritmo backpropagation para o treinamento da rede neural.

Em 2010, Liu e Yetik (2010) apresentaram um novo modelo estatístico para algoritmo de Classificação de Verossimilhança Máxima (MLC) para melhorar o desempenho da segmentação / classificação de imagem. O MLC é utilizado amplamente em muitas aplicações de classificação. No trabalho abordou-se a resolução do problema de distribuições semelhantes para assuntos diferentes por meio da modelagem da média dos valores das características de cada classe e dos valores que se apresentam como dois grupos de variáveis aleatórias dependentes. Desta forma, a associação de classe depende tanto dos valores de recursos quanto de variáveis aleatória que capturem informações específicas do assunto. O método foi aplicado em um conjunto de imagens de simulação e os resultados experimentais obtidos mostram que o modelo proposto poderia melhorar os resultados de segmentação de MLC supervisionado clássicos quando existem diferenças consideráveis entre os indivíduos.

Em 2014, o trabalho proposto por Agrawal e Bawane (2014) apresenta uma nova metodologia baseada na otimização multiobjetivo por enxame de partículas (MOPSO) para determinar as bandas espectrais (Azul, Verde, Vermelho e Infravermelho) e também o número de nós que cada camada oculta deve ter na rede neural. O resultado obtido com tal rede neural otimizada é comparado com a de classificadores tradicionais como os classificadores MLC e euclidiano.

No ano de 2010, [Barbieri et al. \(2011\)](#) propôs uma segmentação baseada em entropia de imagens. A metodologia é avaliada com relação às imagens de satélite obtidas do Google Earth, com o objetivo de identificar as regiões aquáticas, urbanas e rurais.

Em 2003, o trabalho de [Nepomuceno \(2004\)](#) avaliou as propriedades discriminatórias de dados de radar na banda P para o mapeamento da cobertura de uma área próxima à Floresta Nacional do Tapajós, no Estado do Pará, utilizando a rede neural artificial não-supervisionada Fuzzy-ART (Teoria da Ressonância Adaptativa). Na conclusão do trabalho, relatou-se o bom desempenho da Rede Fuzzy-ART no rápido processamento das imagens de tamanho 950 colunas e 2953 linhas (2,9 MB), com resolução de 2,5m x 2,5m e uma área aproximada de 18  $km^2$ .

No artigo de [Akar, Özlem and Güngör, Oğuz \(2012\)](#) é apresentado a análise do desempenho do algoritmo de *Random Forest* (RF), conhecido como um método de classificação de conjuntos baseado em aprendizagem. Os resultados da classificação de RF foram comparados com os resultados obtidos a partir dos algoritmos Gentil AdaBoost (GAB), Support Vector Machine (SVM) e a classificação de máxima verossimilhança (MLC) para áreas rurais e urbanas. Essa classificação utilizou imagens dos satélite Ikonos e QuickBird, com quatro bandas multiespectrais com diferentes resoluções espaciais. No trabalho ressalta-se que o RF pode ser considerado tanto um método de conjunto quanto a um algoritmo de aprendizado de máquina. Assim, quando comparado com GAB, é considerado como um método de conjunto e quando comparado com SVM considera-se como um algoritmo de aprendizado de máquina. O *Random Forest* também pode ser utilizado entre algoritmos de classificação com base em pixel, pois considera pixels individuais e não grupos de pixels. Desta forma, no trabalho também o compara com o método de MLC, que é amplamente usado como um algoritmo de classificação tradicional de pixel.

Apesar dos diversos trabalhos citados na revisão bibliográfica, é possível constatar que existem poucos ou inexistentes trabalhos na área de monitoramento de florestas que i) utilizem classificadores tipo MLP Fuzzy ou RF; ou ii) sejam especialmente desenvolvidos para um projeto de ciência cidadã. Desta forma, a presente dissertação é oportuna e poderá contribuir de forma significativa ao projeto *ForestWatchers* e a outras iniciativas de ciência cidadã similares.

## 1.2 Motivação e Objetivo

As Florestas Tropicais são o ecossistema de maior biodiversidade do planeta e por isso são de grande importância para a vida do homem. No entanto, apesar de serem

essencial para vida da terra, o homem insiste em desmatá-las e destruí-las constantemente. Desta forma, algumas políticas públicas foram desenvolvidas no país com intenção de monitorar e conter tal destruição. Particularmente no Brasil, essas políticas monitoram a Amazônia Brasileira desde 1988 e emitem alertas anuais de taxas de desmatamento. A Figura 1.1 apresenta a evolução dessas taxas de 1988 até 2015.

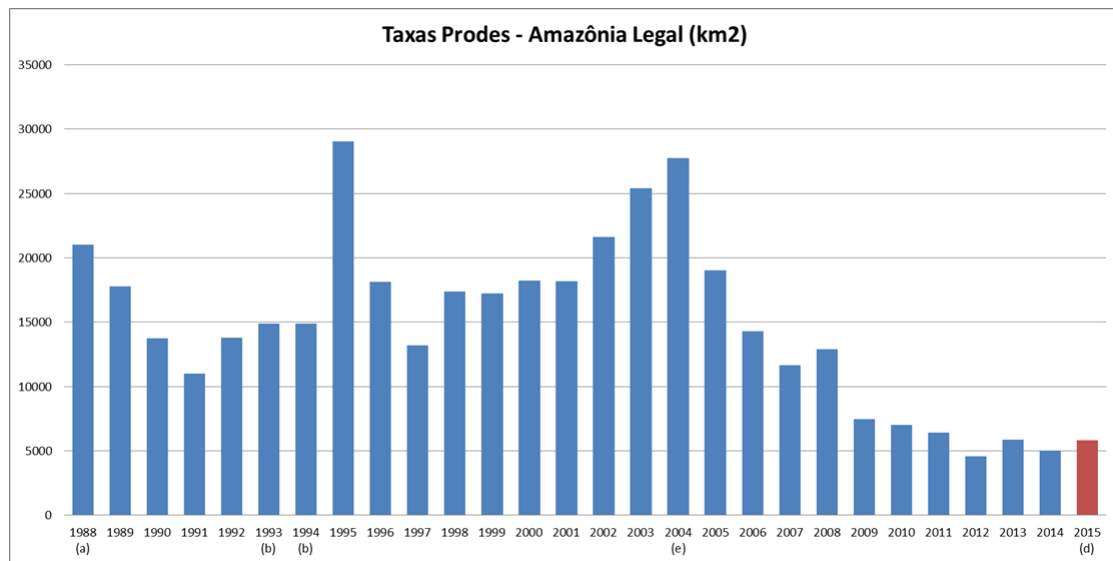


Figura 1.1 - Taxa de desmatamento da Amazônia Legal de 1988 a 2015.

Uma dessas políticas é o PRODES, que desde 1988 emite alertas anuais das taxas de desmatamento da Amazônia Legal utilizando imagens dos satélites da família *LandSat*. Outra política, é o programa de Detecção de Desmatamento em Tempo Real - DETER, que desde de 2004 emite alertas quinzenais de desmatamento da Amazônia Legal. Como observado na Figura 1.1, essas políticas têm obtido bom resultados com a diminuição da taxa de desmatamento nos últimos 5 anos. No entanto, esses programas utilizam especialistas treinados nas suas tarefas mais críticas e isso eleva muito seu custo e aumenta seu tempo de produção. Esses custos dificultam a replicação desses programas em países menos desenvolvidos. Como uma iniciativa para atender esses países, foi desenvolvido um projeto de ciência cidadã denominado ForestWatchers Luz et al. (2014).

O ForestWatchers, em sua configuração atual, utiliza uma Rede Neural MultiLayer Perceptron no lugar dos sistemas utilizados pelos especialistas para a classificação de imagens. Essas imagens classificadas são disponibilizadas por meio de uma interface

Web a voluntários de todo o mundo para a confirmação da classificação da Rede. A Figura 1.2 apresenta um exemplo da imagem disponibilizada aos voluntários, na qual a esquerda é apresentado a imagem do sensor MODIS da NASA e a direita a imagem classificada pela rede neural artificial. Na imagem a direita, uma região da imagem é selecionada, área dentro do retângulo, e o voluntário confirma a classificação Floresta (verde) e Não-Floresta (vermelho) escolhendo o botão Forest (verde) e Non-Forest (vermelho).



Figura 1.2 - Exemplo de imagem disponibilizada aos voluntários para a correção da classificação da Rede Neural selecionando o botão Forest (verde) e Non-Forest (vermelho)

Fonte: Luz et al. (2014)

Com base no contexto apresentado, o objetivo dessa dissertação é investigar classificadores mais especializados para a classificação das imagens disponibilizadas aos voluntários do ForestWatchers. Pretende-se assim reduzir o número de tarefas enviadas aos voluntários. Para isso, as técnicas Rede Neural MultiLayer Perceptron Nebulosa e Florestas Aleatórias serão analisadas. Estas técnicas foram escolhidas pois elas possibilitam a obtenção de uma medida da qualidade da tarefa realizada pelo classificador.

Esta dissertação está organizada da seguinte maneira: o Capítulo 1, Introdução, apresenta a Revisão Bibliográfica com os principais trabalhos desenvolvidos e a Motivação e Objetivo, com as ideias que motivaram o desenvolvimento do trabalho. No Capítulo 2, apresenta-se uma explicação do funcionamento das técnicas e sistemas



utilizados no trabalho. No Capítulo 3 é apresentado a metodologia aplicada no desenvolvimento das técnicas estudadas. O Capítulo 4 mostra os resultados obtidos com a aplicação da metodologia desenvolvida. Já no capítulo 5 é apresenta a conclusão sobre o trabalho desenvolvido e os trabalhos futuros. E por fim é apresentado as referências bibliográficas utilizadas.

## 2 FUNDAMENTAÇÃO TEÓRICA

O aumento do desmatamento das florestas tropicais, principalmente no Brasil, preocupa os estudiosos da área de todo mundo. Assim, a utilização de técnicas para a análise de imagens de satélites das áreas de florestas tropicais é de grande interesse de toda a sociedade. Uma das técnicas analisadas é a rede neural que tem sido utilizada em um grande número de aplicações complexas e tem provado ser eficaz em uma variedade de campos, como reconhecimento de padrões, classificação, visão, sistemas de controle e predição (CARVALHO, 2011). Como visto no capítulo anterior, as redes neurais supervisionadas como a MLP têm sido muito utilizadas na área de processamento de imagens com resultados bastante significativos (EGMONT-PETERSEN et al., 2002). Uma variação da MLP, a *MLP Nebulosa* bem como a técnica de classificação *Random Forest* serão estudadas no projeto para melhorar a classificação das imagens. As imagens utilizadas, neste projeto, na identificação de regiões com desmatamento são do sensor MODIS, dos satélites Acqua e Terra da NASA e do projeto PRODES do INPE. Nas próximas seções apresenta-se com mais detalhes as técnicas e projetos utilizados no trabalhos.

### 2.1 Ciência Cidadã

A ciência cidadã é uma prática que possibilita que a sociedade comum participe de projetos de pesquisa auxiliando de forma voluntária as atividades de pesquisa relacionadas a medição, observação e cálculo das informações analisadas Conrad e Hilchey (2011). O primeiro projeto que se tem conhecimento de utilizar essa prática foi o *Audubon Society* de 1900, no qual 27 voluntários tinham a função de realizar a contagem de passáros, como apresentado em Arcanjo (2014).

De acordo com Arcanjo (2014), os projetos de ciência cidadã podem ser classificados como apresentado a seguir:

- *Volunteer Computing* ou Computação Voluntária: Os voluntários não participam diretamente das tarefas somente disponibilizam os seus recursos computacionais, o que permite realizar as simulações em um ambiente similar ao de um processamento distribuído.
- *Volunteer Thinking* ou Pensamento Voluntário: Os voluntários participam diretamente na realização das tarefas, colaborando na análise das informações por meio dos seus conhecimentos pessoais. Neste projeto utilizamos a ideia de pensamento voluntário

- *Volunteer Sensing* ou Sensoriamento Voluntário: Os voluntários são responsáveis pela coleta dos dados que serão utilizados na pesquisa.

Essa nova forma de desenvolver os projetos científicos envolvendo a sociedade, gera diversos benefícios não só para ciência como para a própria sociedade [Arcanjo \(2014\)](#). Os principais benefícios identificados podem ser observados como segue:

- Benefício para a Comunidade Científica: Melhoria na precisão dos resultados devido a grande extensão das áreas analisadas pelos voluntários gerando uma grande diversidade de dados.
- Benefício para os Voluntários: Possibilidade da aplicação e melhoria dos conhecimentos dos voluntários. Também existem relatos de benefícios ligados a diversão na análise dos dados.
- Benefício para a Educação: A comunidade educacional consegue ter um contato direto com a comunidade científica melhorando e aumentando o seu conteúdo de conhecimento.
- Benefício para a Sociedade: Possibilita-se diminuir a distância entre o cidadão comum e a comunidade científica.

## 2.2 ForestWatchers

O ForestWatchers é um projeto de ciência cidadã que tem como objetivo envolver voluntários do mundo todo na tarefa de monitorar o desmatamento das florestas tropicais. As imagens do sensor Modis da NASA e do program PRODES são classificadas por uma rede neural MultiLayer Perceptron e disponibilizadas em uma interface Web no projeto ForestWatchers. A Figura 2.1 apresenta uma visão geral dos vários níveis de sistemas de softwares utilizados no projeto *Forest Watchers*. Nessa arquitetura, observa-se que os voluntários podem acessar as tarefas, via interface Web, selecionando uma das aplicações (*BestTile*, *Deforestation* e *Correct Classification*) oferecidas pelo projeto *Forest Watchers* [Arcanjo \(2014\)](#). Essas aplicações, desenvolvidas em *PyBossa* são descritas a seguir.

- *BestTile*: apresenta doze opções de frações (*Tiles*) de imagens da mesma região e datas distintas que deverão ser votadas e classificadas como melhor opção para uma análise visual da região apresentada.

- *Deforestation*: apresenta as frações das imagens mais votadas pelo voluntário na aplicação *BestTile* e que são apresentadas em uma interface Web especializada para que os voluntários possam verificar as áreas desmatadas.
- *Correct Classification*: apresenta uma imagem em formato de mosaíco que foi previamente classificada por uma Rede Neural, ao lado de sua respectiva imagem MODIS, em uma interface Web para que os voluntários possam confirmar a classificação de Floresta e Não-Floresta realizada pela Rede Neural.

As imagens manipuladas no *Forest Watchers* são armazenadas em um banco de dados desenvolvido em *PyBossa* e *MapServer* e as informações geradas pelos voluntários no servidor de aplicações são coletadas pelo módulo de análise *Analysis Module*. Arcanjo (2014)

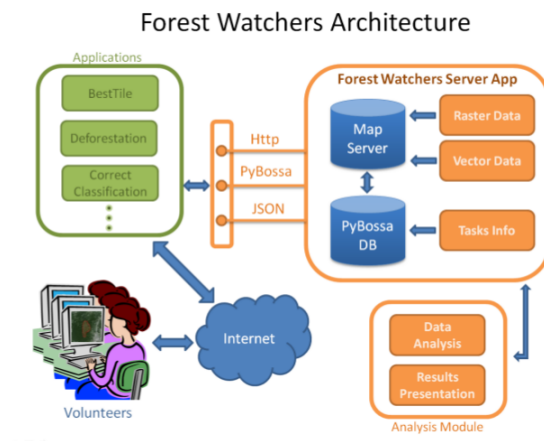


Figura 2.1 - Arquitetura do projeto ForestWatchers.

Fonte: Arcanjo (2014).

A Figura 2.2 apresenta a tela inicial do projeto ForestWatchers Luz e Ramos (2012).



Figura 2.2 - Tela Inicial do projeto ForestWatchers.  
 Fonte: Luz e Ramos (2012).

## 2.3 Dados Utilizados

Os dados de entrada são imagens de baixa resolução espacial obtidas pelo sensor MODIS que está a bordo dos satélites Aqua e Terra da NASA. Esses dados consistem em imagens no formato .hdf e dados no formato shapefile e possuem resolução espacial de 250m. Os dados de saída utilizados são imagens anuais de desmatamento acumulado produzidas pelo projeto PRODES possuem resolução espacial de 120m.

### 2.3.1 Imagens do sensor MODIS que está a bordo dos satélites Aqua e Terra da NASA

As imagens de entrada utilizadas no projeto são disponibilizadas pela NASA em um grupo de imagens de todo o planeta. Essas imagens são fornecidas no formato HDF, em 16 bits e na projeção sinusoidal, como mostra a Figura 2.3.

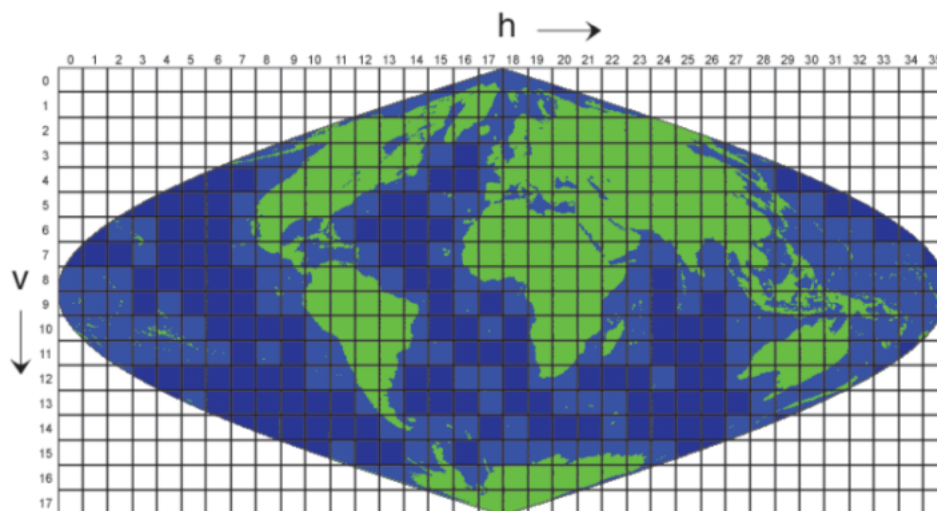


Figura 2.3 - Projeção sinusoidal das imagens da NASA.

Fonte: Luz e Ramos (2012).

O formato HDF é diferente do usado pelo programa DETER do INPE. As imagens HDF tornam-se equivalentes às imagens do DETER (INPE)... (2004), após um conjunto de transformações mostradas na Figura 2.4.

As etapas descritas vão contemplar a elaboração de imagens com base na imagem composta de 8 dias, uma vez que esta imagem será usada para comparar os resultados gerados pelos voluntários. Os passos para a geração de imagens diárias são os mesmos, a diferença reside no banco de dados de HDF. As imagens HDF da NASA foram escolhidas devido a necessidade de se utilizar no projeto imagens com uma resolução temporal muito alta (os satélites Terra e Aqua carregam o sensor MODIS que possui um período de revisita de aproximadamente 1,5 dias).

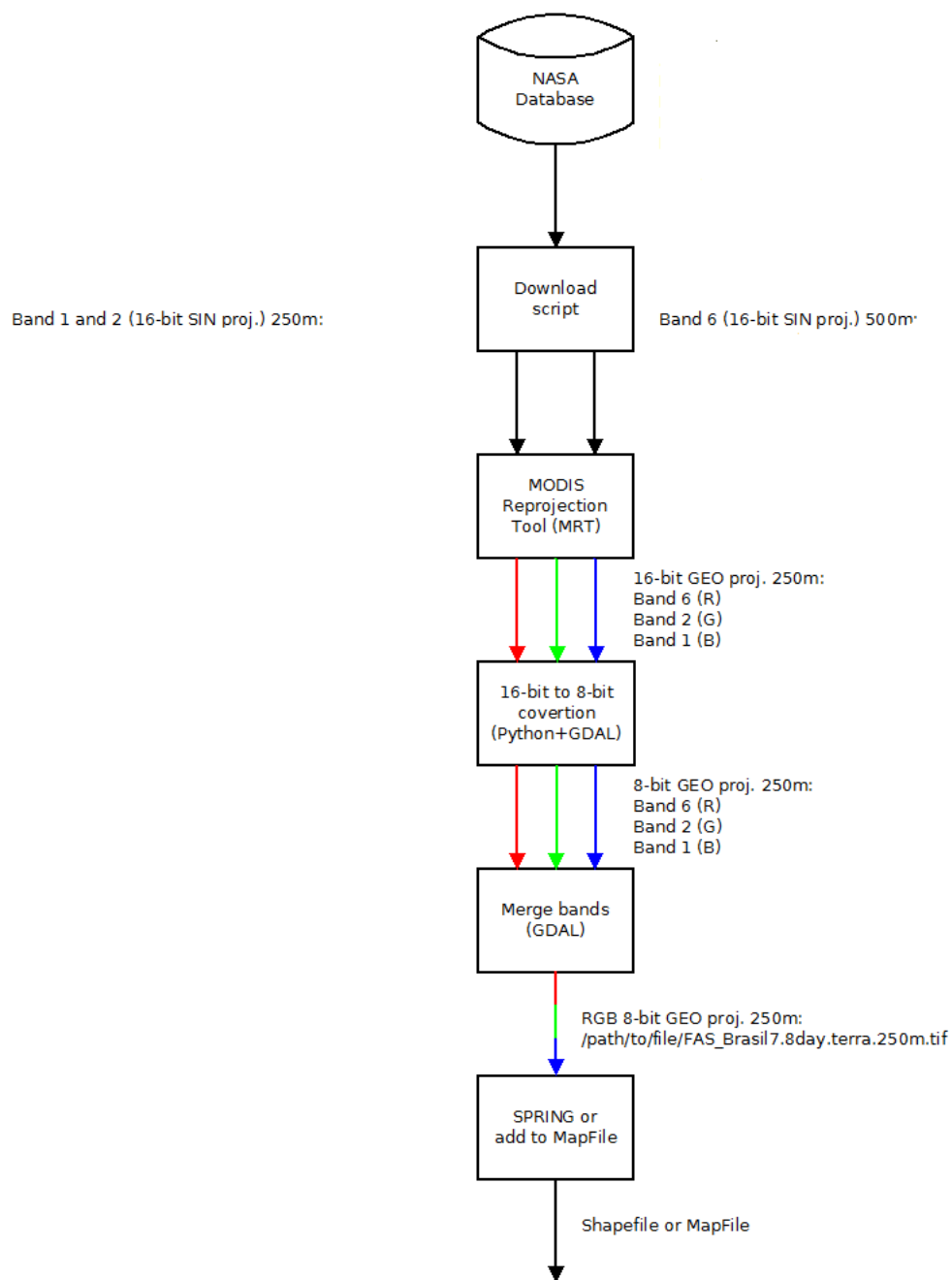


Figura 2.4 - Tranformações das imagens MODIS da NASA.  
 Fonte: adaptação Luz e Ramos (2012).

### 2.3.2 PRODES

O projeto PRODES foi criado em 1988 para gerar taxas anuais de desmatamento da Amazônia Legal. Essas taxas são estimadas a partir dos desmatamentos identificados em cada imagem de satélite que cobre a Amazônia Legal [Instituto... \(1988\)](#). O PRODES utiliza imagens dos satélites da classe Landsat (20 a 30 metros de resolução espacial e taxa de revisita de 16 dias), imagens CCD do CBERS-2, do CBERS-2B, imagens LISS-3, do satélite indiano Resourcesat-1 e imagens do satélite inglês UK-DMC2. Desta forma, utilizando tais imagens, a área mínima mapeada pelo programa é de 6,25 hectares [Instituto... \(1988\)](#). A Figura 2.5 mostra um gráfico com as taxas anuais da Amazônia Legal apresentadas pelo programa PRODES.

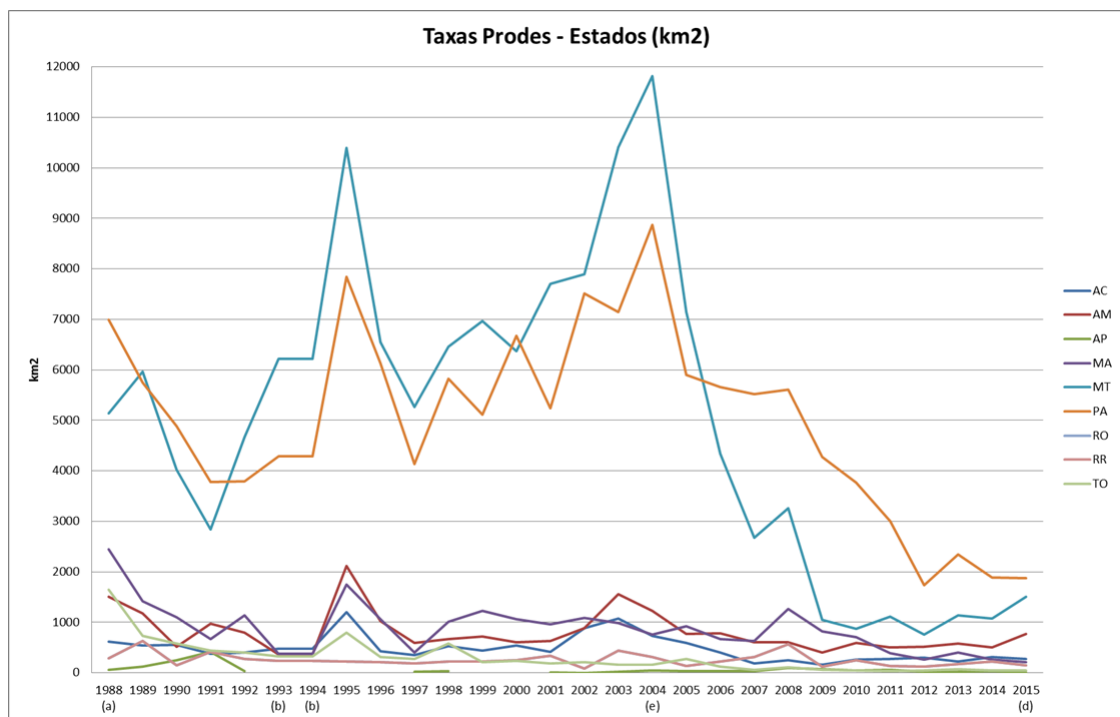


Figura 2.5 - Taxas anuais apresentadas pelo programa PRODES

Fonte: [Instituto... \(1988\)](#).

A classificação das diferentes características das regiões analisadas pelo programa PRODES é basicamente dividida em: floresta, não-floresta (áreas identificadas nas imagens como constituída de vegetação diversa de fisionomia florestal), desflorestamento, hidrografia e nuvem; além da classificação do desmatamento de cada ano. Essa classificação também é exemplificada na Figura 2.6



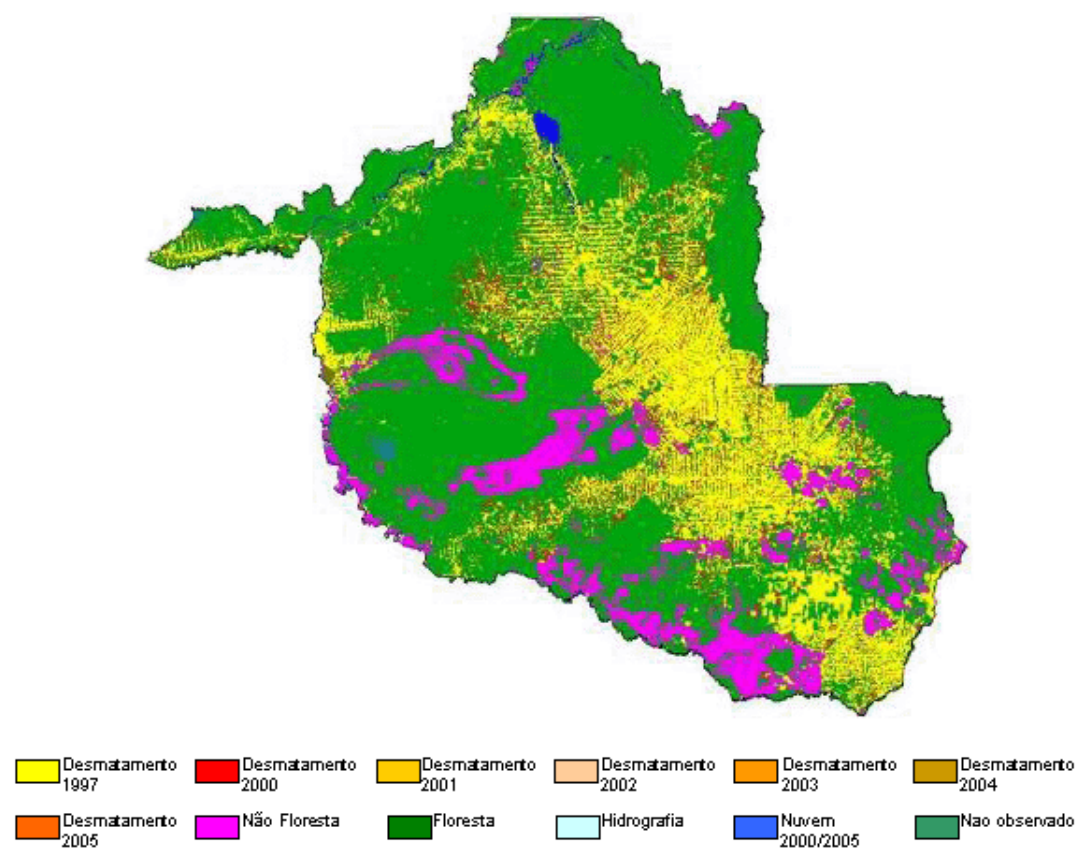


Figura 2.6 - Exemplo de imagem do projeto PRODES para o estado de Rondônia.  
 Fonte: Instituto... (1988).

Neste projeto utiliza-se uma classificação binarizada em Floresta e Não-Floresta da imagem do programa PRODES. Na qual, a classe Floresta (verde) continua Floresta (verde) e as demais classes são consideradas Não-Floresta (vermelho).

### 2.3.3 DETER

O DETER é um sistema de alerta quinzenal de desmatamento realizado pelo INPE e foi operado pela primeira vez em maio de 2004. O sistema utiliza dados do sensor MODIS que está a bordo do satélite Terra/Aqua e do Sensor WFI que está a bordo do satélite CBERS. Os dois sensores possuem resolução espacial de 250 m. Esse sistema apresenta seus dados de forma a facilitar e agilizar as operações de análise para áreas de interesse (INPE)... (2004). A Figura 2.7 uma imagem produto do DETER para o estado de Rondônia.



Figura 2.7 - Exemplo de imagem do programa DETER.

Fonte: (INPE)... (2004).

## 2.4 Redes Neurais

A história das redes neurais inicia-se quase que simultaneamente com a dos computadores eletrônicos programáveis na década de 1940 (KRIESEL, 2007). Durante essa década, várias pesquisas que representavam o cérebro humano com suas formas e sentidos foram divulgadas. Assim, as redes neurais artificiais, segundo (HAYKIN, 2001), são processadores paralelamente distribuídos constituídos de unidades de processamentos simples. Esses processadores possuem a função de armazenar conhecimento específico e disponibilizá-lo para o uso.

O interesse em redes neurais vem de sua capacidade de aprender e responder. Como resultado, as redes neurais têm sido utilizadas em um grande número de aplicações complexas e provaram ser eficazes em uma variedade de campos. Estes incluem o reconhecimento de padrões, classificação, visão, sistemas de controle e predição. A adaptação ou aprendizagem é o principal foco de investigação de RNA atualmente e que fornece um grau de robustez ao modelo. Na modelagem preditiva, o objetivo é mapear um conjunto de padrões de entrada para um conjunto de padrões de saída. Uma RNA realiza essa tarefa através da aprendizagem de uma série de conjunto de dados de entrada/saída apresentados à rede. A rede treinada é então utilizada para aplicar o que foi assimilado durante a aprendizagem para aproximar ou prever a saída correspondente. A estrutura de um neurônio artificial em rede neural é inspirada no conceito de neurônio biológico (CARVALHO, 2011).

O neurônio biológico pode ser visto como o dispositivo computacional elementar do sistema nervoso, composto de muitas entradas e uma saída. As entradas são formadas através das conexões sinápticas que conectam os dendritos aos axônios de outras células nervosas. Os sinais que chegam por estes axônios são pulsos elétricos conhecidos como impulsos nervosos ou potenciais de ação e constituem a informação que o neurônio processa para produzir como saída um impulso nervoso no seu axônio. O elemento básico que forma uma rede neural artificial é o neurônio artificial, conhecido também por nó ou elemento processador (HAYKIN, 2001). Sua criação foi baseada no funcionamento de um neurônio natural, ver figura 2.8(a). A Figura 2.8(b) apresenta um exemplo de um neurônio artificial.

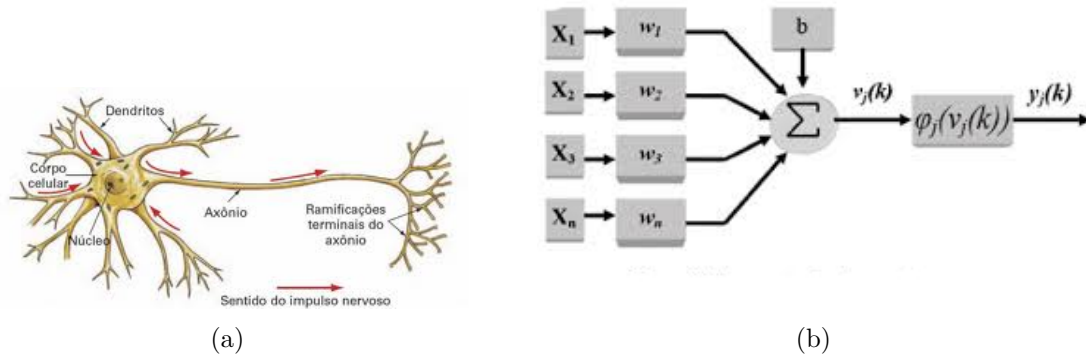


Figura 2.8 - (a) Neurônio Biológico e (b) Neurônio Artificial.

Fonte: [Carvalho \(2011\)](#).

O modelo do neurônio artificial ilustrado na Figura 2.8(b) acima pode ser equacionado da seguinte forma:

$$v_j(k) = \sum_{i=0}^n X_i w_i + b \quad (2.1)$$

$$y_j(k) = \psi_j(v_j(k)) \quad (2.2)$$

onde  $n$  é o número de sinais de entrada do neurônio,  $X_i$  é o  $i$ -ésimo sinal de entrada do neurônio,  $w_i$  é o peso associado com o  $i$ -ésimo sinal de entrada,  $b$  é o limiar de cada neurônio,  $v_j(k)$  é a resposta ponderada do  $j$ -ésimo neurônio em relação ao instante  $k$ ,  $\psi_j$  é a função de ativação para o  $j$ -ésimo neurônio e  $y_j(k)$  é o sinal de saída do  $j$ -ésimo neurônio em relação ao instante  $k$  ([CARVALHO, 2011](#)).

#### 2.4.1 Arquitetura de Redes

As diferentes arquiteturas de RNAs são formadas pela combinação de neurônios artificiais e são definidas pelo tipo de conexão entre as redes. O mais importante caso especial é a arquitetura com alimentação para frente (feedforward) em que os neurônios são divididos em camadas. Cada neurônio transfere o seu sinal apenas para os neurônios que se encontram em uma das camadas subsequentes, ou seja, não há retro acoplamento entre os neurônios. Três tipo de camadas são identificadas:

- Camada de entrada: é a interface de entrada, onde os sinais de entrada externos são alimentados para dentro da rede.

- Camada de saída: os sinais de saída da rede são capturados a partir dos nós de saída.
- Camadas Ocultas: Definem a representação interna do mapeamento e não tem ligação direta com o usuário externo.

### 2.4.2 MLP - MultiLayer Perceptron

Em uma rede MultiLayer Perceptron cada unidade executa uma soma ponderada de suas entradas e transmitem esse nível de ativação através de uma função de transferência para produzir uma saída, e as unidades são arranjadas em uma topologia de camadas. A rede, portanto, tem uma interpretação simples como uma forma de modelo entrada-saída, com os pesos e vieses como parâmetros livres do modelo. Tais redes podem modelar funções de complexidade arbitrária com o número de camadas e o número de unidades em cada camada determinando a complexidade da função. A aprendizagem da rede MLP é realizada utilizando o algoritmo de retropropagação. Esse algoritmo baseia-se em dois passos básicos: a propagação, um padrão de entrada é apresentado e seu resultado é propagado camada a camada. Nessa etapa os pesos sinápticos são fixos e ao final é liberado um conjunto de saída da rede, e a retropropagação, nessa etapa a saída da rede é comparada à saída que se deseja para se ter o parâmetro de correção do erro. Os pesos são ajustados de acordo com o resultado da correção do erro. Esse ajuste é aplicado camada a camada - da camada de saída até a camada de entrada (HAYKIN, 2001).

A Figura 2.9 apresenta um exemplo de um modelo computacional para a rede MLP.

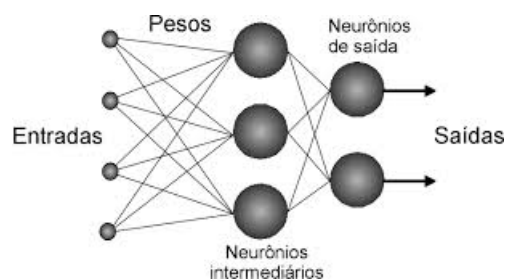


Figura 2.9 - Modelo Computacional da Rede MLP.

Fonte: Adaptação *Carvalho (2011)*.

## 2.5 Sistemas Nebulosos

A teoria dos sistemas nebulosos e da lógica nebulosa iniciou-se com o desenvolvimento das ideias de Loft Zadeh a partir de 1965 para a resolução de problemas que estão sujeitos a incertezas [Sandri e Correa \(1999\)](#). Esses sistemas são compostos por conjuntos nebulosos que segundo [Zadeh \(1965\)](#) são uma classe de objetos caracterizados por uma função  $f_A(x)$  que relaciona cada objeto a um valor  $X$  real ao intervalo  $[0 - 1]$ , no qual se considera  $X$  um espaço de pontos em que  $X = x$ .

### 2.5.1 Rede Neural Nebulosa

A Rede Neural Nebulosa ou sistema Neuro-Fuzzy consiste em uma técnica de aprendizado de máquina que utiliza os parâmetros de um sistema nebuloso com as técnicas de aproximação de redes neurais na classificação de informações [Kruse \(2008\)](#). Os sistemas Neuro-Fuzzy podem ser representados como mostra a Figura 2.10.

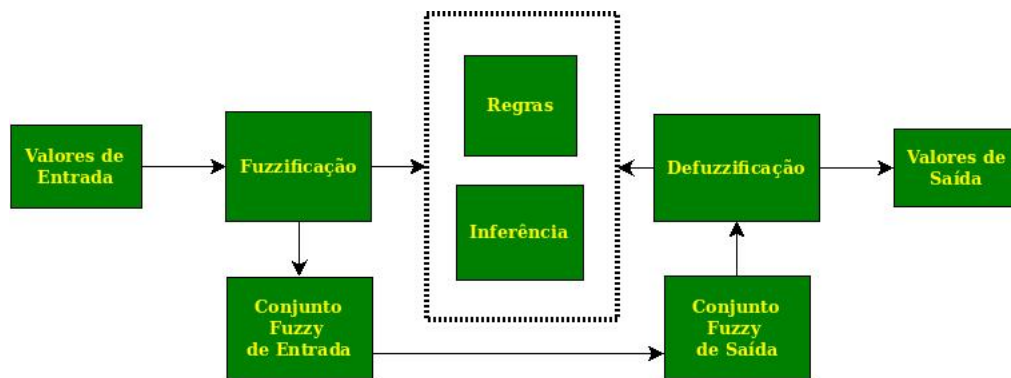


Figura 2.10 - Modelo de um sistema Neuro-Fuzzy.

Fonte: [Rodrigues et al. \(2011\)](#).

Esses sistemas Neuro-Fuzzy são classificados, segundo [Pimentel \(2014\)](#), em:

- **Fuzzificação em nível de rede:** nesse método os neurônios, arquitetura e a saída desejada são difusos.
- **Fuzzificação em nível de aprendizagem:** nesse método a fuzzificação ocorre no algoritmo de aprendizagem, que pode ser chamado de retropropagação difusa.
- **Fuzzificação em nível de rede e de aprendizagem:** nesse método tanto os neurônios, arquitetura e saída desejada quanto o algoritmo de



aprendizagem são difusos.

No perceptron multicamadas difuso a propagação dos erros considera a relação da semelhança exata gerada na saída desejada difusa [Pimentel \(2014\)](#). A utilização das MLP's difusas permite classificar com uma maior precisão os píxels com alto grau de incerteza presentes nas imagens de satélites. A Figura 2.11 apresenta um perceptron multicamadas difuso.

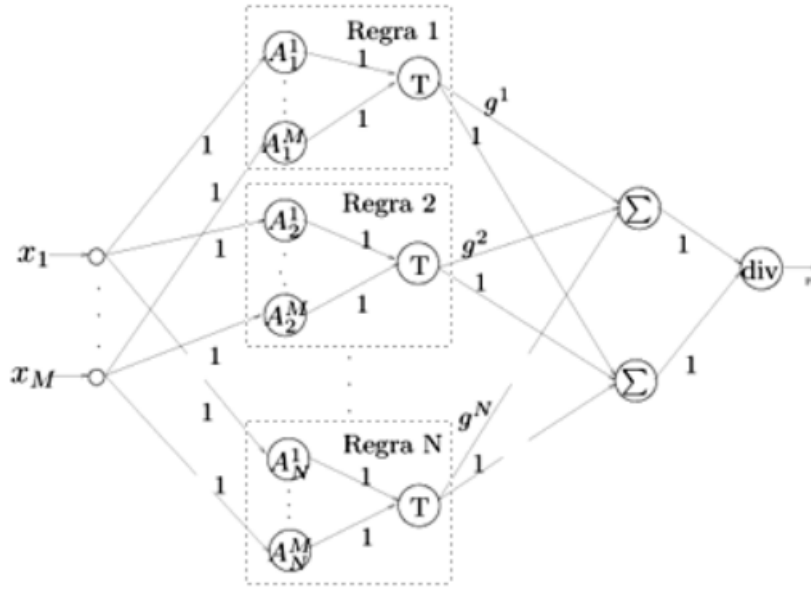


Figura 2.11 - Exemplo de um perceptron multicamadas difuso.  
Fonte: [Ballini e Carneiro \(2003\)](#).

onde,

- $x_{j,j=1,\dots,M}$  – é uma variável fuzzy.
- $M$  – é o número de entradas.
- $A_j^i$  – é um conjunto nebuloso.
- $T$  – é a função de Ativação
- $g^i$  – é um número real definido no espaço de saída.
- $N$  – é o número de regras.

## 2.6 Árvore de Decisão

Segundo [Horning \(2013\)](#), a árvore de decisão é um modelo preditivo que utiliza um conjunto de regras quantitativas ou qualitativas para calcular um valor alvo. Em uma árvore de decisão cada nó não-folha representa um teste de um atributo, os ramos representam os resultados dos testes, os nós folhas ou nós terminais representam uma classe e o nó mais ao topo representa o nó raiz [Han et al. \(2012\)](#). A Figura 2.12 apresenta uma árvore de decisão e seus componentes.

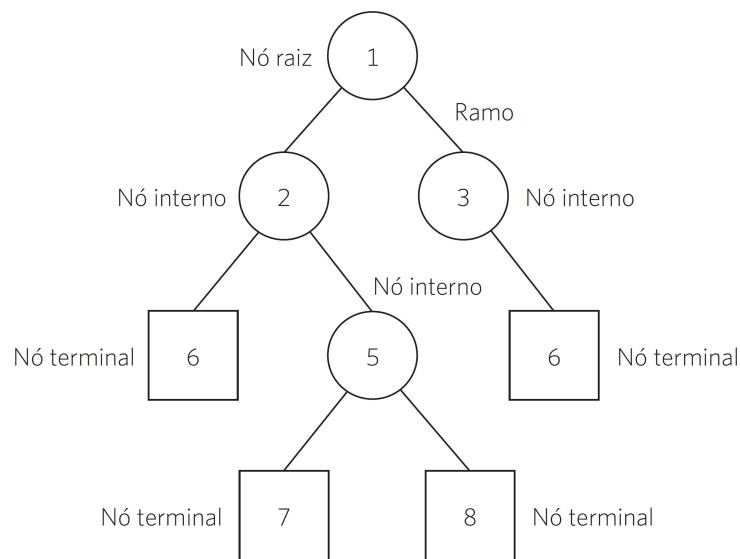


Figura 2.12 - Árvore de decisão e seus componentes.

Fonte: [Medeiros et al. \(2014\)](#).

Neste projeto utilizamos a árvore de decisão para a classificação de imagens. Um exemplo da utilização da árvore de decisão para a classificação de imagens pode ser observado na Figura 2.13. Na árvore de decisão da Figura 2.13, utiliza-se as características *band4* e *band3* e as classes *Water*, *Cloud*, *Non-Forest*, *Forest* e *Scrub*.



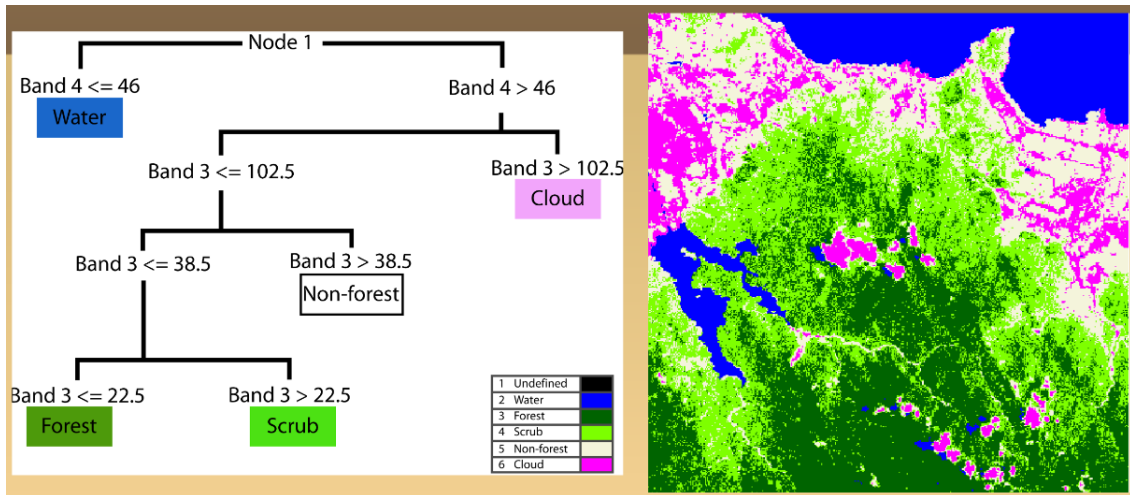


Figura 2.13 - Exemplo da aplicação da Árvore de Decisão.  
 Fonte: [Horning \(2013\)](#).

## 2.7 Árvores Aleatórias ou Random Forest

A *Random Forest* é técnica de aprendizado de máquina desenvolvida por [Breiman \(2001\)](#). Essa técnica permite que se obtenha modelos muito eficazes sem nenhuma preparação de dados ou conhecimento de modelagem [Breiman e Cutler \(2014\)](#). Uma Random Forest pode ser decrita como um classificador formado por um conjunto de árvores de decisão  $\{h(\mathbf{X}, v_k), k, 1, ..\}$ , onde  $v_k$  são vetores aleatórios amostrados de forma independentes, distribuídos igualmente em todas as árvores da floresta. O resultado do processo de classificação é a classe  $X$  com maior número de votos dentre todas as árvores consideradas ([HAN et al., 2012](#)). A Figura 2.14 exemplifica uma Random Forest.

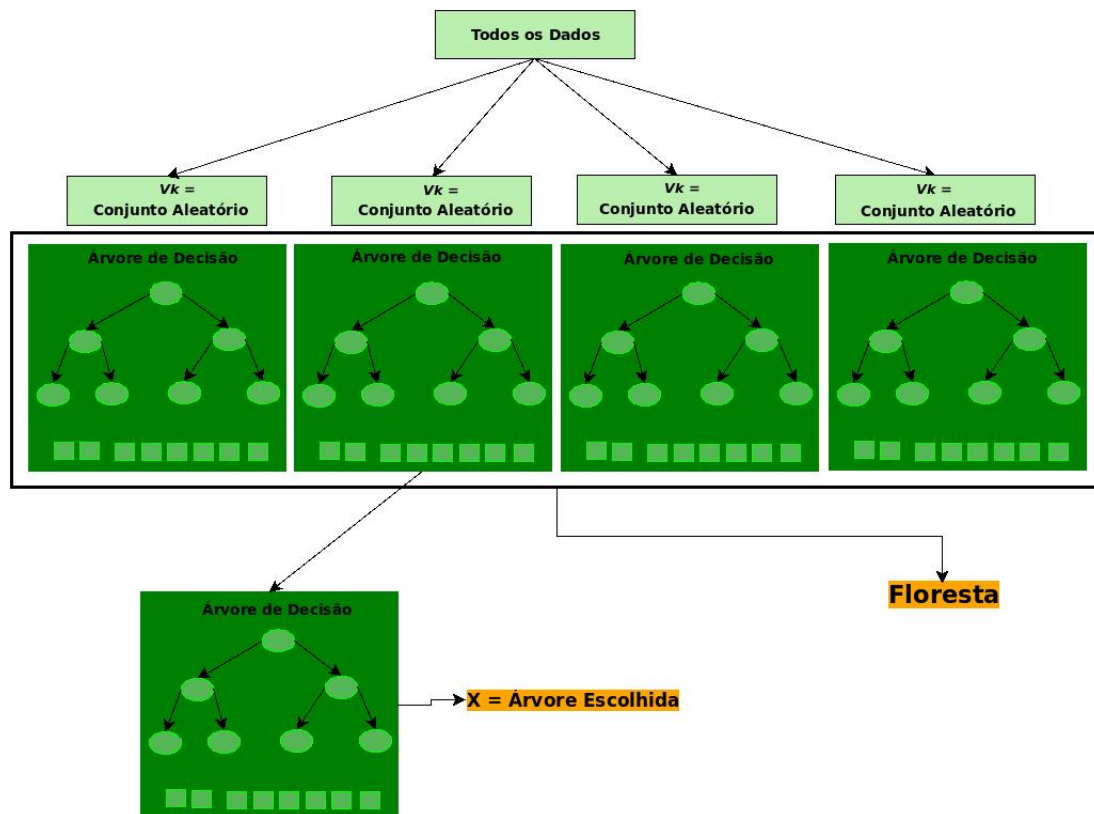


Figura 2.14 - Exemplo de Random Forest.

Fonte: adaptação [Benyamin \(2012\)](#).

Os dados de entrada para o desenvolvimento da técnica *Random Forest* são selecionados de forma aleatória utilizando a técnica *bootstrap*. Segundo [Schmidheiny \(2012\)](#), essa técnica basicamente é utilizada para estimar os parâmetros de interesse e simular a partir de uma distribuição estimada a distribuição assintótica da estatística de interesse. Os próximos passos para o desenvolvimento da *Random Forest* são: realizar a seleção do conjunto de dados para o crescimento das árvores, a seleção das características dos preditores, o crescimento da árvore, o qual é executado até o atendimento de um critério de parada e por fim a estimativa do erro de crescimento da árvore utilizando o *out-of-bag*. No *out-of-bag* utiliza-se o elemento de maior valor do conjunto de dados classificado, que restou da seleção *bootstrap*, e em seguida verifica-se a diferença com o valor da árvore que recebeu o maior número de votos [Breiman \(2001\)](#). A Figura 2.15 ilustra os passos do desenvolvimento da técnica *Random Forest*.

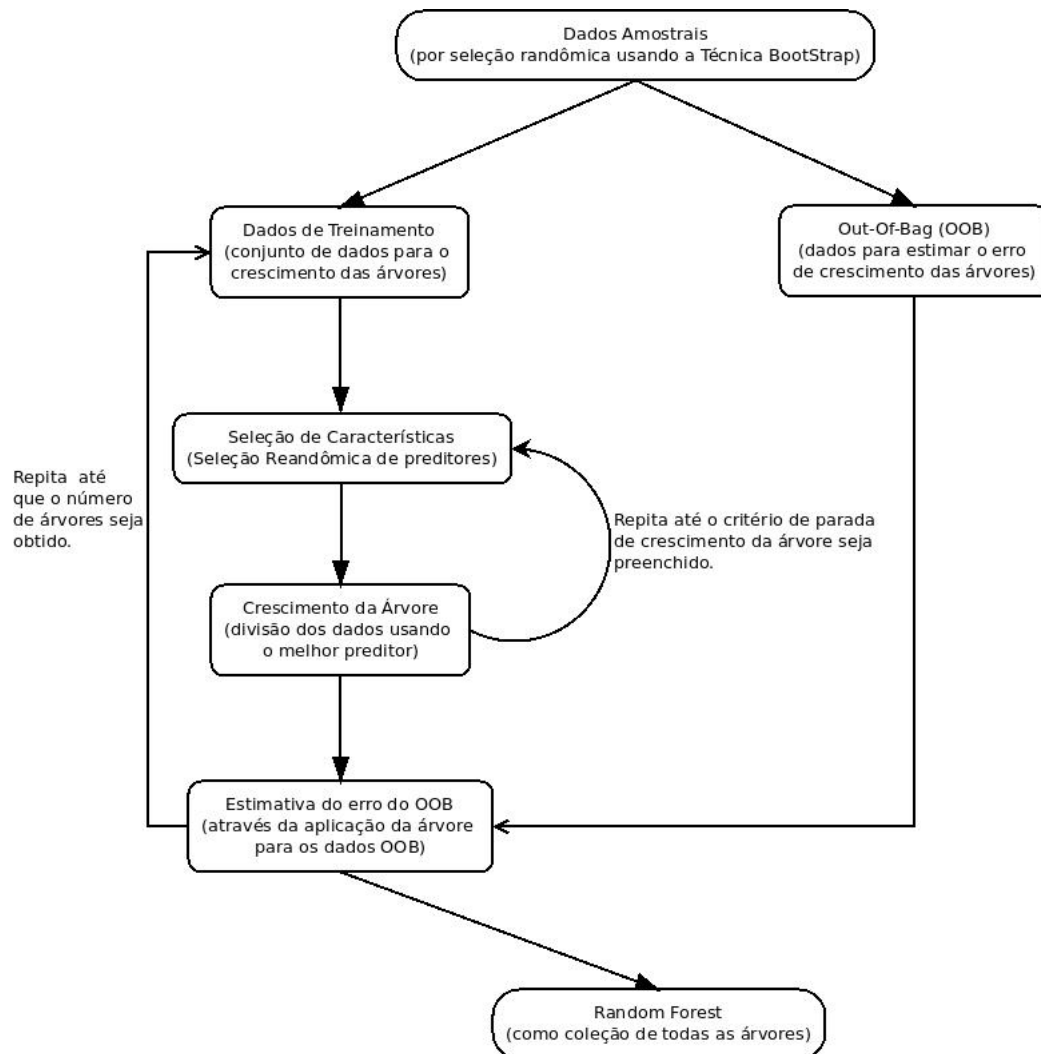


Figura 2.15 - Algoritmo para uma *Random Forest*.  
Fonte: adaptação Barth (2013).

Conforme Breiman e Cutler (2014), o crescimento da árvore é realizado segundo o algoritmo a seguir:

- Se o número de casos no conjunto de treinamento é  $N$ , amostra-se  $N$  casos ao acaso a partir dos dados originais. Essa amostra será o conjunto de treinamento para o crescimento da árvore.
- Se existem  $M$  variáveis de entrada, um número  $m \ll M$  é especificado, de modo que em cada nó  $m$  variáveis sejam selecionadas aleatoriamente dentro do valor de  $M$  e a melhor subdivisão de  $m$  é utilizada para dividir o nó. O valor de  $m$  é mantido constante durante o crescimento florestal.

- Cada árvore é cultivada na maior extensão possível. Não existindo eliminação de nós ou poda. [Breiman e Cutler \(2014\)](#)

Neste projeto iremos utilizar o método de fuzzificação em nível de rede para fazer uma comparação com a técnica *Random Forest* na classificação das imagens, pois nesse método o vetor de saída desejada representa o grau de similaridade do padrão de entrada para as classes de saída. Esse grau de similaridade possibilita uma melhor classificação dos píxels com variação radiométrica aproximada das imagens de satélites. Para o desenvolvimento da fuzzificação em nível de rede pretende-se aplicar os conceitos de conjuntos difusos na camada de neurônios, na qual a utilização de um algoritmo de retropropagação padrão seria empregado para ajustar os parâmetros do sistema difuso. Também se faz necessário a fuzzificação do vetor de saída desejada [Pimentel \(2014\)](#).

## 2.8 Weka - Waikato Environment for Knowledge Analysis

No projeto utiliza-se a ferramenta de mineração Weka para a aplicação da técnica Random Forest. O Weka é um software de mineração de dados, desenvolvido na linguagem de programação Java, que possui um conjunto de algoritmos referentes a linguagem de máquina. Esses algoritmos podem ser utilizados por meio da interface gráfica da ferramenta ou pela utilização de API na linguagem Java no desenvolvimento de software. O Weka permite realizar o pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização, como também o desenvolvimento de sistemas de aprendizagem de máquina [Hall et al. \(2014\)](#). A interface gráfica do Weka é denominada *Weka Gui Chooser* apresenta como principais opções de seleção para o desenvolvimento de suas funções quatro botões (*Explorer*, *Experimenter*, *KnowledgeFlow* e *Simple CLI*) como apresentado na Figura 2.16 [The University Waikato \(2015\)](#).



Figura 2.16 - Exemplo da tela inicial do *Weka*.

Fonte: adaptação [Hall et al. \(2014\)](#).

Os botões do *Weka Gui Chooser* são descritos como segue:

- Explorer: Um ambiente para explorar dados com Weka (Figura 2.17).
- Experimenter: Um ambiente para a realização de experimentos e realização de testes estatísticos entre os esquemas de aprendizagem (Figura 2.18).
- KnowledgeFlow: Este ambiente desenvolve basicamente a mesma função que o Explorer, mas com uma interface drag-and-drop. Uma de suas vantagens é suportar aprendizagem incremental (Figura 2.19).
- SimpleCLI: Fornece uma interface de linha de comando simples para uma execução direta de comandos do Weka para sistemas operacionais que não fornecem sua própria interface de linha de comando (Figura 2.20).

As Figuras a seguir apresentam as opções do Weka.

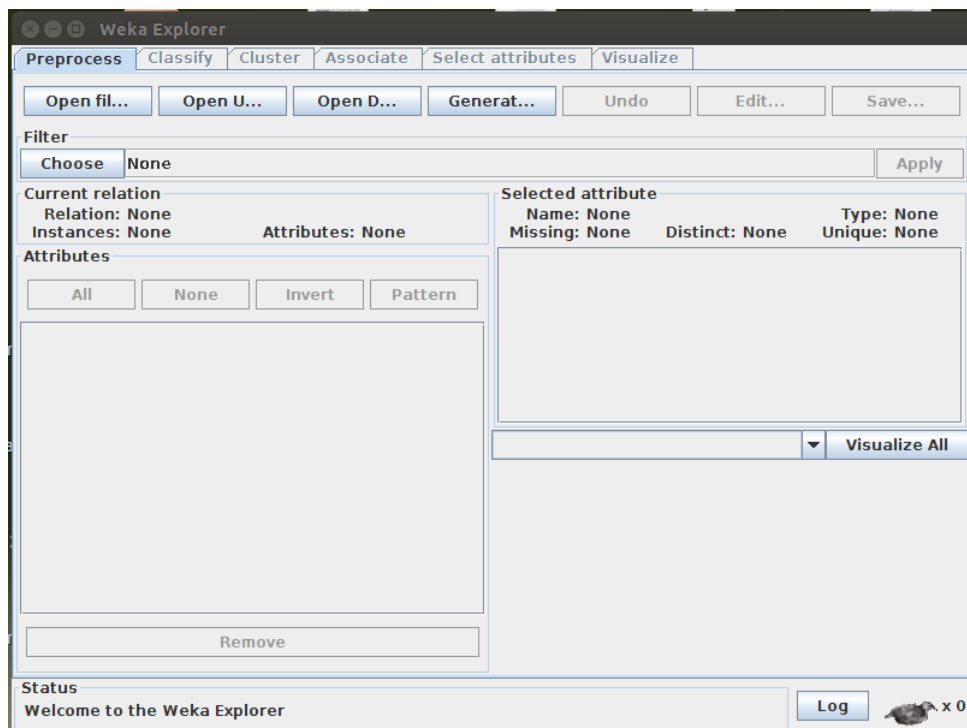


Figura 2.17 - Exemplo da tela Explorer do Weka.

Fonte: adaptação Hall et al. (2014).

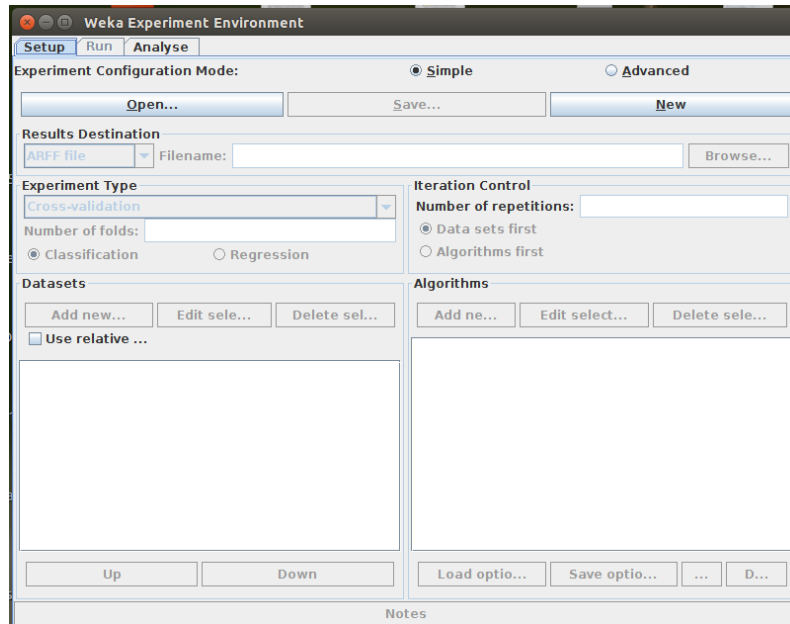


Figura 2.18 - Exemplo da tela Experimenter do Weka.  
 Fonte: adaptação Hall et al. (2014).

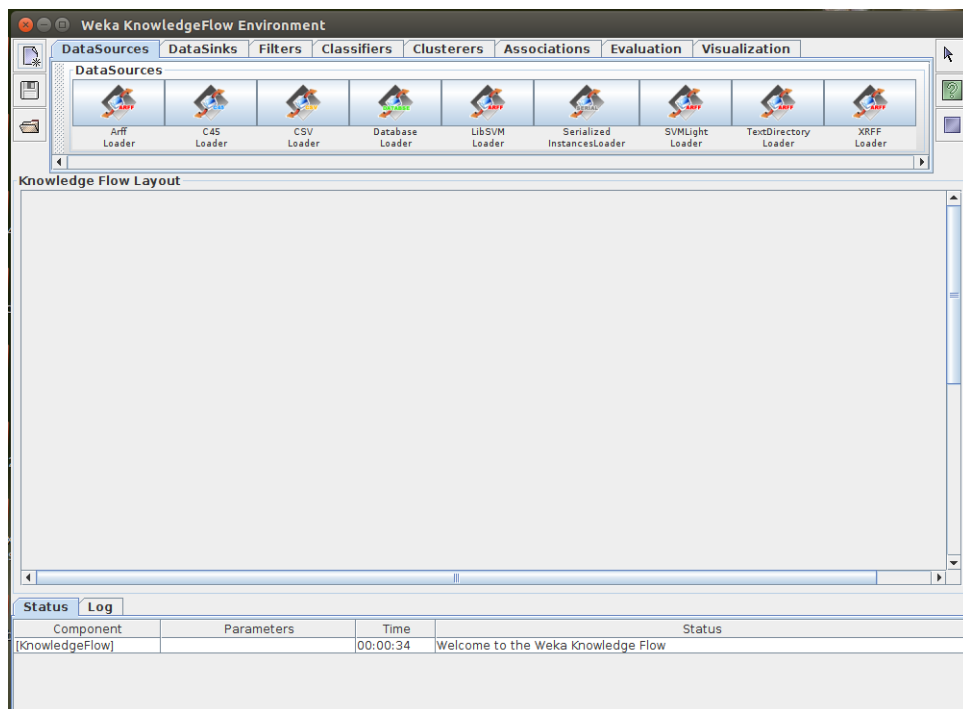


Figura 2.19 - Exemplo da tela KnowledgeFlow do Weka.  
 Fonte: adaptação Hall et al. (2014).

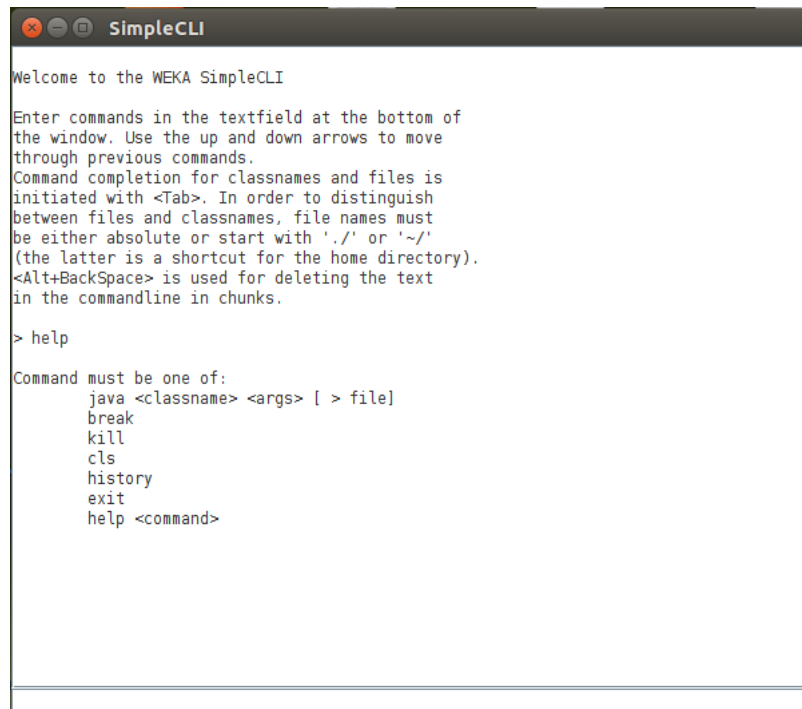


Figura 2.20 - Exemplo da tela SimpleCLI do Weka.

Fonte: *adaptação* Hall et al. (2014).

A API em Java para o desenvolvimento de softwares com as funcionalidade do Weka podem ser encontradas em Hall et al. (2014).

## 2.9 GDAL

A GDAL é uma biblioteca de leitura e escrita de dados geoespaciais referenciados liberada sob a licença X / MIT similar a uma licença da *Open Source Geospatial Foundation*. Essa biblioteca possui dois modelos abstratos de dados: um para raster e um para vetor. A GDAL suporta 142 formatos para o modelo raster e 84 formatos para o modelo vetor. A GDAL também possui um conjunto de funções em linha de comando para processamento e conversão dos dados Geospatial... (2007) e uma API de desenvolvimento de software para a linguagem C++ que foi utilizada no projeto. Neste trabalho a biblioteca GDAL é utilizada para realizar a leitura e/ou em RGB dos valores de cada píxel das imagens de satélite analisadas.

## 2.10 Matriz de Confusão

A matriz de confusão é uma ferramenta útil para analisar e verificar como o classificador utilizado está reconhecendo de forma precisa tuplas de diferentes classes. Na construção da matriz de confusão utiliza-se os termos TP(True Positive), TN(True

Negative), FP(False Positive) e FN(False Negative) que são definidos como segue Han et al. (2012):

- TP: Referem-se as tuplas positivas que foram corretamente classificadas pelo classificador.
- TN: Referem-se as tuplas negativas que foram corretamente classificadas pelo classificador.
- FP: Referem-se as tuplas negativas que foram incorretamente classificadas como positivo.
- FN: Referem-se as tuplas positivas que foram classificadas de forma errada como negativas.

A Figura 2.21 apresenta um exemplo de Matriz de Confusão.

		preditivo		
atual		TP	FN	$P=TP+FN$
		FP	TN	$N=FP+TN$
		$P'=TP+FP$	$N'=FN+TN$	$Total=P+N=P'+N'$

Figura 2.21 - Exemplo de Matriz de Confusão.

Fonte: *adaptação* Han et al. (2012).

A Matriz de Confusão é utilizada no projeto para a validação da classificação das imagens.

## 2.11 Entropia da Imagem

A entropia de imagens representa a irregularidade ou despadronização da imagem analisada Brasil (2010). Neste trabalho estamos utilizando a técnica de Entropia de Kullback-Leibler para calcular o grau de incerteza entre as imagens resultantes e a verdade PRODES.



### 2.11.1 Divergência ou Entropia de Kullback-Leibler

A Entropia de Kullback-Leibler representa a distância entre duas funções discretas de probabilidade  $p$  e  $q$  [Weisstein \(2000\)](#). A Entropia Kullback-Leibler é representada pela fórmula 2.3.

$$d = \sum_k p_k \log_2 \frac{p_k}{q_k} \quad (2.3)$$

em que, nesse caso

$p_k$  = probabilidade da função discreta  $p$

$q_k$  = probabilidade da função discreta  $q$

No Capítulo 3 será apresentado a metodologia desenvolvida para a aplicação dos dados e teorias apresentados nesse capítulo.

### 3 METODOLOGIA

Neste trabalho são estudadas diferentes abordagens de arquiteturas de redes neurais e técnicas utilizadas em reconhecimento de padrões e publicadas na literatura. Em particular, redes do tipo MLP, MLP-difusa e a técnica *Random Forest* são investigadas. A metodologia utilizada para a classificação das imagens em regiões de Floresta e Não-Floresta utilizando as técnicas pode ser observada no fluxo de atividades apresentado na Figura 3.1. As etapas apresentadas nesse fluxo são detalhadas nas próximas seções.

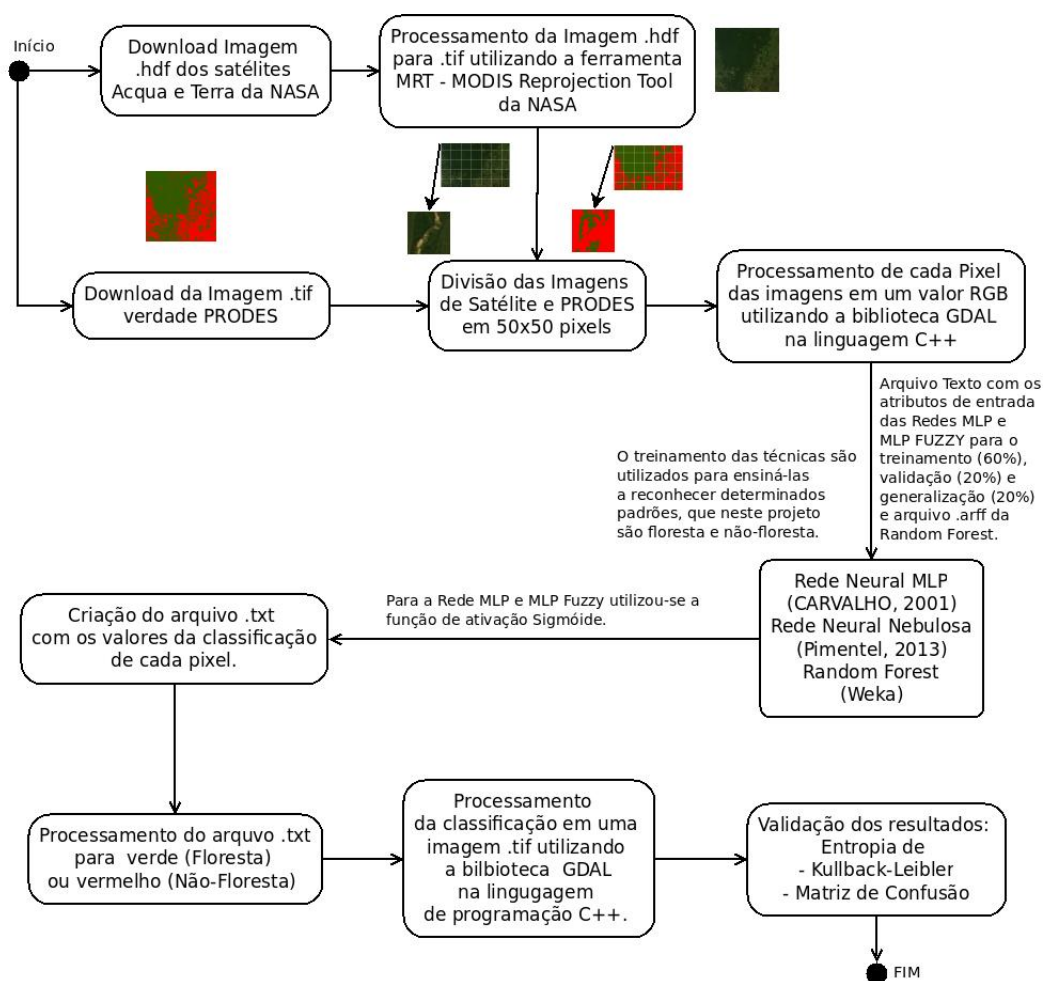


Figura 3.1 - Metodologia utilizada no trabalho.

### 3.1 Aquisição e Reprojeção das Imagens

#### 3.1.1 Imagem do sensor Modis dos Satélites Acqua e Terra da NASA

A aquisição das imagens dos sensor Modis dos Satélites Acqua e Terra é feita realizando o download no site da NASA. Após o download as imagens são reprojetadas de 16 bits para 8 bits e para 250 *m* utilizando a ferramenta da NASA MRT (Figura 3.2).

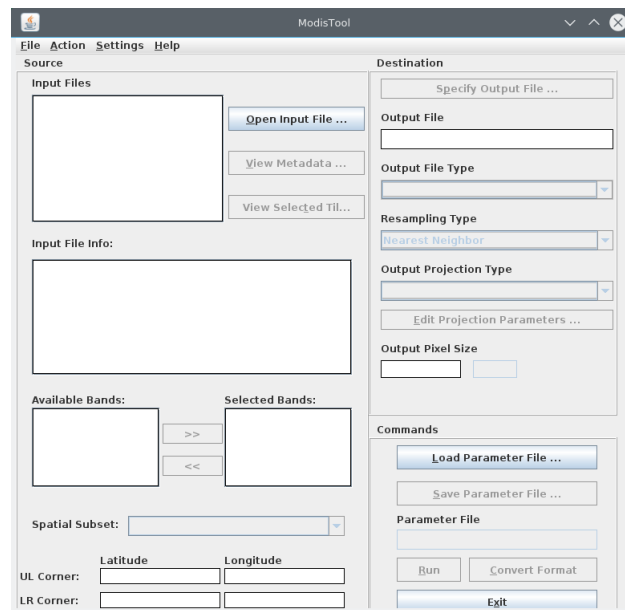


Figura 3.2 - Janela da Ferramenta Modis Reprojection Tool - MRT.

#### 3.1.2 Imagem PRODES

Depois de realizado o download das imagens PRODES, elas são binarizadas em duas classes, Floresta (verde) e não-Floresta (vermelho) e reprojetadas para 250 *m*. Para ambos os processos, binarização e reprojeção, são utilizados a biblioteca *GDAL Geospatial...* (2007). A Figura 3.3 apresenta a comparação de uma imagem binarizada PRODES com a imagem original PRODES.

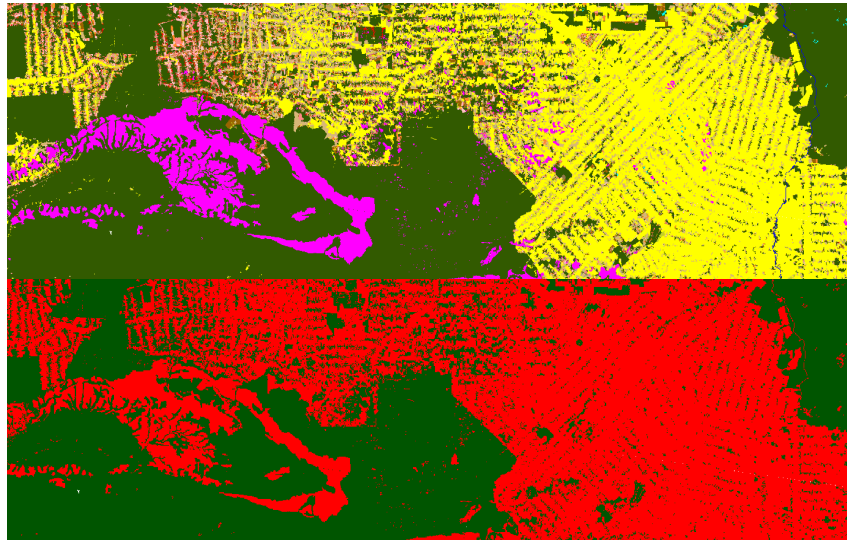


Figura 3.3 - Comparação de uma imagem binarizada com a imagem original PRODES.

### 3.2 Subdivisão das Imagens

As imagens da NASA e do PRODES possuem dimensões aproximadas de 5000x6000 e 3200x2605 pixels. Para que o processamento dessas imagens tivesse um melhor desempenho computacional, elas foram divididas em imagens com dimensões de 50x50 pixels para os testes de pequena e média escala. Para o processo de leitura e divisão das imagens utilizou-se a biblioteca de georreferenciamento *GDAL Geospatial...* (2007). A Figura 3.4 apresenta um exemplo de uma imagem dividida Modis e PRODES.

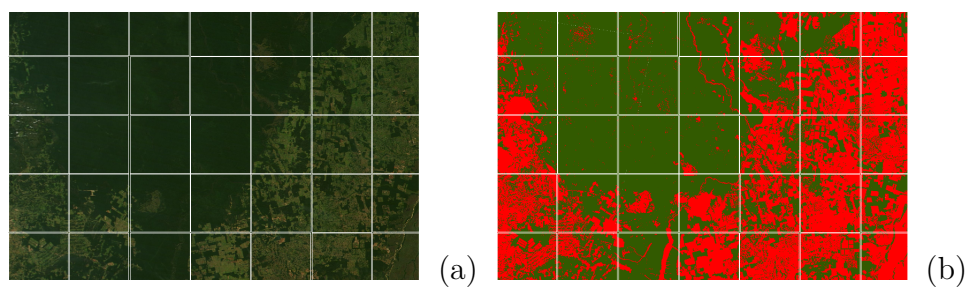


Figura 3.4 - Exemplo da divisão da imagens Modis e PRODES.

### 3.3 Transformação dos Pixels em valores RGB

As imagens divididas em 50x50 são processadas por um software, desenvolvido utilizando a biblioteca GDAL na linguagem C++, para a obtenção do valor RGB de cada pixel. O resultado desse processamento é um arquivo .txt contendo, em cada linha, o valor RGB do pixel da imagem do MODIS seguido pela representação do valor RGB do pixel correspondente da imagem PRODES (Tabela 3.1). As funções utilizadas da biblioteca GDAL são apresentadas de forma simplificada no trecho de código na linguagem C++ a seguir.

```
//Registro dos drivers de todos formatos suportados
GDALAllRegister();

//Abertura do arquivo da imagem
poDataset = (GDALDataset *) GDALOpen( name.c_str(), GA_ReadOnly );
if( poDataset == NULL )
    cout<<"Erro na abertura do arquivo "<<nomearquivo<<"!"<<endl;

//Leitura das bandas Red,Green e Blue da imagem
poDataset->GetGCPs()->dfGCPPixel;

//Coleta das Informações para a banda Red
poBandR = poDataset->GetRasterBand( 1 );
poBandR->GetBlockSize( &nBlockXSize, &nBlockYSize );
nXSize = poBandR->GetXSize();
nYSize = poBandR->GetYSize();
pafScanlineR = (float *) CPLMalloc(sizeof(float)*nXSize*nYSize);
poBandR->RasterIO( GF_Read, 0, 0, nXSize,nYSize,pafScanlineR,
nXSize, nYSize, GDT_Float32,0, 0 );

//Armazenamento dos valores da banda Red
for(int sc=0; sc < nXSize*nYSize; sc++)
    this->setRed(pafScanlineR[sc], sc);
```

Tabela 3.1 - Dados dos píxels em valores RGB

R	G	B	Verdade Prodes
46	55	8	1
46	56	8	1
49	65	20	1
49	65	20	1
49	62	19	1
58	55	9	1
58	53	9	0
47	53	9	0
47	36	9	1
71	47	6	1
71	52	9	1
71	52	9	0
68	50	11	0

### 3.4 Processamento dos dados RGB na Rede MLP, Rede MLP Nebulosa e Random Forest

O arquivo .txt com as informações RGB dos píxels das imagens possui uma configuração própria para cada técnica utilizada na classificação das imagens. Nas próximas seções será apresentado a configuração de cada arquivo e sua aplicação nos métodos.

#### 3.4.1 Processamento na Rede MLP

A Rede Neural MLP utilizada no projeto foi desenvolvida por [Carvalho \(2011\)](#). Para a utilização na classificação das imagens diversas configurações de entrada da rede foram testadas e apresentadas na Tabela 3.2 juntamente com melhor configuração encontrada (em vermelho).

Tabela 3.2 - Dados dos testes com a configuração da Rede MLP

Entrada	Camadas Escondidas	Neurônios de Saída	Momentum Momentum	Função de Ativação
3	2	3	0.2	Sigmóide
3	2	3	0.7	Sigmóide
3	2	3	0.90625	Sigmóide
3	2	3	0.75	Tangente
3	2	1	0.2	Sigmóide
3	2	1	0.7	Sigmóide
3	2	1	0.90625	Sigmóide
3	2	1	0.75	Tangente

A divisão dos dados para o treinamento, validação e generalização da rede é feita utilizando a teoria de [Haykin \(2001\)](#) adaptada para a quantidade de pixels de cada arquivo .txt de entrada da rede.

- Treinamento - 60 % da quantidade de pixels de entrada.
- Validação - 20 % da quantidade de pixels de entrada.
- Generalização - 20 % da quantidade de pixels de entrada.

Um exemplo do arquivo de entrada com as configurações da Rede Neural MLP é apresentado na [Figura 3.6](#).

```

saveNetworkFile=NetS.ser
inputTrainFile=/home/marilyn/Documentos/Mestrado/resultados/MLP/15Imagens/MOD09.A2011169.h11v10_9511.tif_7
inputValidFile=/home/marilyn/Documentos/Mestrado/resultados/MLP/15Imagens/MOD09.A2011169.h11v10_9511.tif_7
inputGenFile=/home/marilyn/Documentos/Mestrado/resultados/MLP/15Imagens/MOD09.A2011169.h11v10_9511.tif_7
resultFile=/home/marilyn/Documentos/Mestrado/resultados/MLP/ResultadoRede15/
resultado_MOD09.A2011169.h11v10_9511.tif_7.txt
outputFileValidation=outputValid_s1.txt
outputFileTraining=outputTraining_s1.txt
numEpochs=5000
minErrorValue=0.0001
numTrainingPatterns=2000
numValidationPatterns=500
minInputValue=0
maxInputValue=255
minOutputValue=0
maxOutputValue=1
firstRowTraining=1
lastRowTraining=2000
firstRowValidation=2001
lastRowValidation=2500
firstRowGeneralization=1
lastRowGeneralization=2500
firstColInput=1
lastColInput=3
firstColTarget=4
lastColTarget=4
inputNeurons=3
outputNeurons=1
numMaxLayers=2
nBitsLR=6
nBitsMM=5
nBitsNumNeurons=4
nBitsAFTType=2
randomInit=false
LR=0.2
Momentum=0.7
AFTType=0
numHiddenLayers=2
hiddenNeurons1=5
hiddenNeurons2=5
dynamicAnnealing_Rate=10
dynamicAnnealing_Step=15

```

Figura 3.5 - Exemplo do arquivo de configuração da Rede.

A configuração do arquivo .txt com as informações dos valores dos pixels em RGB para a Rede Neural MLP é a mesma apresentada na Tabela 3.1.

Como resultado da classificação da Rede Neural MLP é gerado um arquivo .txt com o modelo de informação apresentada na Figura 3.3

Tabela 3.3 - Dados dos testes com a configuração da Rede MLP

Resultado da Classificação dos Pixels
0.991226356671217
0.9917118403318257
0.988823449978975
0.8714325696801124



### 3.4.2 Processamento da Rede MLP Nebulosa

A Rede Neural MLP Nebulosa utilizada no projeto foi desenvolvida por [Pimentel \(2014\)](#). A mesma configuração de entrada da Rede Neural é utilizada para a Rede Neural MLP Nebulosa na classificação dos pixels em Floresta e Não-Floresta. Para o processamento da Rede MLP Nebulosa utiliza-se o mesmo arquivo .txt com as informações dos valores RGB dos pixels utilizado pela Rede Neural MLP - Tabela 3.1.

O resultado do processamento do arquivo de entrada com os valores dos pixels em RGB da Rede Neural Nebulosa é um arquivo .txt com a classificação da cada pixel em Floresta (1) e Não-Floresta (0).

### 3.4.3 Processamento da Random Forest

No projeto, a técnica Random Forest foi aplicada utilizando a API de desenvolvimento Java da ferramenta de mineração de dados *Weka*. O arquivo com as informações de entrada da Random Forest possui a configuração apresentada a seguir.

```
%      Entrada para Random Forest
%      R - Red
%      G - Green
%      B - Blue

@relation deflorest

@attribute Red integer
@attribute Green integer
@attribute Blue integer
@attribute truePRODES {0,1}

@data
74,58,14,0
56,55,13,0
56,54,8,0
51,53,6,1
51,53,6,1
51,59,10,0
70,63,14,0
70,61,15,0
71,61,15,0
```

Figura 3.6 - Exemplo do arquivo de entrada da Random Forest.

Como resultado do processamento dos dados de entrada obtém-se um lista com a classificação de cada pixel em Floresta (1) e Não-Floresta (0). Posteriormente, os pixels classificados são processados em um program desenvolvido na linguagem de programação C++ utilizando a biblioteca GDAL.

### 3.5 Processamento para a criação das Imagens Classificadas

A próxima etapa para a metodologia utilizada é a criação das imagens resultantes com a classificação dos pixels. Para a realização desse processamento, desenvolveu-se um software na linguagem Java utilizando biblioteca *imageio*. Os arquivos com a classificação dos pixels em Floresta e Não-Floresta de cada técnica utilizada são processados nesse software e gerada as imagens classificadas. Os resultados da classificação com as três técnicas são apresentados no Capítulo 4. A Figura 3.7 apresenta um exemplo de uma imagem classificada em Floresta (verde) e Não-Floresta (Vermelho) pela Rede Neural MLP.



Figura 3.7 - Exemplo de uma Imagem Classificada pela Rede Neural MLP em Floresta (verde) e Não-Floresta (Vermelho).

### 3.6 Validação das Imagens Classificadas

Para a validar os resultados obtidos calculou-se para as imagens classificadas o valores referentes a aplicação das técnicas de Entropia de Kullback Leiber e Matriz de Confusão. Nas próximas seções apresenta-se como as técnicas são aplicadas no projeto.

#### 3.6.1 Entropia ou Divergência de Kullback Leiber

Para a aplicação da Entropia ou Divergência de Kullback Leiber desenvolveu-se um software na linguagem C++. Como essa Divergência representa a distância entre duas funções discretas de probabilidade  $p$  e  $q$ , considerou-se para este projeto que,

$$\begin{aligned} p &= \text{número de pixels classificados como Floresta} \\ q &= \text{número de pixels classificados como Não-Floresta} \end{aligned}$$

e como probabilidades  $p_k$  e  $q_k$ ,

$p_k$  = probabilidade do píxel  $p_k$  ocorrer na imagem resultante da classificação das técnicas.

$q_k$  = probabilidade do píxel  $q_k$  ocorrer na imagem da Verdade PRODES.

A Entropia de Kullback Leiberé utilizada para a validação dos resultados da classificação da imagens.

O Capítulo 4, apresenta os resultados do cálculo da Entropia de KullBack Leiber para as imagens classificadas para as técnicas Rede MLP, Rede MLP Nebulosa e Random Forest.

### 3.6.2 Matriz de Confusão

Na aplicação da Matriz de Confusão desenvolveu-se um software na linguagem C++. Para o cálculo da Matriz de Confusão no programa, considerou-se para o cálculo de TP(True Positive) e TN(True Negative) os pixels classificados como Não-Floresta e para o cálculo de FP(False Positive) e FN(False Negative) os pixels classificados como Floresta. Os resultados do cálculo da Matriz de confusão para as imagens classificadas para cada Técnica (Rede MLP, Rede MLP Nebulosa e Random Forest) utilizada no projeto, é apresentado no Capítulo 4.

Tabela 3.4 - Matriz de Confusão

a	b	
TP	FP	a = NF
FN	TN	b = F

A Matriz de Confusão é utilizada para a validação dos resultados da classificação da imagens.

## 4 RESULTADOS OBTIDOS

Os resultados do projeto foram obtidos aplicando a metodologia desenvolvida no Capítulo 4 aos modelos de testes denominados de pequena escala com 37500 píxels, média escala com 500.000 píxels e grande escala com 8.341.210 píxels. Esses testes serão detalhados nas próximas seções.

### 4.1 Testes de Pequena Escala

Os primeiros resultados foram obtidos com a realização dos testes de pequena escala para cada técnica abordada. Esses testes foram realizados com 15 imagens MODIS do estado de Rondônia com tamanho de 50 x 50 píxels, totalizando 37.500 píxels. Os resultados estão apresentados nas Tabelas 4.1, 4.2 e 4.3. Por simples inspeção visual, em comparação com a verdade PRODES, e, mais rigorosamente, pelos resultados fornecidos pela entropia de Kullback-Leibler e pelas matrizes de confusão, observa-se que a técnica MLP teve um resultado satisfatório na classificação, enquanto a MLP Nebulosa apresentou um bom desempenho. Já a técnica de Floresta Aleatória apresentou o melhor resultado dentre as técnicas analisadas, desempenho este que melhora muito pouco com o aumento do número de árvores utilizados.

Tabela 4.1 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala


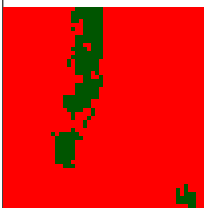



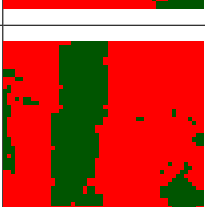

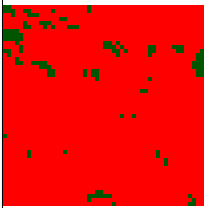

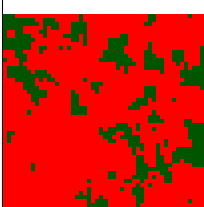
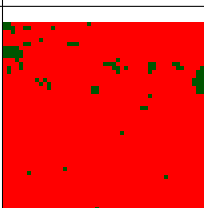
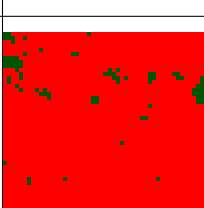
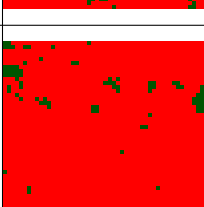

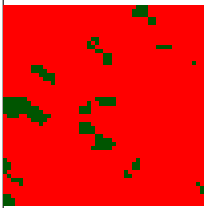
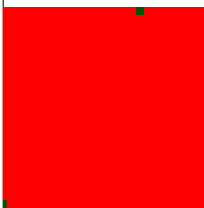

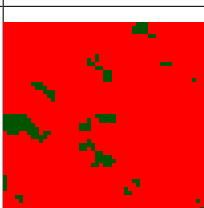

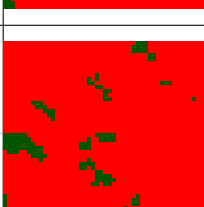
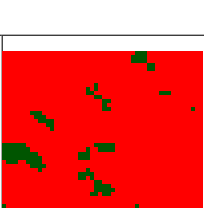
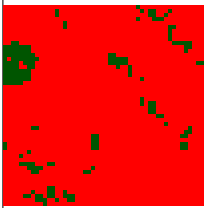
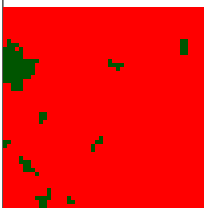
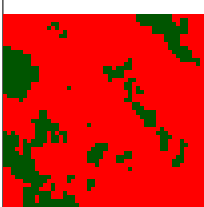
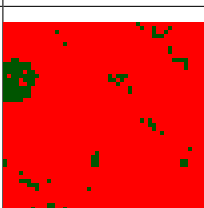
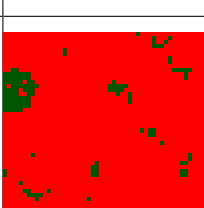
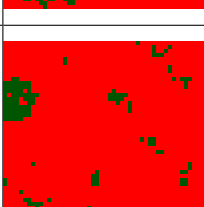
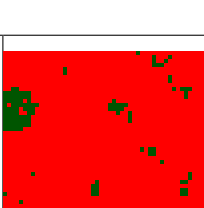
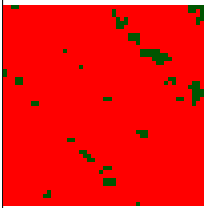

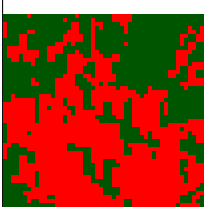
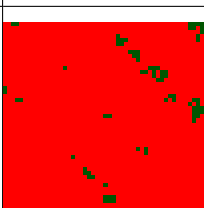
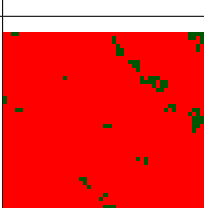
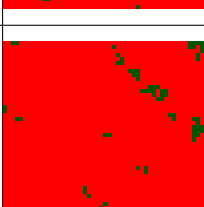
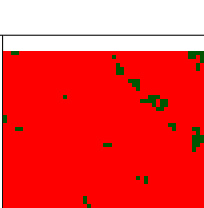
Número	Verdade Prodes	MLP	MLP Nebulosa	Random Forest			
				10 Árvores	100 Árvores	500 Árvores	1000 Árvores
1							
2							
3							
4							
5							

Tabela 4.2 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala

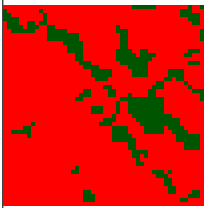
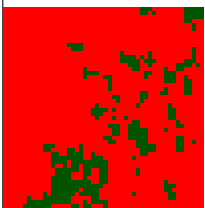
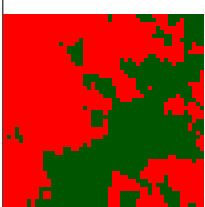
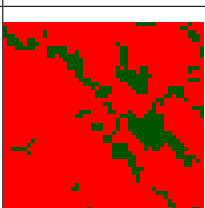
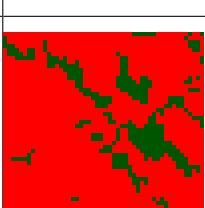
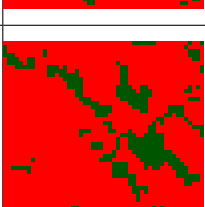
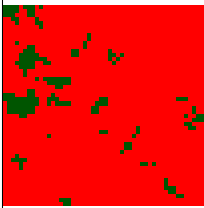
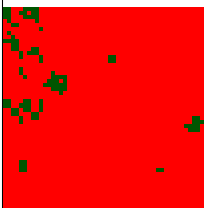
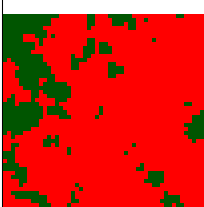
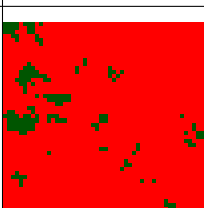
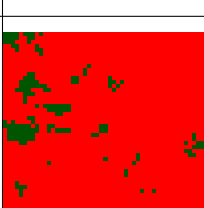
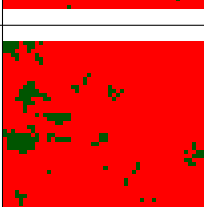
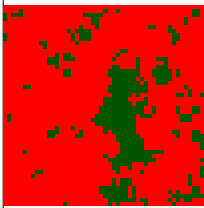
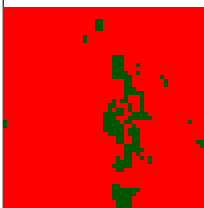
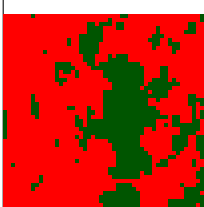
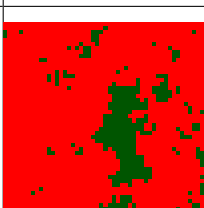
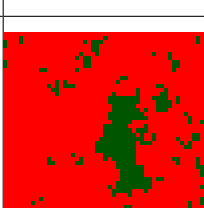
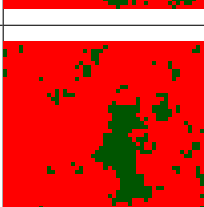
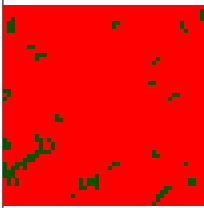

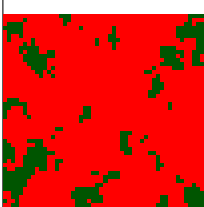
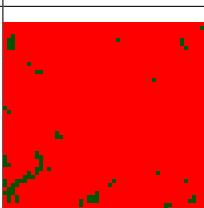
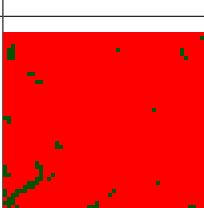
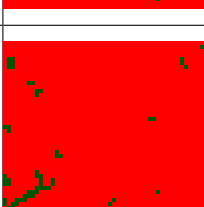
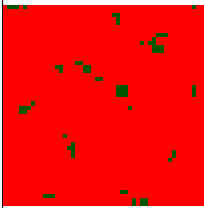
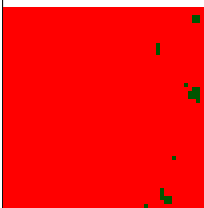
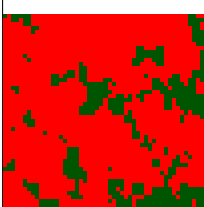
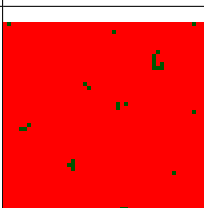
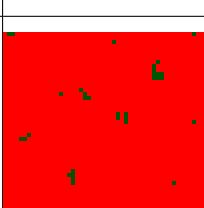
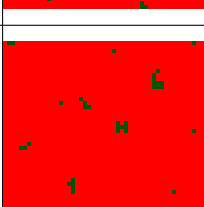
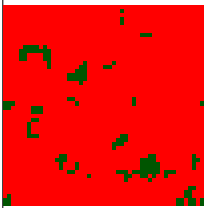
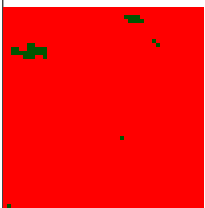
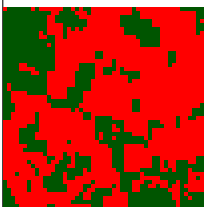
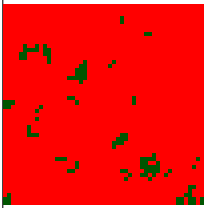
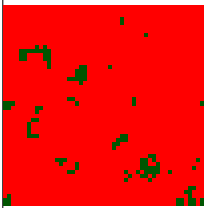
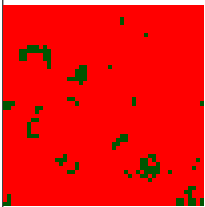
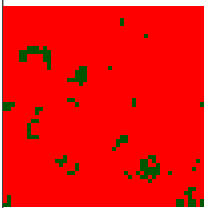

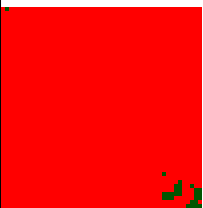
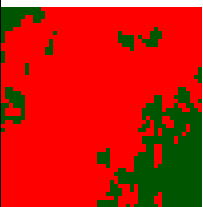





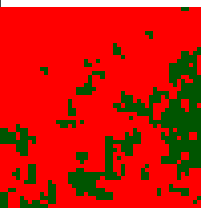
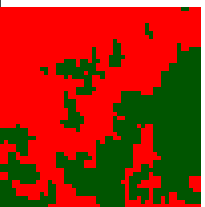
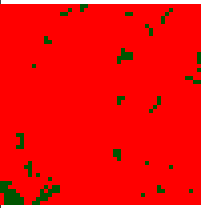



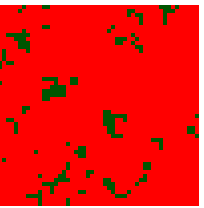
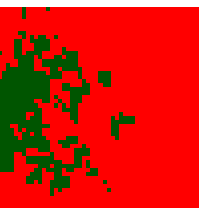
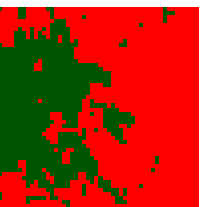
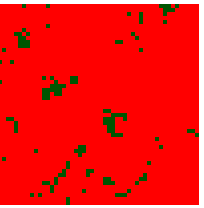
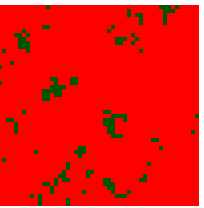
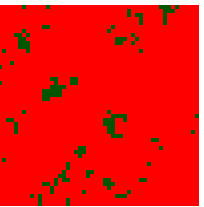
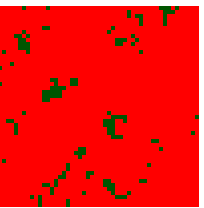
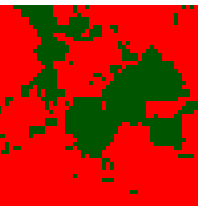
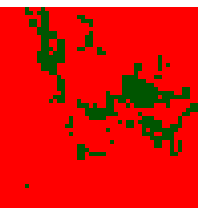
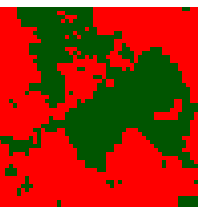
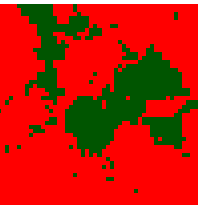
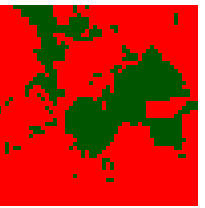
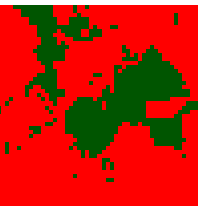
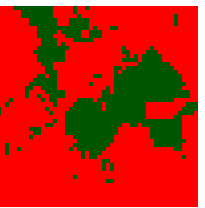







Número	Verdade Prodes	MLP	Random Forest			
			MLP Nebulosa 10 Árvores	100 Árvores	500 Árvores	1000 Árvores
6						
7						
8						
9						
10						

Tabela 4.3 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de pequena escala

Número	Verdade Prodes	MLP	MLP Nebulosa	Random Forest			
				10 Árvores	100 Árvores	500 Árvores	1000 Árvores
11							
							
12							
							
13							
							
14							
15							

## 4.2 Validação dos Resultados da Classificação com os Testes da Pequena Escala

A validação dos resultados obtidos na classificação das imagens foi realizada utilizando as técnicas MLP, MLP Nebulosa e Random Forest e aplicando a Entropia ou Divergência de Kullback-Leiber e Matriz de Confusão, como apresentado nas seções 4.4.1 e 4.4.2.

### 4.2.1 Entropia ou divergência de Kullback-Leiber

A Tabela 4.4 apresenta o cálculo dos valores de Entropia ou Divergência de Kullback-Leiber para as imagens das Tabelas 4.1, 4.2 e 4.3 na mesma ordem de apresentação das imagens. Os resultados obtidos com os cálculos da entropia de Kullback-Leiber confirmam o que pode-se observar visualmente nas Tabelas 4.1, 4.2 e 4.3 sobre os resultados alcançados com as técnicas MLP, MLP Nebulosa e Floresta Aleatória, nos quais a técnica de Floresta Aleatória apresentou os melhores resultados da classificação dos pixels das imagens.

Tabela 4.4 - Resultados do Cálculo da Entropia de Kullback-Leiber para os testes de Pequena Escala

Número	MLP	MLP Nebulosa	Random Forest			
			10 Árvores	100 Árvores	500 Árvores	1000 Árvores
1	0.003	0.644	0.507	0.519	0.519	0.519
2	0.772	0.420	0.849	0.845	0.846	0.846
3	0.03	0.228	0.050	0.047	0.048	0.045
4	0.457	0.321	0.017	0.029	0.027	0.024
5	0.231	0.628	0.495	0.507	0.479	0.487
6	0.0336	0.202	0.098	0.873	0.873	0.873
7	0.873	0.325	0.030	0.032	0.033	0.033
8	0.487	0.612	0.232	0.232	0.231	0.231
9	0.024	0.101	0.109	0.107	0.103	0.107
10	0.500	0.249	0.129	0.121	0.119	0.124
11	0.071	0.664	0.005	0.003	0.005	0.003
12	0.123	0.451	0.072	0.070	0.071	0.071
13	0.519	0.775	0.039	0.039	0.041	0.042
14	0.846	0.664	0.792	0.767	0.771	0.771
15	0.107	0.266	0.046	0.045	0.046	0.046



### 4.2.2 Matriz de Confusão

As Tabelas 4.5, 4.6 e 4.7 apresentam, na mesma ordem, em cada coluna/linha o cálculo dos valores da Matriz de Confusão para cada imagem apresentada nas Tabelas 4.1, 4.2 e 4.3. Para essas matrizes, a diagonal principal apresenta a quantidade de píxels classificados corretamente como  $a=F$  (F - Floresta) e  $b=NF$  (NF - Não-Floresta). Da mesma forma que o cálculo da entropia, observa-se para a Matriz de Confusão os melhores resultados para a técnica de Floresta Aleatória.

Tabela 4.5 - Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala

Número	MLP		MLP Nebulosa		Random Forest											
					10 Árvores				100 Árvores				500 Árvores			
1	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2131	211	1584	758	1710	632	1698	644	1699	643	1698	644	1699	643	1698	644
	131	27	96	62	102	56	102	56	101	57	101	57	101	57	101	57
2	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2368	1	1943	426	2369	0	2369	0	2369	0	2369	0	2369	0	2369	0
	124	7	26	105	38	93	35	96	36	95	36	95	36	95	36	95
3	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2332	0	2103	229	2332	0	2332	0	2332	0	2332	0	2332	0	2332	0
	162	6	22	146	21	147	18	150	19	149	17	151	17	149	17	151
4	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2281	29	1986	324	2310	0	2310	0	2310	0	2310	0	2310	0	2310	0
	100	90	25	165	44	146	34	156	35	155	38	152	35	155	38	152
5	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2393	1	1348	1046	2394	0	2394	0	2394	0	2394	0	2394	0	2394	0
	106	0	10	96	26	80	23	83	30	76	28	78	30	76	28	78

Tabela 4.6 - Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala

Número	MLP		MLP Nebulosa		Random Forest							
					10 Árvores		100 Árvores		500 Árvores		1000 Árvores	
6	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1818	264	1383	699	2082	0	2082	0	2082	0	282	0
	317	101	115	303	61	357	50	368	50	368	53	365
7	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2256	48	1973	331	2304	0	2304	0	2304	0	2304	0
	131	65	28	168	26	170	28	168	29	167	29	167
8	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2030	29	1722	337	2059	0	2059	0	2059	0	2059	0
	293	148	70	371	77	364	77	364	73	368	77	364
9	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2376	15	2053	338	2391	0	2391	0	2391	0	2391	0
	88	21	20	89	34	75	32	77	29	80	32	77
10	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	5	2	1423	351	1748	26	1745	29	1743	31	1746	28
	18	16	556	170	719	7	715	11	715	11	716	10

Tabela 4.7 - Resultados do Cálculo da Matriz de Confusão para os testes de Pequena Escala

Número	MLP		MLP Nebulosa		Random Forest							
					10 Árvores		100 Árvores		500 Árvores		1000 Árvores	
11	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2395	33	1575	853	2309	119	2306	122	2305	123	2306	122
	71	1	51	21	70	2	70	2	70	2	70	2
12	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2366	13	1912	467	2379	0	2379	0	2379	0	2379	0
	106	15	34	87	25	96	23	98	29	97	24	97
13	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1890	467	1472	885	2357	0	2357	0	2357	0	2357	0
	89	54	39	104	37	106	38	105	35	108	34	109
14	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1859	452	1565	746	2311	0	2311	0	2311	0	2311	0
	129	60	54	135	59	130	42	147	45	144	45	144
15	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1727	46	1451	322	1773	0	1773	0	1773	0	1773	0
	460	267	73	654	73	654	52	675	59	668	58	669

### 4.3 Testes de Média Escala

Os próximos resultados foram obtidos com a realização dos testes de média escala para cada técnica abordada. Esses testes foram realizados com 200 imagens MODIS do estado de Rondônia com resolução de 50 x 50 pixels, totalizando 500.000 pixels. Devido ao grande número de imagens, selecionou-se 15 imagens de 50 x 50 pixels para a apresentação dos testes realizados, como observado nas Tabelas 4.8, 4.9 e 4.10. Da mesma forma que nos testes de pequena escala, pode-se observar visualmente que, novamente, o melhor desempenho foi obtido pela técnica de Floresta Aleatória, mesmo para um número de árvores limitado a 10.

Tabela 4.8 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala

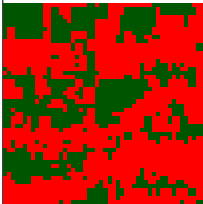
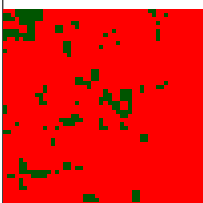
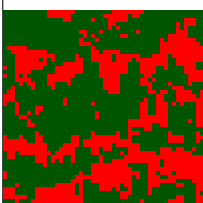
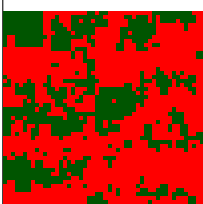
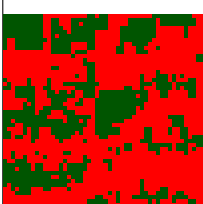
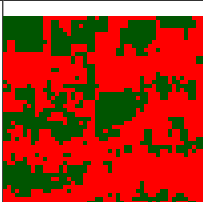
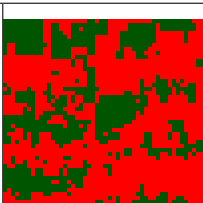
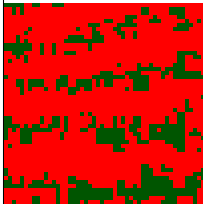
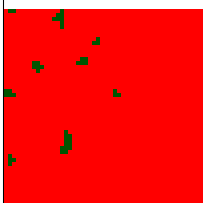
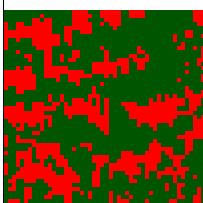
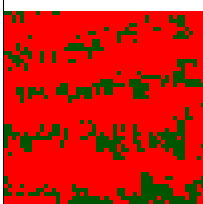
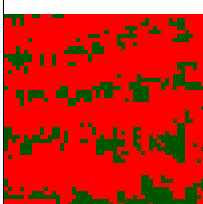
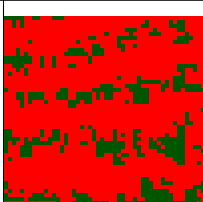
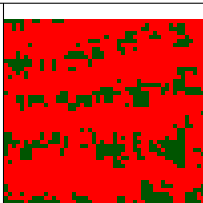
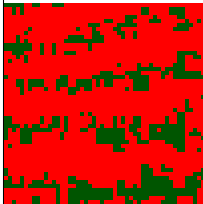
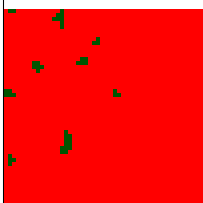
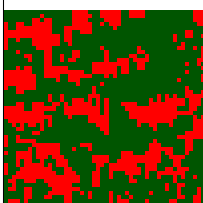
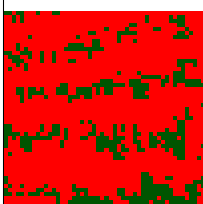
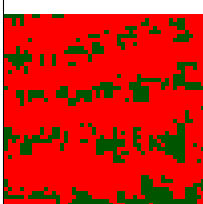
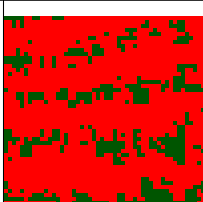
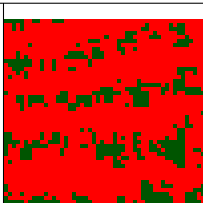
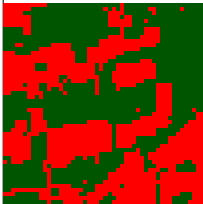
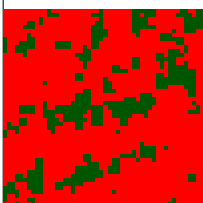
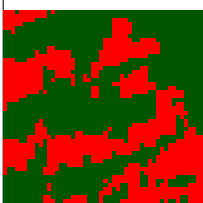
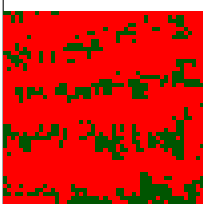
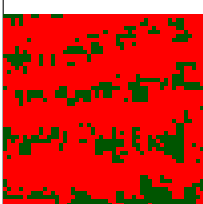
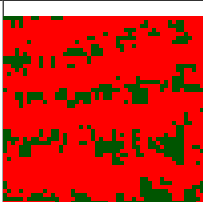
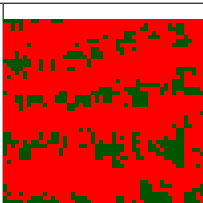
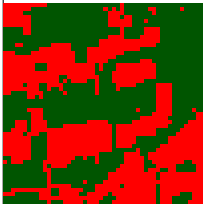
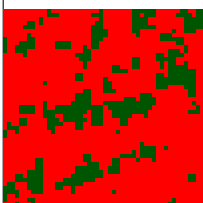
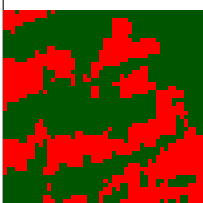
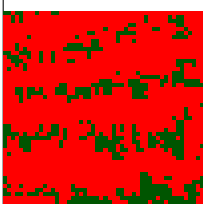
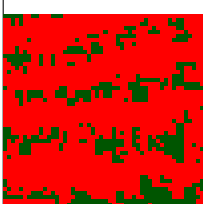
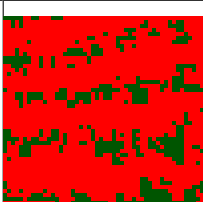
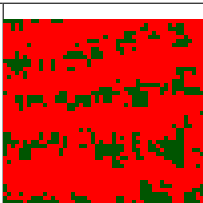
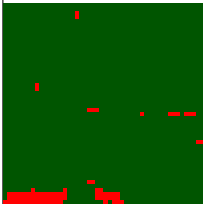
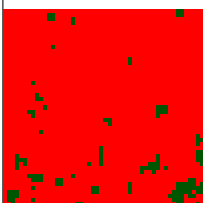
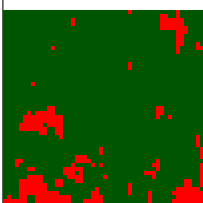
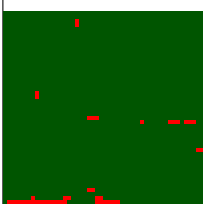
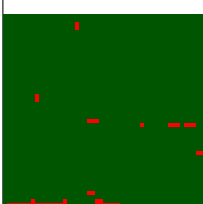
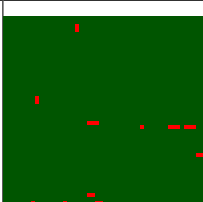
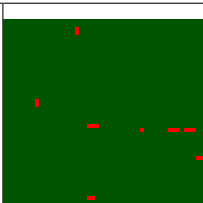
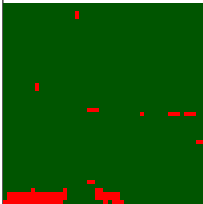
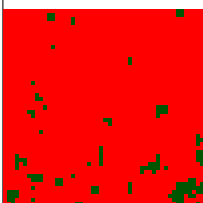
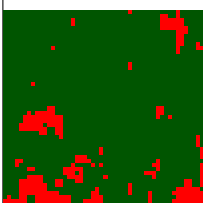
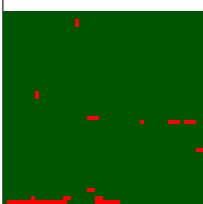
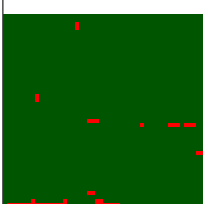
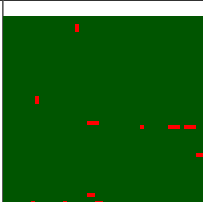
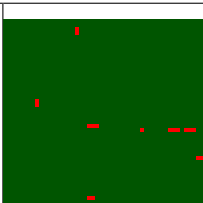

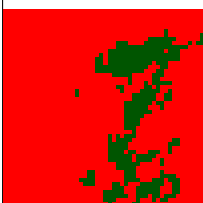
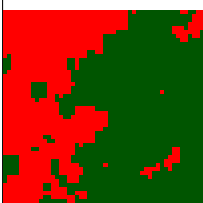



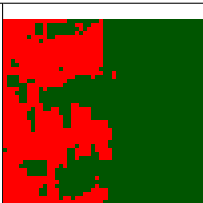

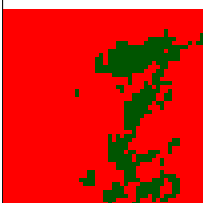
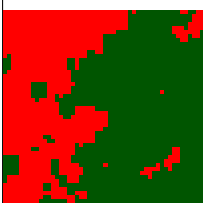



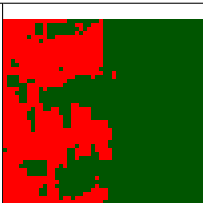

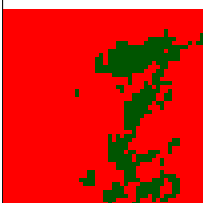
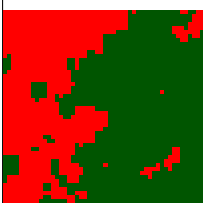



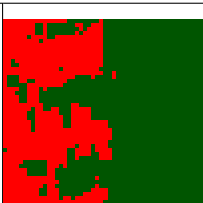
Número	Verdade Prodes	MLP	MLP Nebulosa	Random Forest			
				10 Árvores	100 Árvores	500 Árvores	1000 Árvores
1							
							
2							
							
3							
							
4							
							
5							
							

Tabela 4.9 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala


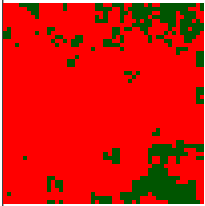
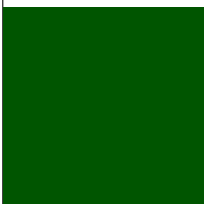

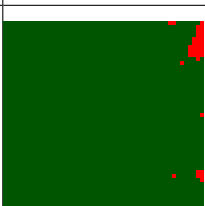
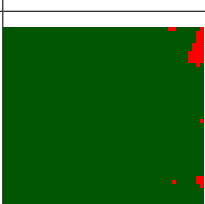

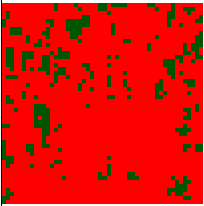


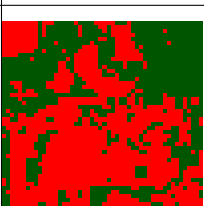

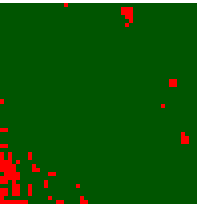
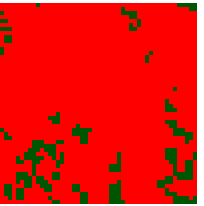
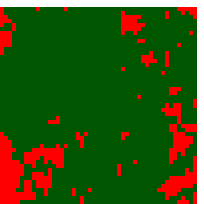

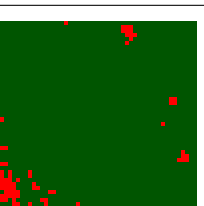
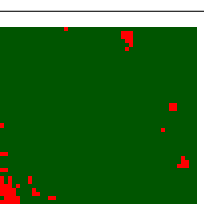

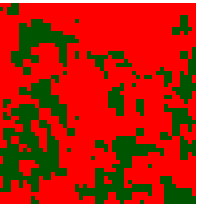
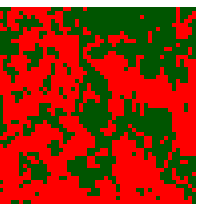

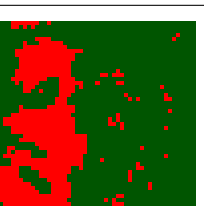
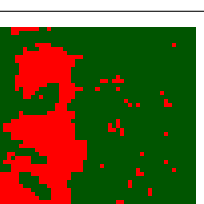
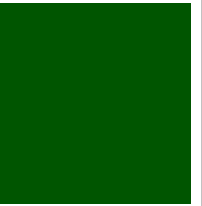
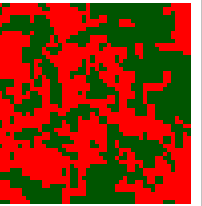
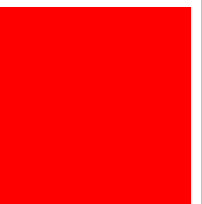
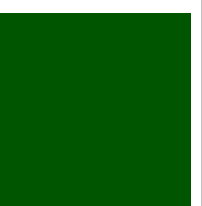
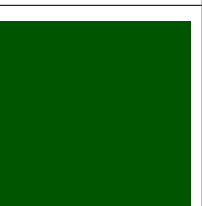
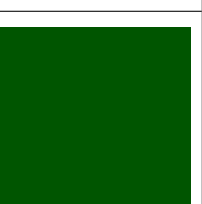







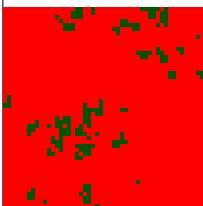
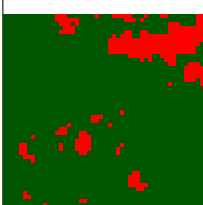
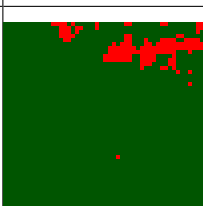
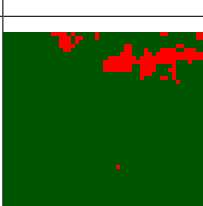

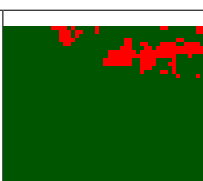

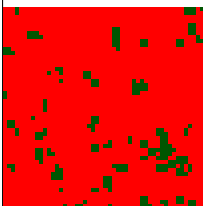
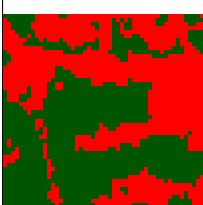


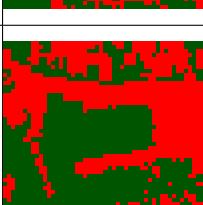

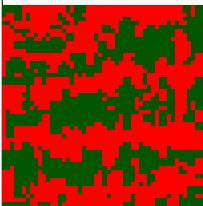
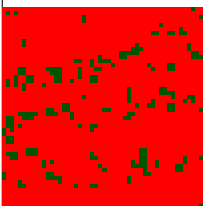
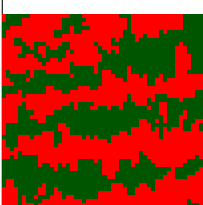
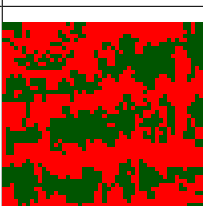
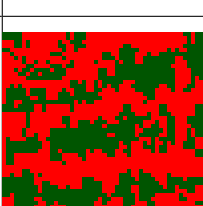

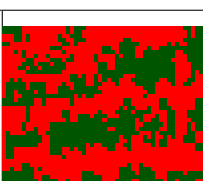
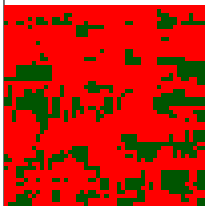

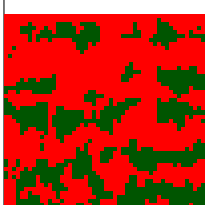
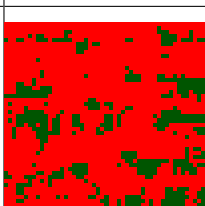
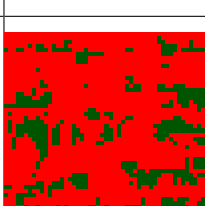

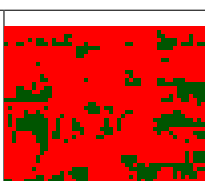
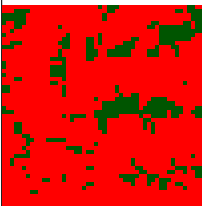

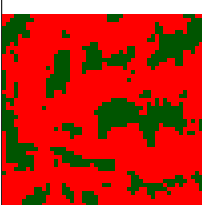
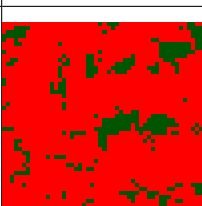
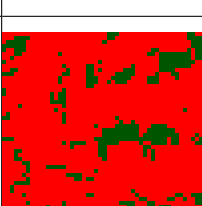

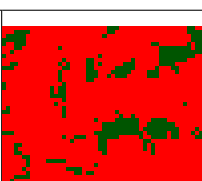
Número	Verdade Prodes	MLP	Random Forest			
			MLP Nebulosa 10 Árvores	100 Árvores	500 Árvores	1000 Árvores
6						
						
7						
						
8						
						
9						
10						

Tabela 4.10 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Média Escala

Número	para uma linha horizontal						
	Verdade Prodes	MLP	MLP Nebulosa	Random Forest			
				10 Árvores	100 Árvores	500 Árvores	1000 Árvores
11							
12							
13							
14							
15							



## 4.4 Validação dos Resultados da Classificação com os Testes da Média Escala

A validação dos resultados obtidos na classificação das imagens utilizando as técnicas MLP, MLP Nebulosa e Random Forest foi realizada aplicando a Entropia ou Divergência de Kullback-Leiber e Matriz de Confusão, como apresentado nas seções 4.4.1 e 4.4.2.

### 4.4.1 Entropia ou divergência de Kullback-Leiber

A Tabela 4.11 apresenta o cálculo dos valores de Entropia ou Divergência de Kullback-Leiber para as imagens das Tabelas 4.8, 4.9 e 4.10 na mesma ordem de apresentação das imagens. Os resultados obtidos com os cálculos da entropia de Kullback-Leiber confirmam o que pode-se observar visualmente nas Tabelas 4.8, 4.9 e 4.10 sobre os resultados alcançados com as técnicas MLP, MLP Nebulosa e Floresta Aleatória, nos quais a técnica de Floresta Aleatória apresentou os melhores resultados da classificação das imagens.

Tabela 4.11 - Resultados do Cálculo da Entropia de Kullback-Leiber para os testes de Média Escala

Número	MLP	MLP Nebulosa	Random Forest			
			10 Árvores	100 Árvores	500 Árvores	1000 Árvores
1	0.154	0.0100	0.251	0.284	0.283	0.283
2	0.764	0.993	0.426	0.461	0.466	0.464
3	0.000	0.856	0.199	0.236	0.251	0.259
4	0.183	0.309	0.284	0.289	0.289	0.289
5	0.395	0.182	0.192	0.196	0.196	0.197
6	0.507	0.000	0.153	0.150	0.147	0.147
7	0.067	0.085	0.019	0.019	0.019	0.019
8	0.713	0.395	0.449	0.449	0.460	0.449
9	0.670	0.651	0.447	0.447	0.467	0.467
10	0.880	0.500	0.437	0.437	0.437	0.437
11	0.500	0.518	0.461	0.446	0.446	0.447
12	0.853	0.947	0.398	0.363	0.368	0.362
13	0.940	0.145	0.046	0.043	0.443	0.441
14	0.219	0.066	0.018	0.021	0.021	0.021
15	0.647	0.442	0.164	0.174	0.163	0.163

#### 4.4.2 Matriz de Confusão

As Tabelas 4.12, 4.13 e 4.14 apresentam, na mesma ordem, em cada coluna/linha o cálculo dos valores da Matriz de Confusão para cada imagem das Tabelas 4.8, 4.9 e 4.10. Para essas matrizes, a diagonal principal apresenta a quantidade de pixels classificados corretamente como a=F (F - Floresta) e b=NF (NF - Não-Floresta). Da mesma forma que o cálculo da entropia, observa-se para a Matriz de Confusão os melhores resultados para a técnica de Floresta Aleatória.

Tabela 4.12 - Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala

Número	MLP		MLP Nebulosa		Random Forest											
					10 Árvores				100 Árvores				500 Árvores			
1	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2194	200	849	1275	1557	775	1527	805	1528	804	1525	807	1525	807	1525	807
	100	6	121	255	93	75	90	78	90	78	90	78	90	78	90	78
2	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1753	21	735	1334	1910	518	1875	553	1870	558	1872	556	1870	558	1872	556
	699	27	157	274	64	8	61	11	61	11	61	11	61	11	61	11
3	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1831	501	408	818	912	861	897	876	892	881	891	882	892	881	891	882
	129	39	597	677	297	430	297	430	296	431	294	433	296	431	294	433
4	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2006	107	58	25	72	2238	70	2240	70	2240	70	2240	70	2240	70	2240
	354	33	289	2128	14	176	14	176	14	176	14	176	14	176	14	176
5	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1640	386	956	1434	910	1484	901	1493	899	1495	899	1495	899	1495	899	1495
	461	13	52	58	12	94	12	94	12	94	11	95	12	94	11	95

Tabela 4.13 - Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala

Número	MLP		MLP Nebulosa		Random Forest							
					10 Árvores		100 Árvores		500 Árvores		1000 Árvores	
6	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1931	379	0	2457	28	2054	27	2055	26	2056	26	2056
	164	26	0	43	9	409	9	409	9	409	9	409
7	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1808	274	872	1388	1310	994	1287	1017	1276	1028	1279	1025
	361	57	72	168	118	78	112	84	114	82	112	84
8	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2057	247	379	1991	119	1940	110	1949	109	1950	109	1950
	185	11	28	102	3	438	3	438	3	438	3	438
9	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1463	596	1381	996	628	1763	615	1776	608	1783	609	1782
	285	156	63	60	40	69	40	69	39	70	39	70
10	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	1262	1129	1557	0	0	2369	0	2369	0	2369	0	2369
	67	42	943	0	0	131	0	131	0	131	0	131

Tabela 4.14 - Resultados do Cálculo da Matriz de Confusão para os testes de Média Escala

Número	MLP		MLP Nebulosa		Random Forest							
					10 Árvores		100 Árvores		500 Árvores		1000 Árvores	
11	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2204	165	266	2005	123	1651	115	1659	116	1658	117	1657
	119	12	19	210	20	706	21	705	20	706	20	706
12	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2212	216	364	293	1099	1243	1078	1264	1076	1266	1078	1264
	58	14	837	1006	54	104	52	106	52	106	52	106
13	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2144	198	860	667	1419	960	1402	977	1405	974	1404	975
	140	18	526	447	77	44	78	43	78	43	78	43
14	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2241	138	854	300	1868	489	1849	508	1851	506	1851	506
	114	7	762	584	109	34	110	33	110	33	110	33
15	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F	a = NF	b = F
	2290	67	1268	472	1995	316	1975	336	1980	331	1979	332
	135	8	570	190	170	19	168	21	168	21	168	21

## 4.5 Testes de Grande Escala

Os testes de Grande Escala foram realizados utilizando imagens do estado de Rondônia para os anos de 2010 e 2011. Essas imagens possuem uma dimensão de 3202x2605 pixels, totalizando um número de 8.341.210 pixels cada imagem. Nesse teste utilizou-se apenas a técnica Random Forest, devido aos seus ótimos resultados na classificação das imagens nos testes de pequena e média escala. Os testes foram realizados para uma RF com 10 árvores (Figura 4.15), 100 árvores (Figura 4.16), 500 árvores (Figura 4.17) e 1000 árvores (Figura 4.18).

Tabela 4.15 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala

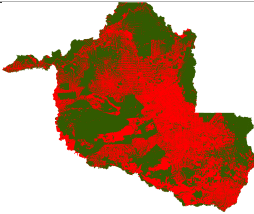
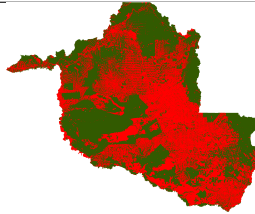
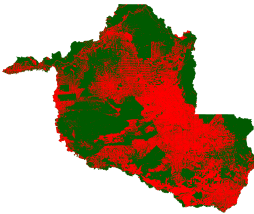
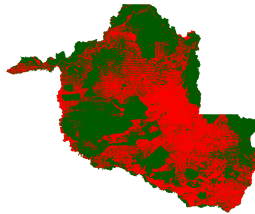
	2010	2011
Verdade Prodes		
10 árvores		

Tabela 4.16 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala

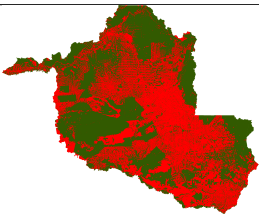
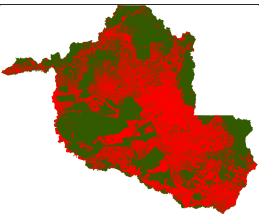
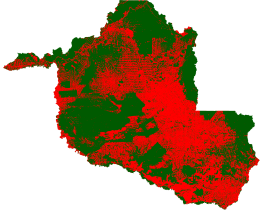
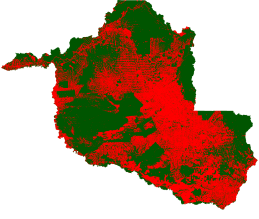
	2010	2011
Verdade Prodes		
100 árvores		

Tabela 4.17 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala

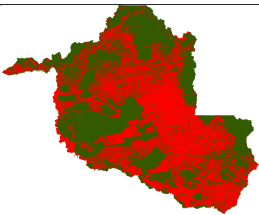
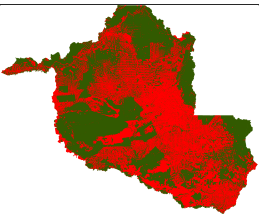
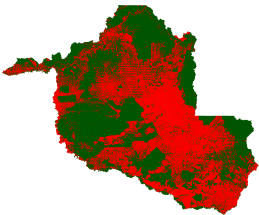
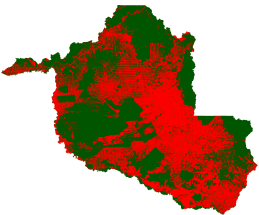
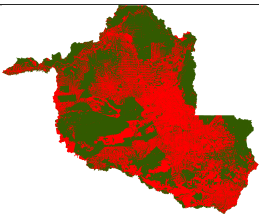
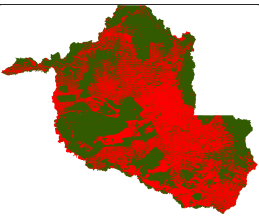
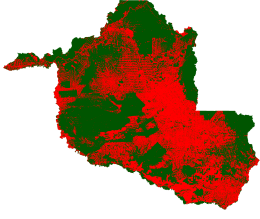
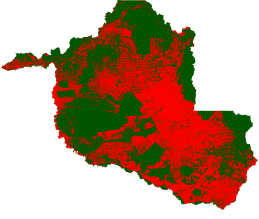
	2010	2011
Verdade Prodes		
500 árvores		



Tabela 4.18 - Resultados da Classificação das Imagens em Floresta e Não-Floresta com testes de Grande Escala

	2010	2011
Verdade Prodes		
1000 árvores		

#### **4.6 Validação da Área classificada como Floresta na imagem resultante do estado de Rondônia**

Para a validação das imagens classificadas de Rondônia, para os anos de 2010 e 2011, foi realizada uma comparação entre a área classificada como floresta (verde) na imagem resultante do estado de Rondônia e área de classe Floresta (Verde) da verdade PRODES. Como as imagens classificadas possuem uma relocação espacial diferente das imagens PRODES, precisou-se determinar um fator de correção para transformação do número de pixels classificados em uma determinada classe (F ou NF) para um valor em hectares, passível de ser comparado com o valor fornecido pelo projeto PRODES. Este fator de conversão foi obtido uma única vez via calibração com a verdade PRODES, e utilizados em todas as outras simulações. O valor encontrado para esse fator foi 0.41, como pode ser observado nas Tabelas [4.19](#) e [4.20](#).

• 2010

Tabela 4.19 - Resultados da Validação da Área classificada como Floresta na imagem resultante do ano de 2010 do estado de Rondônia

PRODES	PROGRAMA (Limiar)									
	0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49
122315,7	121824,75	12124,49	120735,81	120291,19	119779,69	119258,25	118756,31	118264,0	117692,81	116895,0

• 2011

Tabela 4.20 - Resultados da Validação da Área classificada como Floresta na imagem resultante do ano de 2011 do estado de Rondônia

PRODES	PROGRAMA (Limiar)									
	0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49
122236,3	122829,06	122352,06	121976,69	121582,25	121106,75	120570,63	120001,69	119424,31	118715,06	117872,06

#### 4.6.0.1 Validação do Limiar encontrado para o cálculo da área de Floresta

A validação do Limiar 0,41 encontrado para o cálculo da área de Floresta das imagens do estado de Rondônia foi realizada calculando a área de Floresta para os municípios de Rondônia. Os testes foram aplicados em sete municípios e calculado os erros entre as áreas verificadas. A Figura 4.1 apresenta a sobreposição dos municípios analisados no estado de Rondônia com os valores das áreas e erros.

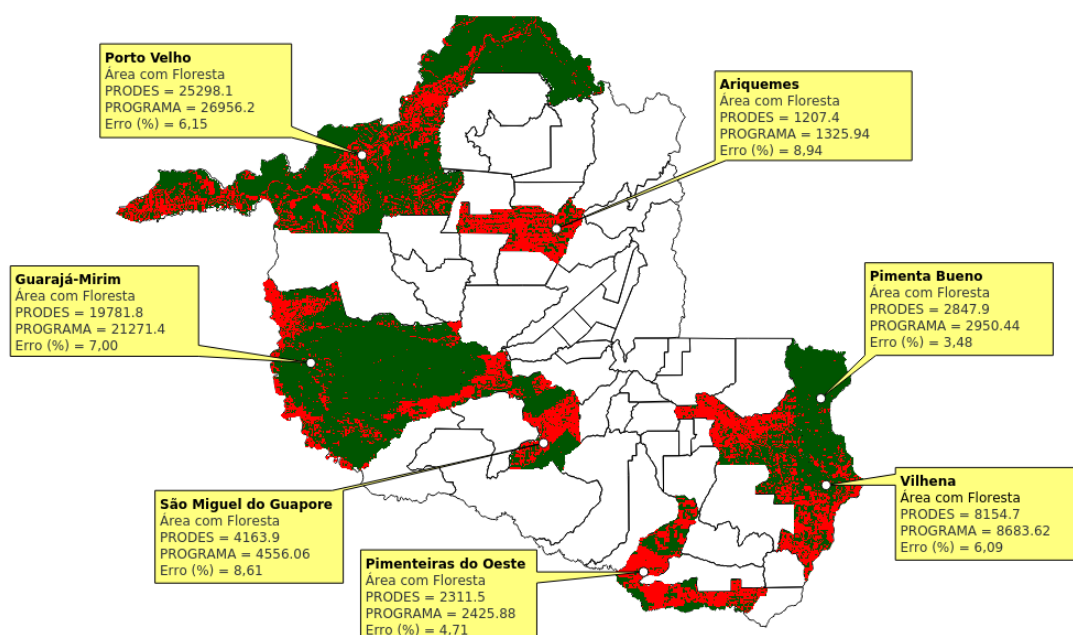


Figura 4.1 - Área de Floresta dos Municípios PRODES e das Imagens Classificadas e erro do cálculo das áreas.

#### 4.7 Cálculo do Grau de Incerteza das imagens classificadas para a diminuição dos trabalhos dos voluntários

Para a redução do número de pixels a serem enviados para a análise dos voluntários foi gerado uma imagem em escala de cinza utilizando as informações da porcentagem de acerto das árvores da RF na determinação da classe vencedora, Floresta e Não-Floresta. Assim, um pixel onde todas as árvores da RF convergiram para uma mesma classe tem uma incerteza nula (1 ou branco) e, logo, não deverá ser submetido a um voluntário para reanálise. Já um pixel cuja classificação dividiu ao meio as árvores da RF possui máxima incerteza (0 ou preto) e deverá ser necessariamente inspecionado visualmente por um ou mais voluntários. Note que a informação sobre

a porcentagem de árvores que convergiram para uma dada classe (F, por exemplo) é fornecida automaticamente pelo pacote *Weka*. De posse desta informação, é possível calcular a incerteza associada á classificação de um dado píxel:

$$I = \frac{|P - 0.5|}{0.5} \quad (4.1)$$

onde,  $I$  = valor da incerteza do píxel classificado

$P$  = porcentagem de árvores na classe vencedora do píxel em floresta

Um exemplo dos dados gerado pelo *Weka* é mostrado na Tabela 4.21.

Tabela 4.21 - Dados de porcentagem de árvores que convergiram para uma dada classe fornecida pelo pacote *Weka*

F	NF
0.1	0.9
0.3	0.7
1.0	0.0
1.0	0.0
1.0	0.0
0.83	0.17
0.51	0.49
0.71	0.29
0.71	0.29
0.63	0.37
1.0	0.0
1.0	0.0
1.0	0.0

Para a identificação da quantidade ideal de píxels a serem enviados para os voluntários foi realizado validações com diversos limiares de incerteza aplicados aos resultados da eq. (1). Para um dado limiar, são enviados para inspeção por voluntários somente os píxels cuja classificação possui uma incerteza igual ou superior a este valor. Os resultados são apresentados para os testes de pequena escala nas Tabelas 4.22, 4.23, 4.24 e 4.25, para os testes de média escala nas Tabelas 4.26, 4.27, 4.28 e 4.29 e para os testes de grande escala nas Tabelas 4.30, 4.31, 4.32 e 4.33 para o ano de 2010 e nas Tabelas 4.34, 4.35, 4.36 e 4.37 para o ano de 2011. Também se calculou o coeficiente de redução do número de tarefas a serem enviadas aos voluntários, como calculado por:

$$CR = \left(1 - \frac{PEV}{PTI}\right) * 100 \quad (4.2)$$

onde,

$CR$  =coeficiente de redução.

$PEV$  =número de píxels enviados aos voluntários.

$PTI$  =número de píxels total da imagem classificada.

Estes resultados mostram que o uso de um limiar de incerteza permite reduzir significativamente a carga de trabalho dos voluntários (entre 45% a mais de 98% de redução). Esta redução depende pouco do número de árvores utilizado na RF, e muito da complexidade e do tamanho da imagem original. Imagens muito complexas, com um número grande de áreas desmatadas em meio a áreas com floresta, são mais difíceis de classificar, devido à baixa resolução espacial da imagens MODIS, e, portanto, exigem uma maior carga de trabalho dos voluntários.

Tabela 4.22 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 10 árvores

- Pequena Escala: 37.500 pixels
- Número de árvores: 10

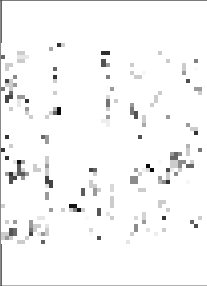

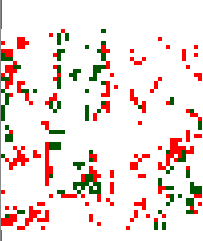
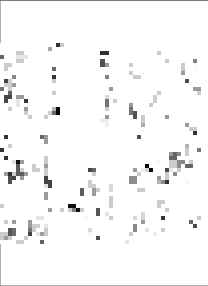

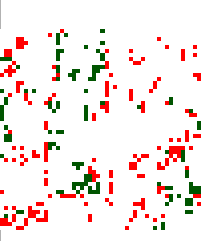
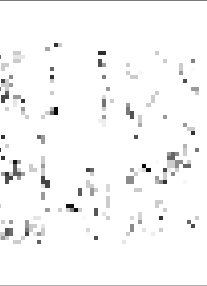

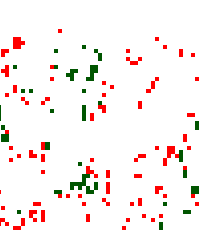
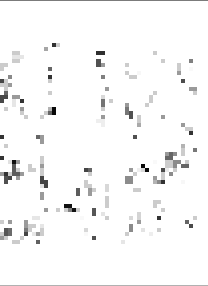

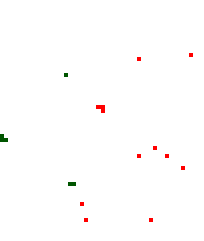
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				83.84%
0.40				86.76%
0.70				91.80%
0.90				99.16%

Tabela 4.23 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 100 árvores

- Pequena Escala: 37.500 pixels
- Número de árvores: 100

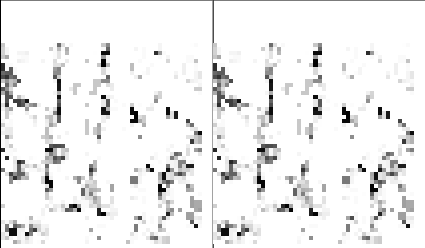
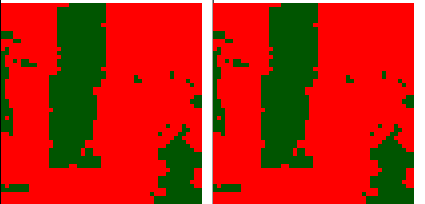
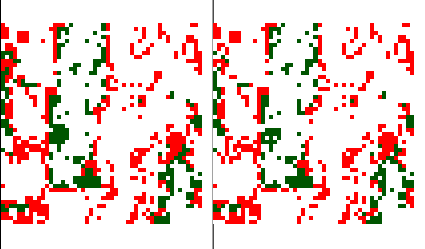
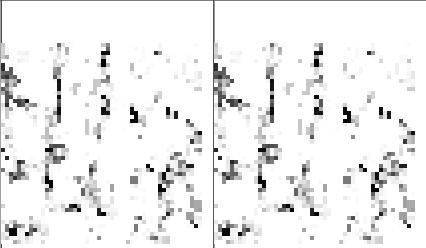
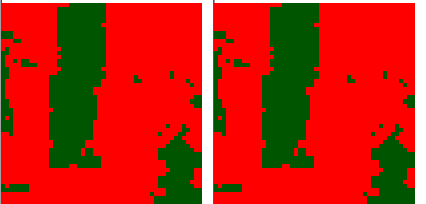
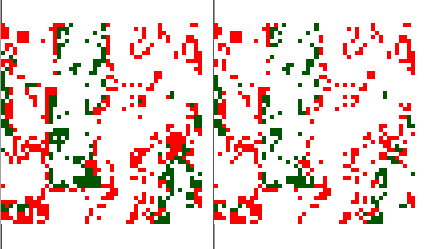
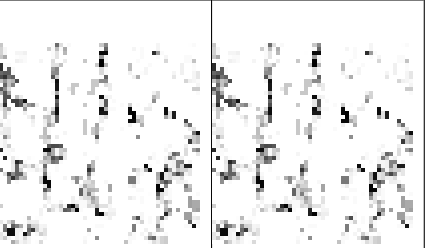
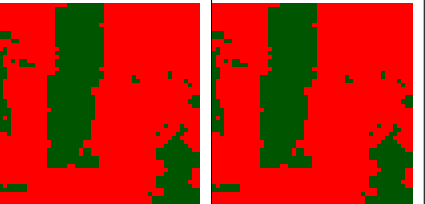
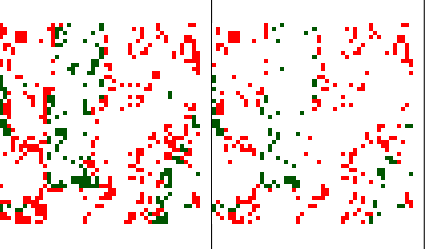
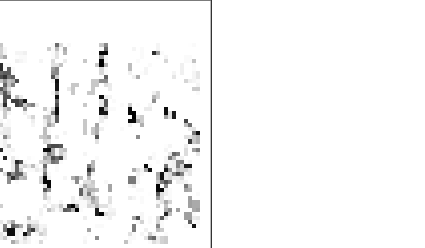
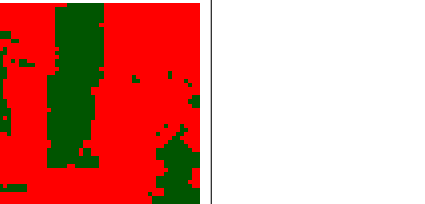
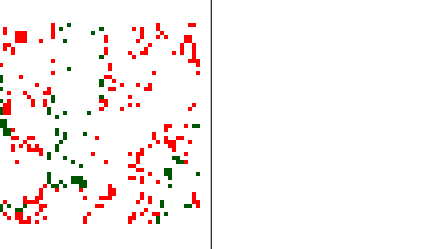
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				71.92%
0.40				74.08%
0.70				79.60%
0.90				87.64%



Tabela 4.24 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 500 árvores

- Pequena Escala: 37.500 pixels
- Número de árvores: 500



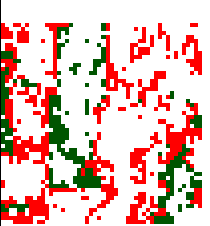


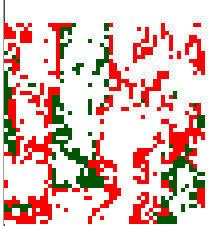


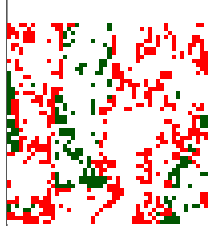


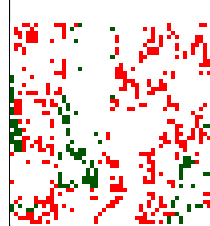
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				65.04%
0.40				67.2%
0.70				72.48%
0.90				81.57%

Tabela 4.25 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de pequena escala com 100 árvores

- Pequena Escala: 37.500 pixels
- Número de árvores: 1000



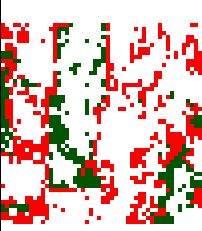


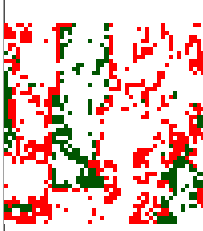


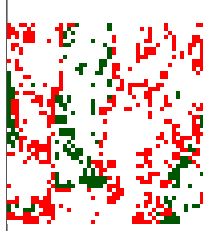


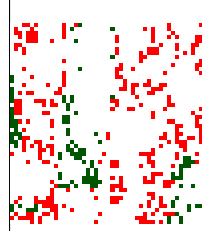
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				66.08%
0.40				68.16%
0.70				73.52%
0.90				81.48%

Tabela 4.26 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 10 árvores

- Média Escala: 500.000 pixels
- Número de árvores: 10


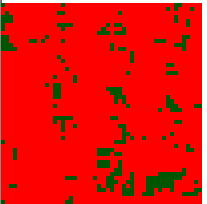
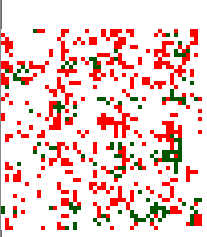

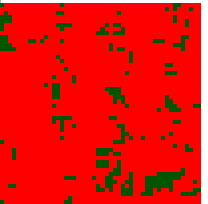
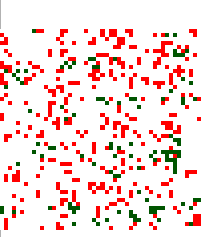

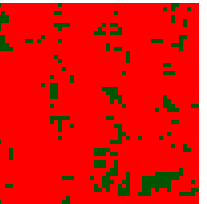
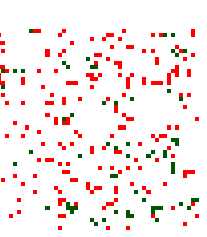

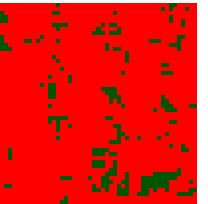
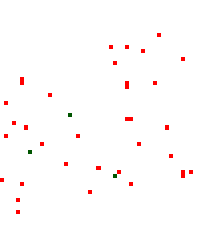
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				75.4%
0.40				81.46%
0.70				89.64%
0.90				98.48%

Tabela 4.27 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 100 árvores

- Média Escala: 500.000 pixels
- Número de árvores: 100


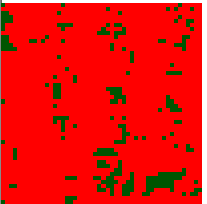
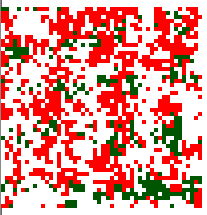

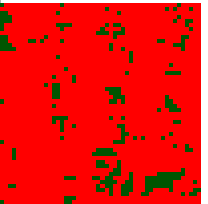
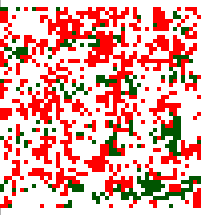

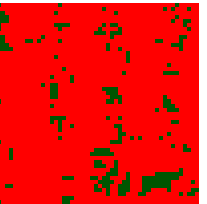
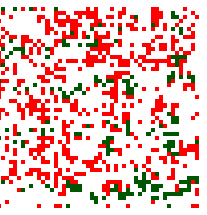

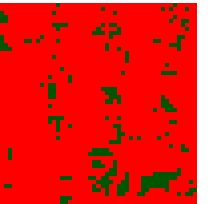
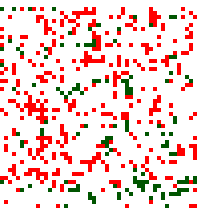
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				52.56%
0.40				59.32%
0.70				68.60%
0.90				78.84%

Tabela 4.28 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 500 árvores

- Média Escala: 500.000 pixels
- Número de árvores: 500


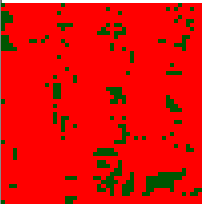
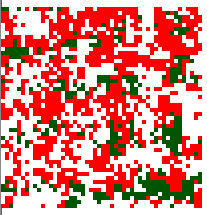

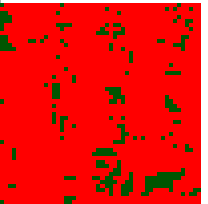
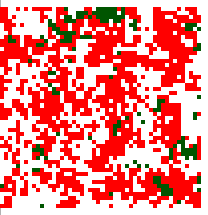

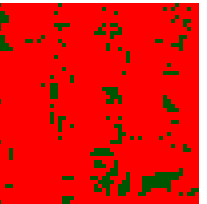
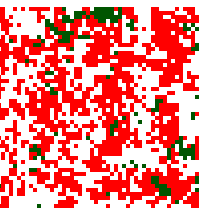

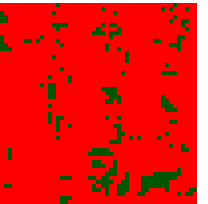
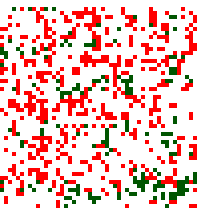
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				45.08%
0.40				52.16%
0.70				60.80%
0.90				71.40%

Tabela 4.29 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de média escala com 1000 árvores

- Média Escala: 500.000 pixels
- Número de árvores: 1000


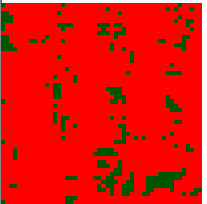
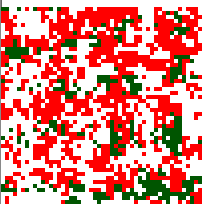

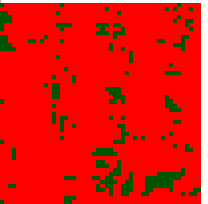
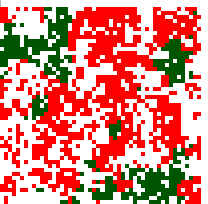

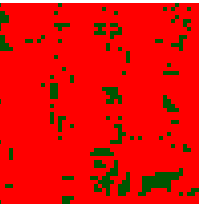
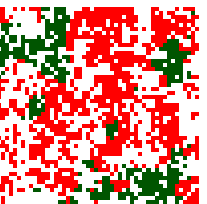

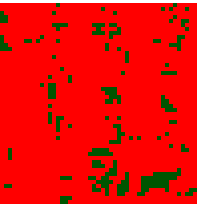
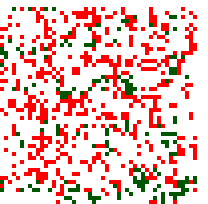
Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				45.88%
0.40				53.64%
0.70				61.68%
0.90				71.72%

Tabela 4.30 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2010 com 10 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 10

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				69.84%
0.40				72.97%
0.70				78.62%
0.90				86.54%

Tabela 4.31 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rôndônia de 2010 com 100 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 100

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				74.99%
0.40				77.82%
0.70				83.19%
0.90				90.38%



Tabela 4.32 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondonia de 2010 com 500 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 500

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				69.61%
0.40				72.72%
0.70				78.40%
0.90				86.31%

Tabela 4.33 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondonia de 2010 com 1000 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 1000

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				78,24%
0.40				81.24%
0.70				86.74%
0.90				94.31%

Tabela 4.34 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 10 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 10

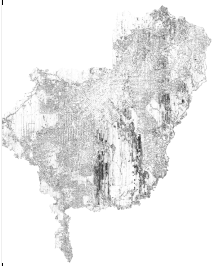
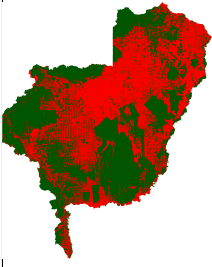

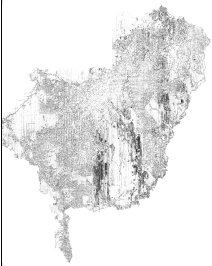
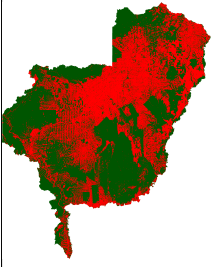
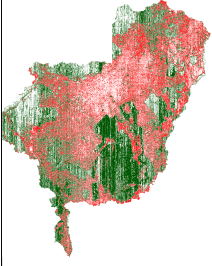
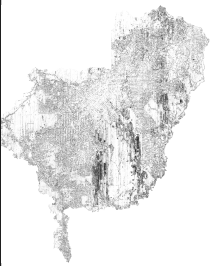
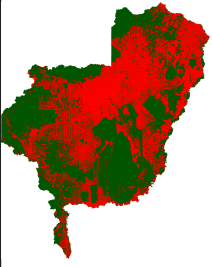

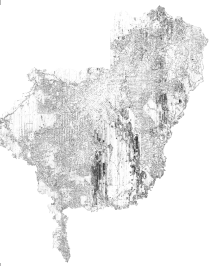
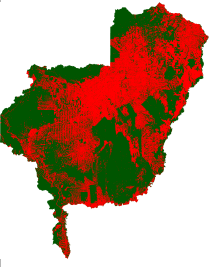

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				69.84%
0.40				72.97%
0.70				78.92%
0.90				86.54%

Tabela 4.35 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 100 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 100

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				71.93%
0.40				74.90%
0.70				80.73%
0.90				88.72%

Tabela 4.36 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondonia de 2011 com 500 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 500

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coefficiente de Redução
0.20				72.20%
0.40				75.17%
0.70				80.38%
0.90				87.5134%

Tabela 4.37 - Mapa de incerteza para a redução do número de tarefas enviadas aos voluntários para os testes de grande escala da imagem do estado de Rondônia de 2011 com 1000 árvores

- Grande Escala: 8.341.210 pixels
- Número de árvores: 1000

Limiar de Incerteza	Mapa de Incerteza	Imagem Classificada	Imagem Enviada	Coeficiente de Redução
0.20				82.57%
				85.47%
				90.43%
0.40				
0.70				



#### 4.8 Estimativa do Tempo de Processamento da Random Forest

Considerando a grande área de florestas a serem inspecionados, é importante avaliar como varia o tempo de processamento do classificador RF com o número de pixels e com o número de árvores. A estimativa de tempo de processamento para a técnica de Random Forest foi calculada em segundos para os testes de pequena (37.500 pixels), média (500.000) e grande escala (8.341.210 pixels) considerando a quantidade de 10, 100, 500 e 1000 árvores, como apresentado na Tabela 4.38.

Tabela 4.38 - Tempo de processamento em segundos para a técnica de Random Forest para o número de 10, 100, 500 e 1000 árvores.

	Número de Árvores			
Teste (pixels)	10	100	500	1000
37500	5.33	15.36	70.39	151.64
50000	29.40	231.85	1156.21	2533.95
8341210	432.79	3075.63	16130.60	41451.31

As Figuras 4.2 e 4.3 apresentam os gráficos Tempo x Número de pixels e Tempo x Número de Árvores, respectivamente. Observa-se que nos dois casos o tempo de computação varia linearmente (curva preta tracejada), o que indica um bom desempenho computacional para um classificador a ser utilizado em imagens potencialmente muito grandes. Assim, considerando todos os resultados obtidos nesta dissertação, o classificador RF com 10 árvores é provavelmente a melhor escolha para o projeto *ForestWatchers*.

### Tempo de Processamento da Random Forest (Floresta Aleatória)

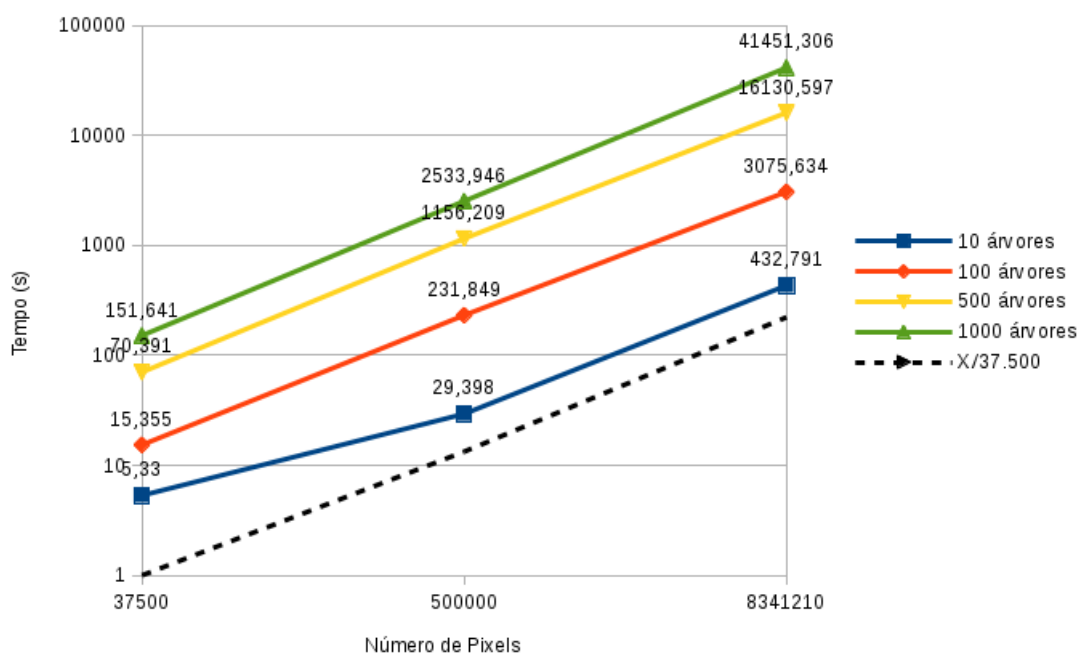


Figura 4.2 - Gráfico da Estimativa de tempo considerando Tempo x Número de pixels.

### Tempo de Processamento da Random Forest (Floresta Aleatória)

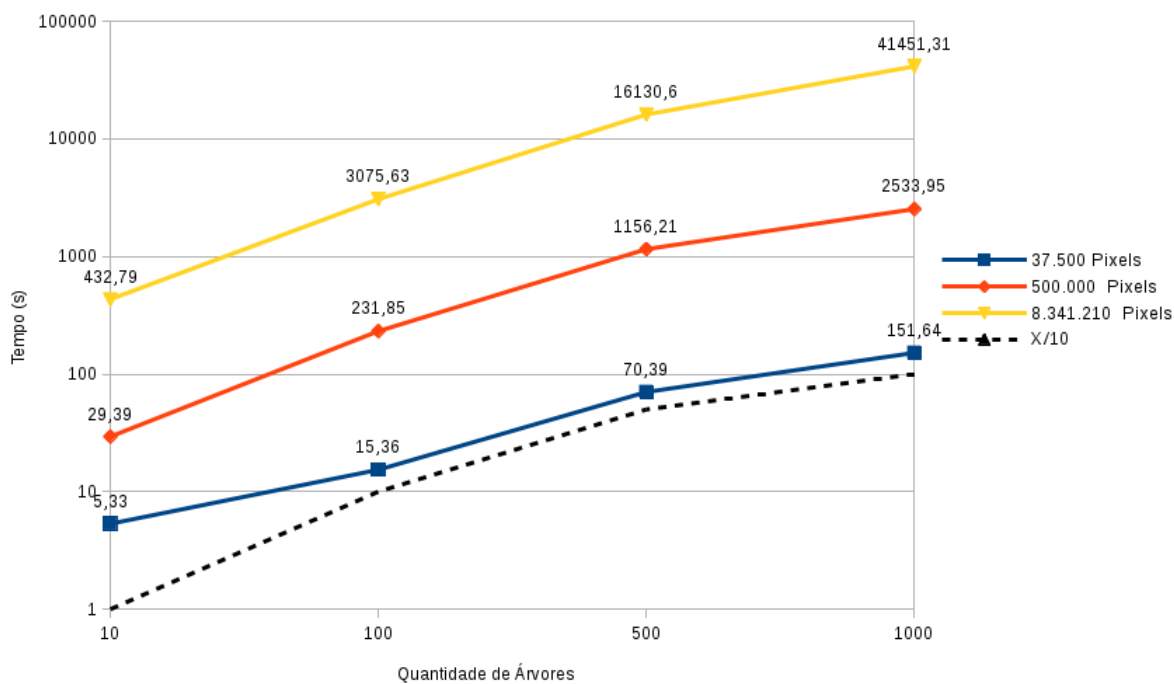


Figura 4.3 - Gráfico da Estimativa de tempo considerando Tempo x Número de Árvores.





## 5 CONCLUSÕES

As florestas tropicais são de grande importância para o planeta devido aos inúmeros serviços ambientais por elas fornecidos. Nas últimas décadas no Brasil, foram desenvolvidas no políticas públicas para a proteção destes valiosos ecossistemas. Essas políticas, apesar dos seus bons resultados nos últimos anos, possuem um custo operacional muito elevado, dificultando a sua utilização em países menos desenvolvidos.

Nesse contexto, foi desenvolvido um projeto de ciência cidadã, chamado *ForestWatchers*, que possibilita que voluntários leigos ajudem a monitorar o desmatamento das florestas tropicais por meio de uma interface Web. Nesta dissertação foram analisadas as técnicas de classificação por Rede Neural MLP, Rede Neural MLP Nebulosa e Floresta Aleatória no monitoramento de desmatamento. O objetivo é melhorar a etapa de pré-classificação das imagens publicadas no *ForestWatchers* e reduzir a carga de trabalho dos voluntários.

Para as análises das três técnicas foram realizados testes de pequena escala com 37500 pixels, média escala com 500.000 pixels e grande escala com 8.341.210 pixels. De um modo geral, o classificador RF apresentou melhor desempenho, independente do número de árvores utilizados. A extração de uma métrica de incerteza da classificação fornecida pela RF permitiu obter uma redução bastante significativa da carga de trabalho dos voluntários. Assim, para um limiar de 0.90 de incerteza, foi possível obter um coeficiente de redução CR de pelo menos 86% (dependendo do número de árvores utilizado), no teste de grande escala, com mais de 8 milhões de pixels. Para imagens menores e menos complexas, os ganhos foram mais expressivos ainda.

Para avaliar o desempenho computacional do classificador, calculou-se para a técnica Floresta Aleatória o tempo de processamento para os testes com 10, 100, 500 e 1000 árvores e para o número de 37.500, 500.000 e 8.341.210 pixels. Desse cálculo, observou-se que o tempo de processamento varia linearmente com o número de árvores e com o número de pixels. Assim, considerando todos os resultados, o classificador RF com 10 árvores é uma solução adequada para o projeto *ForestWatchers*.

Na dissertação também foi gerado um Coeficiente de Redução que possibilitou reduzir de forma considerável o número de pixels a serem enviados para os voluntários realizarem a correção das imagens resultantes da classificação das imagens.

Para os trabalhos futuros, seria interessante investigar a paralelização dos algoritmos de classificação para uma maior redução do tempo de processamento da tarefa de

classificação. Sugere-se também analisar o algoritmo Floresta Aleatória com a técnica *bootstrap* linear, que atualmente não se encontra disponível no pacote Weka. Assim, seria possível minimizar o problema da repetição dos dados de entrada das árvores do algoritmo *bootstrap* aleatório utilizado neste trabalho.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGRAWAL, R. K.; BAWANE, N. G. Optimized neural network for classification of multispectral images. **International Journal on Signal and Image Processing**, Association of Computer Electronics and Electrical Engineers (ACEEE), v. 5, n. 1, p. 65, 2014. 4
- AHMADI, F. F.; ZOEJ, M. J. V.; EBADI, H.; MOKHTARZADE, M. The application of neural networks, image processing and cad-based environments facilities in automatic road extraction and vectorization from high resolution satellite images. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, Beijing, v. 37, n. Part B3b, p. 585–592, 2008. Disponível em: <[http://www.isprs.org/proceedings/XXXVII/congress/3b\\_pdf/109.pdf](http://www.isprs.org/proceedings/XXXVII/congress/3b_pdf/109.pdf)>. Acesso em: 10 março de 2013. 3
- Akar, Özlem and Güngör, Oğuz. Classification of multispectral images using random forest algorithm. **Journal of Geodesy and Geoinformation**, HKMO, v. 1, n. 106, p. 105–112, November 2012. 5
- ANDRADE, A. O.; MARTINS, W. **Redes neurais aplicadas ao processamento de imagens de satélite**. Goiás: Universidade Federal de Goiás (UFGO) - Escola de Engenharia Elétrica, 1997. 4 p. Grupo PIRENEUS. Disponível em: <<http://www.aoandrade.eletrica.ufu.br/Documents/PaperImagemDeSat%E9lite.pdf>>. Acesso em: 05 de abril de 2013. 2
- ARCANJO, J. S. **Desenvolvimento e teste de uma ferramenta de avaliação da qualidade dos dados e da confiabilidade dos voluntários para uma projeto de ciência cidadã**. 90 p. Dissertação de Mestrado em Computação Aplicada — Universidade Federal de São Paulo (UNIFESP), São José dos Campos, 2014. 9, 10, 11
- BALLINI, R.; CARNEIRO, A. A. F. M. Interpretação linguística da operação ótima de uma usina hidroelétrica através de redes neurais nebulosas adaptáveis. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, SciELO Brasil, v. 14, n. 3, p. 330–337, 2003. 22
- BARBIERI, A. L.; ARRUDA, G. D.; RODRIGUES, F. A.; BRUNO, O. M.; COSTA, L. da F. An entropy-based approach to automatic image segmentation of

satellite images. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 390, n. 3, p. 512–518, 2011. 5

BARTH, F. J. **Identificação de span utilizando random forest**. Faculdade BandTec e VAGAS Tecnologia, 2013. 11 p. Disponível em: <<http://fbarth.net.br/materiais/docs/aula08.pdf>>. Acesso em: 25 de maio de 2014. 26

BENYAMIN, D. **A gentle introduction to random forests, ensembles, and performance metrics in a commercial system**. November 2012. Disponível em: <<https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/>>. Acesso em: 11 de maio de 2014. 25

BRASIL, A. L. **Entropia de imagens**. Rio de Janeiro: Universidade Federal Fluminense - UFF, 2010. 3 p. Relatório técnico. Disponível em: <<http://www2.ic.uff.br/~aconci/entropia.PDF>>. Acesso em: 29 de junho de 2014. 31

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. 24, 25

BREIMAN, L.; CUTLER, A. **An introduction to random forest for beginners**. 1. ed. Califórnia, Estados Unidos: Salford Systems, 2014. 71 p. 24, 26, 27

BRUNO, R.; FOLLADOR, M.; PAEGELOW, M.; RENNO, F.; VILLA, N. Integrating remote sensing, gis and prediction models to monitor the deforestation and erosion in peten reserve, guatemala. **Annual Conference on Quantitative Geology from Multiple Sources (IAMG'2006)**, Belgium:Université de eLiège, Belgium, p. 1–6, 2006. 4

CARVALHO, A. R. **Uso da redes neurais otimizadas para recuperação do perfil de concentração de gases traço atmosféricos a partir de dados de satélites**. 208 p. Tese (Tese Doutorado em Computação) — Instituto Nacioando de Pesquisas Espaciais (INPE), Laborat 'orio de Computação Aplicada - LAC Curso de Pós-Graduação em Computação Aplicada, São José dos Campos, 2011. Disponível em: <<http://urlib.net/8JMKD3MGP7W/39R454E>>. Acesso em: 15 de maio de 2012. 9, 18, 19, 20, 37

CONRAD, C. C.; HILCHEY, K. G. A review of citizen science and community-based environmental monitoring: issues and opportunities.

**Environmental Monitoring and Assessment**, Springer, v.176, n. 4, p. 273–291, 2011. 9

EGMONT-PETERSEN, M.; RIDDER, D. de; HANDELS, H. Image processing with neural networks a review. **The Journal of the Pattern Recognition Society**, v. 35, n. 10, p. 2279–2301, October 2002. 2, 9

GEOSPATIAL Data Abstraction Library (GDAL). Open source geospatial foundation, 2007. Disponível em: <<http://www.gdal.org/>>. Acesso em: 06 de junho de 2013. 30, 34, 35

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. . **The WEKA data mining software: an update**. April 2014. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 09 de julho de 2014. 27, 28, 29, 30

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Estados Unidos: Morgan Kaufmann and Elsevier, 2012. 673 p. 23, 24, 31

HAYKIN, S. **Redes neurais princípios e práticas**. 3. ed. Canadá: McMaster University, 2001. 902 p. ISBN 978-85-7307-718-6. 18, 20, 38

HORNING, N. **Introduction to decision trees and random forests**. American Museum of Natural Historys and Center Biodiversity and Conservation, 2013. Disponível em: <[http://whrc.org/wp-content/uploads/2016/02/DecisionTrees\\_RandomForest\\_v2.pdf](http://whrc.org/wp-content/uploads/2016/02/DecisionTrees_RandomForest_v2.pdf)>. Acesso em: 10 abril de 2014. 23, 24

INSTITUTO Nacional De Pesquisas Espaciais (INPE): **Projeto PRODES Monitoramento da Floresta Amazônica por Satélite**. junho 1988. Disponível em: <<http://www.obt.inpe.br/prodes/index.php>>. Acesso em: 30 de março de 2013. 15, 16

INSTITUTO Nacional De Pesquisas Espaciais (INPE): **Sistema Deter Detecção de desmatamento em Tempo Real**. June 2004. Disponível em: <<http://www.obt.inpe.br/deter/>>. Acesso em: 30 de março de 2013. 13, 17

JIANG, J.; ZHANG, J.; YANG, G.; ZHANG, D.; ZHANG, L. Application of back propagation neural network in the classification of high resolution remote sensing image: Take remote sensing image of beijing for instance. In: INTERNATIONAL CONFERENCE ON GEOINFORMATICS, 18., 2010, Beijing. **Proceeding ...** Beijing: Insitute of Electrical and Eletronics Engeneers, 2010. p. 1–6. 4

KRIESEL, D. **A brief introduction to neural networks**. 1. ed. Alemanha: dkriesel.com, 2007. 286 p. Download  
location:[http : //www.dkriesel.com/en/science/neural<sub>n</sub>etworks](http://www.dkriesel.com/en/science/neural_networks). 18

KRUSE, R. Fuzzy neural network. **Physica A: Statistical Mechanics and its Applications**, Scholarpedia, v. 3, n. 11, p. 6043, 2008. 21

LIU, X.; YETIK, I. S. A maximum likelihood classification method for image segmentation considering subject variability. In: INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS - IEEE. **Proceedings...** Austin, TX: IEEE SOUTHWEST SYMPOSIUM ON IMAGE ANALYSIS INTERPRETATION (SSIAI), 2010. p. 125–128. ISBN 978-1-4244-7801-9. 4

LUZ, E. F. P.; CORRE, F. R.; GONZÁLEZ, D. L.; GREY, F.; RAMOS, F. M. The forestwatchers: a citizen cyberscience project for deforestation monitoring in the tropics. **Humam Computation - A Transdiciplinary Journal**, International Society for Optics and Photonics, v. 1, n. 2, p. 1–9, 2014. xi, xiii, 1, 6, 7

LUZ, E. F. P.; RAMOS, F. M. **Forest whatchers.net technical report**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2012. 30 p. Relatório técnico. Disponível em: <[\\$http://forestwatchers.net/\\$](http://forestwatchers.net/)>. Acesso em: 03 de março de 2014. 11, 12, 13, 14

MAS, J.; PUIG, H.; PALACIO, J.; SOSA-LÓPEZ, A. Modelling deforestation using gis and artificial neural networks. **The Journal of the Pattern Recognition Society**, v. 19, n. 5, p. 461—471, May 2004. 3

MATHUR, P.; GOVIL, R. Detecting temporal changes in satellite imagery using ann. In: INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN SPACE TECHNOLOGIES (RAST 2005), 2., 2005, Gramado, Brazil. **Proceedings ...** Gramado: Insitute of Electrical and Eletronics Engeneers, 2005. p. 645–647. ISBN 0-7803-8977-8. 3

MEDEIROS, A. R. C.; ARAÚJO, Y. B. d.; VIANNA, R. P. T.; MORAES, R. M. Modelo de suporte à decisão aplicado à identificação de indivíduos não aderentes ao tratamento anti-hipertensivo. **Saúde em Debate**, SciELO Brasil, v. 38, n. 100, p. 104–118, 2014. 23

NEAGOE, V.-E.; NEGHINA, M.; DATCU, M. A neural network approach for land-cover change detection in multi-temporal multispectral remote-sensing imagery. In: WORLD SCIENTIFIC AND ENGINEERING ACADEMY AND

SOCIETY (WSEAS), 11., 2011, Stevens Point, Wisconsin, USA. **Proceedings...** Stevens Point, Wisconsin, USA: Institute of Electrical and Electronics Engineers, 2011. p. 221–226. 4

NEAGOE, V.-E.; ROPOT, A.-D. A new neural approach for pattern recognition in space imagery. In: NATO ADVANCED RESEARCH WORKSHOP ON DATA FUSION TECHNOLOGIES FOR HARBOUR PROTECTION. **Proceedings...** Tallinn, Estonia: Springer and NATO Public Diplomacy Division, 2005. p. 283–289. 3

NEPOMUCENO, A. **Uso de rede neural artificial não supervisionada na classificação de dados de radar na banda-p para mapeamento de cobertura da terra em floresta tropical.** 197 p. (INPE-10236-TDI/901). Dissertação de Mestrado em Sensoriamento Remoto — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2004. 5

PEREIRA, L. A. d. A.; LACERDA, W. S.; VIEIRA, T. G. **Classificação automática de áreas cafeeiras em imagens de satélite, utilizando redes neurais artificiais.** Lavras, Minas Gerais, 2013. 98 p. Disponível em: <[http://repositorio.ufla.br/bitstream/1/5198/1/TCC\\_Classificacao\\_automatica\\_de\\_areas\\_cafeeiras\\_em\\_imagens\\_de\\_satelite\\_utilizando\\_redes\\_neurais\\_artificiais.pdf](http://repositorio.ufla.br/bitstream/1/5198/1/TCC_Classificacao_automatica_de_areas_cafeeiras_em_imagens_de_satelite_utilizando_redes_neurais_artificiais.pdf)>. Acesso em: 05 de janeiro de 2014. 4

PIMENTEL, T. R. G. **Classificação de padrões temporais de uso do solo e cobertura da terra em séries temporais de índice de vegetação utilizando um sistema neuro-difuso.** 102 p. Dissertação (Mestrado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2014. Disponível em: <[http://www.dpi.inpe.br/terrama2/lib/exe/fetch.php?media=docs:papers:xii\\_srh\\_natal\\_2014\\_alan.pdf](http://www.dpi.inpe.br/terrama2/lib/exe/fetch.php?media=docs:papers:xii_srh_natal_2014_alan.pdf)>. Acesso em: 30 de setembro de 2014. 21, 22, 27, 40

RIDDER, D. D.; DUIN, R. P.; EGMONT-PETERSEN, M.; VLIET, L. J. V.; VERBEEK, P. W. Nonlinear image processing using artificial neural networks. **Advances in Imaging and Electron Physics**, Academic Press, v. 126, n. 10, p. 352–450, October 2003. 3

RODRIGUES, J.; BEZERRA, L.; LIMA, O. A. **Simulação baseada em software livre de um sistema robótico fuzzy.** Fortaleza, 2011. Disponível em: <<http://mathworld.wolfram.com/RelativeEntropy.html>>. Acesso em: 10 de julho de 2013. 21



SANDRI, S.; CORREA, C. Lógica nebulosa. **V escola de redes neurais**, conselho nacional de redes neurais, v. 1, n. 5, p. C073–c090, 1999. 21

SAPKAL, A. T.; BOKHARE, C.; TARAPORE, N. Satellite image classification using the back propagation algorithm of artificial neural network. **technical article**, India, p. 1–4, 2006. Disponível em:

<[https://static.aminer.org/pdf/PDF/000/347/478/a\\_neural\\_network\\_classifier\\_for\\_occluded\\_images.pdf](https://static.aminer.org/pdf/PDF/000/347/478/a_neural_network_classifier_for_occluded_images.pdf)>. Acesso em: 14 de agosto de 2013. 3

SCHMIDHEINY, K. **The bootstrap**. 2012. Short guides to microeconometrics - Universität Basel. Disponível em:

<<http://kurt.schmidheiny.name/teaching/bootstrap2up.pdf>>. Acesso em: 5 de junho de 2014. 25

THE UNIVERSITY WAIKATO. **WEKA Manual for Version 3-7-13**. Nova Zelândia: IKGDD Explorations, 2015. 327 p. GNU General Public License.

Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 09 de julho de 2014. 27

WEISSTEIN, E. W. **Relative Entropy**. June 2000. Disponível em:

<<http://mathworld.wolfram.com/RelativeEntropy.html>>. Acesso em: 25 de setembro de 2014. 32

ZADEH, L. A. Fuzzy sets. **Information and Control**, Elsevier, v. 8, n. 3, p. 338–353, 1965. 21

## **PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE**

### **Teses e Dissertações (TDI)**

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

### **Manuais Técnicos (MAN)**

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

### **Notas Técnico-Científicas (NTC)**

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

### **Relatórios de Pesquisa (RPQ)**

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

### **Propostas e Relatórios de Projetos (PRP)**

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

### **Publicações Didáticas (PUD)**

Incluem apostilas, notas de aula e manuais didáticos.

### **Publicações Seriadas**

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

### **Programas de Computador (PDC)**

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

### **Pré-publicações (PRE)**

Todos os artigos publicados em periódicos, anais e como capítulos de livros.