



EVELINE APARECIDA PEREIRA

**DETECÇÃO DE MUDANÇAS EM ÁREAS DE CERRADO
USANDO INTELIGÊNCIA ARTIFICIAL**

**LAVRAS – MG
2020**

EVELINE APARECIDA PEREIRA

**DETECÇÃO DE MUDANÇAS EM ÁREAS DE CERRADO USANDO
INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

Prof. Dr. Luís Marcelo Tavares de Carvalho
Orientador

Prof. Dr. Fausto Weimar Acerbi Júnior
Coorientador

**LAVRAS – MG
2020**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Pereira, Eveline Aparecida.

Detecção de Mudanças em Áreas de Cerrado Usando Inteligência Artificial / Eveline Aparecida Pereira. - 2020.

45 p. : il.

Orientador(a): Luís Marcelo Tavares Carvalho.

Coorientador(a): Fausto Weimar Acerbi Júnior.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras,
2020.

Bibliografia.

1. Aprendizado de Máquinas. 2. OBIA. 3. Cerrado. I. Carvalho, Luís
Marcelo Tavares. II. Acerbi Júnior, Fausto Weimar. III. Título.

EVELINE APARECIDA PEREIRA

**DETECÇÃO DE MUDANÇAS EM ÁREAS DE CERRADO USANDO
INTELIGÊNCIA ARTIFICIAL
CHANGE DETECTION IN SAVANNA AREAS USING ARTIFIAL INTELLIGENCE**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

Aprovada em 27 de fevereiro de 2020.

Dr. Danton Diego Ferreira UFLA

Dr. Fausto Weimar Acerbi Júnior UFLA

Prof. Dr. Luís Marcelo Tavares de Carvalho
Orientador

**LAVRAS – MG
2020**

*Aos meus pais,
Antônio Roberto e Antônia Rita Ferreira.*

Dedico

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA), que através do Departamento de Ciências Florestais (DCF) me proporcionou a oportunidade de conclusão do mestrado em Engenharia Florestal.

Ao professor Luís Marcelo Tavares de Carvalho (Passarinho), pela dedicação, confiança, apoio e amizade. Obrigada por toda compreensão e ensinamento. Me sinto imensamente grata por tê-lo como orientador.

Ao meu coorientador Fausto Acerbi Júnior, agradeço pela amizade, incentivo e apoio. Sua coorientação foi fundamental, obrigada por tudo.

Aos meus amigos do LabGeo, obrigada por compartilharem comigo os melhores e também os não tão melhores momentos dessa dissertação.

Ao LEMAF/FUNDECC, agradeço imensamente pelo tempo que pude trabalhar, aprender, gerar e compartilhar conhecimento ao lado de profissionais incríveis e assim poder concluir com sucesso esta etapa da minha vida acadêmica e profissional, sou muito grata.

Ao professor Danton Diego Ferreira, pelo apoio e ensinamento, saiba que você acrescentou muito a este trabalho, desde sua concepção.

Aos meus pais, saibam que devo tudo a vocês, que me ensinaram a mais linda lição, nunca desistir. Obrigada por fazerem parte da minha vida e me apoiarem em todo esse percurso. À vocês dedico este trabalho, com todo meu coração.

À minha família e amigos, pelo carinho, amor e constante apoio, vocês fizeram dessa caminhada um pouco mais leve.

À Deus, autor da vida, por sempre guiar meus passos, me dando força, discernimento, sabedoria e fé.

RESUMO

O bioma cerrado está sob constante pressão antrópica e poucos esforços tem sido feitos no âmbito do monitoramento dessas mudanças no uso do solo. O sensoriamento remoto aliado a inteligência artificial fornecem ferramentas eficientes e rápidas para detectar mudanças. As questões científicas abordadas neste estudo foram: qual tipo de atributo (espacial ou espectral) e ou a combinação deles melhor diferenciam as mudanças sazonais dos processos de antropização em imagens NDVI-bitemporais? E qual o efeito da intensidade amostral do monitoramento no desempenho dos classificadores utilizando inteligência artificial? O estudo explorou ambas informações, espaciais e espectrais, derivadas de imagens NDVI bi-temporais Landsat na análise do monitoramento, empregando classificadores de alto desempenho: *Neural Network Multilayer Perceptron* (MLP), *Support Vector Machine* (SVM) e *Random Forest* (RF). Foi analisado a intensidade amostral e a robustez dos algoritmos em cada um dos conjuntos de atributos da classificação, a rede MLP obteve a melhor generalização com 75,16% de acurácia global e maior robustez em relação a variação da intensidade amostral. O algoritmo Multilayer Perceptron (MLP) treinado foi aplicado numa área contígua, detectando as mudanças com a precisão de 56% indicando algumas limitações do método. Portanto, os atributos espaciais, derivados de imagens bi-temporais NDVI são capazes de detectar com precisão os desmatamentos e queimadas ocorridos no cerrado, sendo insensíveis as mudanças causadas pelo período sazonal do ambiente.

Palavras-chave: Aprendizado de Máquinas. OBIA. Cerrado.

ABSTRACT

Brazil contains large tracts of native vegetation, including large areas of tropical Brazilian Savannas biome, which has been threatened due to the expansion of anthropic activities. In the last years, Remote Sensing (RS) data combined with Artificial Intelligence (AI) have been used to identify the dynamic of the Land use/Land Cover Change (LULCC) of these areas, producing LULCC maps with high accuracy. However, the choice of the AI algorithm and the selection data attributes for the learning process are crucial steps, especially in environments influenced by seasonal variations. Considering these circumstances, the study focus in the following questions: a) what type of attribute (spatial or spectral) or their combination could better differentiate the seasonal changes produced by weather conditions, from atrophic changes in RS images; b) what is the effect of the training sample size into different AI classifiers to produce change maps. Thus, spatial and spectral information were extract for objects generated from Landsat NDVI images in a Tropical Savanna area, acquired at different seasonal periods. The Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Random Forest (RF) algorithms were compared. The MLP produced the most accurate change map, with 75,16% of global accuracy and greater robustness in relation to the variation of the sample intensity. In order to evaluate the generalization capacity of the algorithm, the trained MLP was used to detect changes in contiguous Landsat tiles. The results showed a decrease to 56% of global accuracy, which indicates a limitation of the method. Therefore, the spatial attributes were capable of accurately differentiate deforestation and fires sites, from seasonal changes.

Keywords: Machine Learning. OBIA. Cerrado.

LISTA DE FIGURAS

Figura 1 –	Representação gráfica do semivariograma experimental, $\gamma(h)$ – estimador da semivariância para cada distância h ; (σ^2) – Sill; (ϕ) – Range; (τ^2) – Nugget.	18
Figura 2 –	Fluxograma da metodologia. (1) Aquisição e pré-processamento das imagens; (2) Segmentação multiresolução; (3) Cálculo da imagem diferença; (4) Sobreposição dos objetos na imagem diferença; (5) Geração dos atributos espectrais e espaciais; (6) Detecção de mudanças; (7) Avaliação da acurácia.	21
Figura 3 –	Área de estudo. Localização da área de estudo no Brasil, no estado de Minas Gerais - MG.	22
Figura 4 –	Exemplos de valores do semivariograma extraído da imagem NDVI diferença (a) segmento de vegetação natural em 2015; (b) segmento com sazonalidade da vegetação em 2016 – valores de semivariância constantes; (c) segmento de vegetação natural em 2015 e (d) segmento com desmatamento em 2016 – aumento no valor de semivariância.	32
Figura 5 –	Exemplos em (a) áreas afetadas pela sazonalidade que apresentaram padrão homogêneo nos atributos do semivariograma; (b) áreas de mudanças nas quais há um acréscimo nos atributos do semivariograma. .	33
Figura 6 –	Mapa final em graduação de cores variando de verde (áreas de não mudança e sazonais) a vermelho (áreas de desmatamento e queimadas). Em (a) Polígono de uma área em 2015; (b) O mesmo polígono no ano de 2016 e (c) A segmentação do polígono desmatado.	36
Figura 7 –	Área de abrangência. Localização da área de abrangência no grid de imagens Landsat, no estado de Minas Gerais - MG.	38

LISTA DE TABELAS

Tabela 1 –	Características espectrais do sensor Operational Land Imager (OLI).	24
Tabela 2 –	Índices do semivariograma descritos por Balaguer et al. (2010) agrupados em: (A) Parâmetros do semivariograma próximos à origem; (B) Parâmetros usando valores até o primeiro máximo.....	25
Tabela 3 –	Desempenho dos classificadores nos conjuntos de atributos da detecção de mudanças; em que MD (classe mudança) e NM (classe não mudança).....	30
Tabela 4 –	Capacidade de generalização dos classificadores nos conjuntos de atributos da detecção de mudanças na imagem Landsat 8; em que MD (classe mudança) e NM (classe não mudança).	31
Tabela 5 –	Variação da intensidade amostral, total de amostras coletadas, número de amostras de treinamento e teste.....	34
Tabela 6 –	Acurácia da detecção para os três classificadores em cada uma das intensidades amostrais.	34
Tabela 7 –	Acurácia da detecção para a área de estudo.....	36
Tabela 8 –	Acurácia da detecção para a área de abrangência.	38

LISTA DE SIGLAS

ANN	<i>Artificial Neural Network</i>
BagTE	<i>Ensemble</i>
CE	Computação Evolutiva
EROS	<i>Earth Observation and Science</i>
ERTS	<i>Earth Resources Technology Satellite</i>
FDO	Derivada próxima da origem
FML	Primeiro valor máximo de distância
GEOBIA	<i>Geographic Object-based Image Analysis</i>
KNN	<i>K - Nearest Neighbor</i>
MD	Classe mudança
MFM	Média do semivariograma até o primeiro máximo
MLP	<i>Neural Network Multilayer Perceptron</i>
Mtry	Número de variáveis de entrada
NB	<i>Naive Bayes</i>
NM	Classe não mudança
NDVI	<i>Normalized Difference Vegetation Index</i>
Ntree	Número de árvores de decisão
OLI	<i>Operational Land Imager</i>
RBF	Função de base radial
RF	<i>Random Forest</i>
RMM	Razão entre a semivariância no primeiro máximo e a média do semivariograma até esse máximo
RNAs	Redes Neurais Artificiais
RVF	Semivariância na primeira distância
SDF	Diferença de segunda ordem entre o primeiro intervalo e o primeiro máximo
SE	Sistema de Enxames
SF	Sistemas Fuzzy
SIA	Sistemas Imunes
SVM	<i>Support Vector Machine</i>
USGS	<i>United States Geological Survey</i>
VFM	Variância dos valores de semivariância

LISTA DE SÍMBOLOS

h_{sc}	Parâmetro de escala
l	Comprimento da borda do objeto
lr	Menor comprimento possível dado por um retângulo de pixels
$N(h)$	Número de pares de pontos separados pela distância h
n_b	Número de bandas espectrais
np	Número de pixels
w_{cp}	Parâmetro espacial
w_m	Pesos dos neurônios
w_{sp}	Parâmetro espectral
x_m	Entradas dos neurônios
$\hat{\gamma}(h)$	Estimador da semivariância para cada distância h
$Z(x)$	Valor da variável regionalizada no ponto x
$Z(x+h)$	Valor no ponto $x+h$
σ^2	Sill
σ_b	Variância dos valores dos pixels dentro do objeto para a banda espectral b
τ^2	Nugget
φ	Range

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA	15
2.1	O bioma Cerrado	15
2.2	Detecção de mudanças	15
2.3	Análise de imagens baseada em objetos	16
2.4	Índices espaciais do semivariograma	17
2.5	Algoritmos machine learning	19
3	METODOLOGIA	21
3.1	Área de estudo	21
3.2	Dados	23
3.3	Segmentação multiresolução	23
3.4	Análise espectral	24
3.5	Análise espacial	25
3.6	Detecção de mudanças	26
3.7	Classificadores de inteligência artificial	26
3.8	Análise da acurácia da detecção	27
3.9	Análise da abrangência da detecção	27
4	RESULTADOS E DISCUSSÃO	29
4.1	Análise dos conjuntos de atributos	29
4.2	Análise da intensidade amostral	33
4.2.1	Análise do desempenho dos classificadores nas diferentes intensidades amostrais	34
4.3	Análise da acurácia	35
4.4	Análise da abrangência da detecção	37
5	CONCLUSÃO	39
	REFERÊNCIAS	40

1 INTRODUÇÃO

O Cerrado brasileiro está entre os biomas mais ameaçados do planeta devido às altas taxas de desmatamento e à escassez de áreas legalmente protegidas (HOEKSTRA et al., 2005), sendo a principal causa atribuída a expansão da agricultura e pecuária (FRANÇOSO et al., 2015). As áreas de Cerrado são compostas por um mosaico de tipos de cobertura de solo, com características da forte sazonalidade no clima o que torna um desafio o monitoramento das mudanças por sensoriamento remoto (ACERBI JÚNIOR et al., 2015; PEREIRA et al., 2019; SILVEIRA et al., 2018a). Os efeitos da fenologia da vegetação podem ser captados como áreas desmatadas, fornecendo estimativas pouco precisas (LU et al., 2016).

O monitoramento de mudanças baseado em objetos fornece melhor precisão nos resultados (ACERBI JÚNIOR et al., 2015; HUSSAIN et al., 2013; RUIZ et al., 2011; SILVEIRA et al., 2017b; WU et al., 2015), visto que permite a incorporação da informação espacial, como os semivariogramas, além de dispensar o uso de séries temporais densas (SILVEIRA et al., 2017a).

O semivariograma derivado de imagens NDVI (*Normalized Difference Vegetation Index* – NDVI) tem sido empregado com precisão em estudos de detecção de mudanças no Cerrado (ACERBI JÚNIOR et al., 2015; COSTANTINI et al., 2012; SILVEIRA et al., 2017c). Estudos de mudanças em áreas sazonais também empregaram os índices extraídos do semivariograma que identificam pontos singulares na curva do semivariograma e melhoram as informações contidas na correlação de semivariância em distâncias curtas, garantindo bons resultados nas classificações (PEREIRA et al., 2019; SILVEIRA et al., 2017c).

A amostragem em sensoriamento remoto ainda é muito discutida e não há um consenso sobre o assunto. Zheng, Wang e Cui (2019), em estudo de classificação, demonstraram bons resultados de acurácia usando pequenas amostras de treinamento. Heydari e Mountrakis (2018) analisaram o desempenho dos classificadores *Naive Bayes* (NB), *Support Vector Machine* (SVM), *K – Nearest Neighbor* (KNN), *Ensemble* (BagTE) e *Artificial Neural Network* (ANN) em diferentes tamanhos de amostras de treinamento, constatando a sensibilidade de alguns dos classificadores as amostras pequenas e à sua distribuição espacial.

Embora os semivariogramas tenham sido aplicados para o monitoramento de mudanças poucos estudos têm se concentrado na abordagem da detecção comparando os contextos espacial e espectral, a fim de definir os melhores atributos a serem utilizados. Outro ponto crítico e de suma importância a influência da intensidade amostral usada no monitoramento no desempenho dos classificadores.

Assim, o principal objetivo deste estudo foi explorar ambos os atributos, espectrais e espaciais, derivados de imagens NDVI bi-temporais e a intensidade amostral empregada na análise do monitoramento, para determinar uma metodologia que seja precisa na detecção de mudanças em áreas com forte influência da fenologia na vegetação, tais como o Cerrado brasileiro, garantindo eficiência na intensidade amostral. As questões científicas abordadas neste estudo foram:

- (1) Qual tipo de atributo (espacial ou espectral) e ou a combinação deles melhor diferenciam as mudanças sazonais dos processos de antropização em imagens NDVI-bitemporais?
- (2) Qual o efeito da intensidade amostral no desempenho dos classificadores utilizando inteligência artificial?
- (3) E qual a acurácia da abrangência da detecção, ou seja, acurácia de uma imagem contígua classificada com um algoritmo treinado na área de estudo?

2 REVISÃO DE LITERATURA

2.1 O bioma Cerrado

Os ecossistemas de savana representam 20% da superfície terrestre global, predominando principalmente na Austrália, África e nas Américas (SCHWIEDER et al., 2016). Inserido no bioma das savanas o Cerrado brasileiro é responsável por abrigar 30 % da biodiversidade brasileira, considerado um dos hotspots para a conservação da biodiversidade mundial, possuindo a mais rica flora entre as savanas do mundo, caracterizados por alta concentração de espécies endêmicas e elevada perda de habitat (OLIVEIRA-FILHO; RATTER, 2002).

Caracterizado por padrões de paisagem de pastagens abertas a densas florestas, definindo um gradiente heterogêneo na estrutura da vegetação em todo o bioma, a combinação de fisionomia lenhosa muito diversificada e composição de espécies de alta biodiversidade torna o Cerrado ponto de acesso de riqueza de flora entre as savanas do mundo (HILL et al., 2016; KLINK; MACHADO, 2005).

O Cerrado do Brasil cobre aproximadamente 45% de todo o país, principalmente nos estados de Goiás, Mato Grosso do Sul, Mato Grosso, Tocantins e Minas Gerais. Os tipos de cerrado totalizam mais de 2.000.000 km² de extensão territorial, sendo que aproximadamente mais da metade dessa área vem sendo substituída em outros usos (KLINK; MACHADO, 2005; MMA, 2015).

O bioma do Cerrado é categorizado como segundo maior bioma brasileiro com mudanças induzidas pelo homem, a maior parte da grande conversão do uso do solo se deve à expansão agrícola, extração predatória de madeira, esgotamento de recursos naturais e baixa taxa de áreas legalmente protegidas, em que 7,44% são protegidos e apenas 2,91% são totalmente preservadas (MMA, 2011).

2.2 Detecção de mudanças

O uso de imagens de sensores orbitais tem se tornado imprescindível no mapeamento e monitoramento de mudanças na estrutura da paisagem (ACERBI JÚNIOR et al., 2015). O NDVI tem sido amplamente utilizado em estudos de detecção de mudanças, trabalhos recentes demonstram a aplicabilidade do NDVI no monitoramento da vegetação (AQUINO; OLIVEIRA, 2012; BAYMA; SANO, 2015; FERREIRA, 2015; SILVEIRA et al., 2017b).

Com o intuito de aprimorar a detecção de mudanças na cobertura do solo, técnicas geoestatísticas, que proveem descrições quantitativas de variáveis regionalizadas, veem sendo aplicadas (GALEANA-PIZAÑA et al., 2014; VIANA et al., 2012). A integração de técnicas geoestatísticas aplicadas ao sensoriamento remoto, utilizando semivariogramas derivados de imagens NDVI, contribue para o aumento da acurácia de detecção de mudanças baseada em objetos (SILVEIRA et al., 2017b).

Acerbi Júnior et al. (2015) utilizaram semivariogramas derivados das imagens NDVI para detecção de mudanças em áreas de cerrado, demonstrando a relação entre as mudanças observadas na imagem NDVI e os índices do semivariograma. Costantini et al. (2012) em estudos nas bacias de lagos vulcânicos na Itália, integraram técnicas geoestatísticas e análises espaciais de NDVI a fim de detectar distúrbios em áreas vegetadas, comprovando a eficiência dos atributos espaciais na indicação da susceptibilidade à mudanças. Silveira et al. (2017a) utilizaram o NDVI para detecção de mudanças após o desastre ambiental em Mariana, Minas Gerais, concluindo que atributos espaciais calculados a partir dos semivariogramas derivados da imagem NDVI são precisos na detecção de mudanças. Balaguer et al. (2010) criaram um conjunto de índices extraídos do semivariograma usando imagens de alta resolução espacial para a classificação de agricultura, demonstrando melhorias na classificação ao combinar atributos espaciais e espectrais.

Os atributos espaciais estão altamente relacionados a variabilidade dos valores de semivariância na imagem NDVI diferença, principalmente os valores alcançados até o primeiro máximo, que são eficazes na detecção de mudanças (SILVEIRA et al., 2017b), sendo capazes de categorizar a mudança causada pelo homem e as mudanças fenológicas da vegetação, o que somente com o NDVI não é possível detectar.

A lacuna existente é a exploração de quais atributos espaciais são melhores a ser utilizadas na detecção de mudanças, em qual contexto, a partir de qual algoritmo e intensidade amostral adotadas, questões a serem discutidas nesse trabalho.

2.3 Análise de imagens baseada em objetos

A análise de imagem baseada em objetos geográficos (*Geographic Object-based Image Analysis - GEOBIA*) tem como centro de estudo os objetos criados através de imagens de sensoriamento remoto, a partir dos quais é possível interpretar informações espaciais, espectrais e temporais (BLASCHKE, 2010).

Segmentação é o processo de assimilação de objetos ou regiões com características similares numa imagem, resultando em segmentos. Portanto, segmento é um conjunto de pixels espectralmente semelhantes e espacialmente adjacentes que representam um único objeto na imagem (GENELETTI; GORTE, 2003).

No processo de criação de objetos os algoritmos de segmentação baseiam-se em duas propriedades dos níveis de quantização dos pixels: a descontinuidade onde os algoritmos buscam mudanças abruptas nos valores de nível de cinza para delimitação dos limites e a similaridade em que os objetos são divididos buscando crescimento de regiões e fusão (GONZALEZ; WOODS, 2000). É um processo preliminar a classificação baseada em objetos, aplicada principalmente em imagens de alta resolução espacial, garantindo melhor distinção dos objetos e mapas mais precisos (YAN et al., 2006).

Na segmentação de imagens, além dos valores de média e desvio padrão da reflectância dos objetos, podem ser obtidas as medidas espectrais (inclusão de bandas multiespectrais e pancromática), medidas espaciais (área, comprimento, largura e direção), medidas morfológicas (forma e textura), medidas contextuais (relação aos vizinhos e análise de proximidade) e medidas temporais (séries temporais) (NAVULUR, 2006). Além de permitirem a incorporação de informação espacial, tais como índices derivados de semivariogramas (uma ferramenta geoestatística) para análise da heterogeneidade espacial (SILVEIRA et al., 2017a).

Os objetos permitem novas análises quando comparados às análises pixel a pixel por possuírem informação espectral adicional e relação espacial mais próxima do mundo real, além do que os objetos são verdadeiros objetos geográficos enquanto que os pixels são representações de valores espectrais (HUSSAIN et al., 2013). Entretanto, mesmo com o avanço de técnicas inovadoras que melhoram a precisão das classificações faz-se necessário um entendimento do contexto da paisagem, principalmente em áreas heterogêneas, sendo a associação da interpretação visual com os classificadores automáticos imprescindíveis na busca melhores precisões (SHALABY; TATEISHI, 2007).

2.4 Índices espaciais do semivariograma

A teoria das variáveis regionalizadas ou geoestatística pressupõe que cada dado é modelado com uma variável aleatória que pode ser expressa pela soma de três componentes: uma componente estrutural, uma componente aleatória e um erro aleatório (MELLO, 2004). O semivariograma é o método geoestatístico empregado para identificação e descrição da

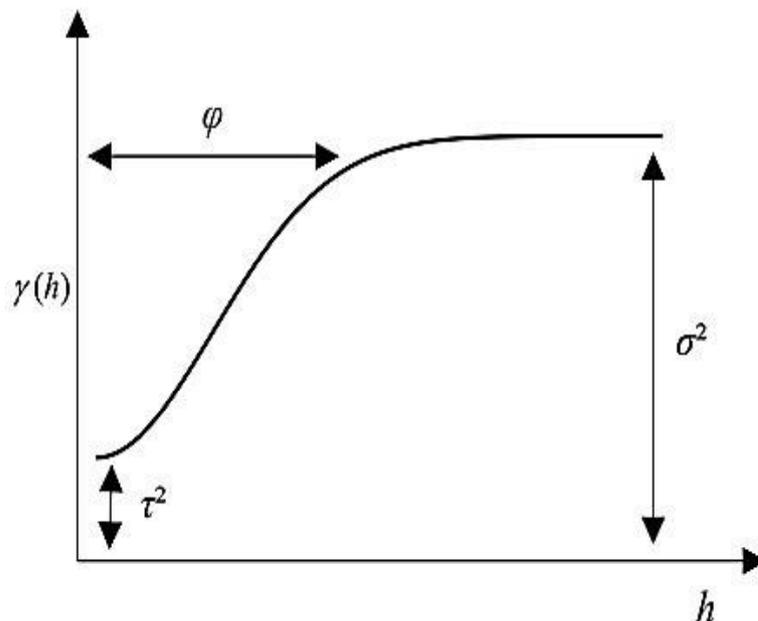
correlação espacial entre unidades amostradas (DUARTE, 2000; MELLO, 2004), configurando uma função de semivariâncias em relação às respectivas distâncias (Equação 1).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (1)$$

Onde $\hat{\gamma}(h)$ é o estimador da semivariância para cada distância h , $N(h)$ é o número de pares de pontos separados pela distância h , $Z(x)$ é o valor da variável regionalizada no ponto x e $Z(x+h)$ é o valor no ponto $x+h$.

A semivariância é definida como uma medida do grau de dependência espacial entre duas variáveis (NOGUEIRA, 2013). É entendida como a metade da variância entre diferentes observações da variável aleatória, sendo caracterizadas por três parâmetros: Sill, Range e Nugget (ISAAKS; SRIVASTAVA, 1989). O Sill (patamar) é o patamar alcançado pelos valores de semivariância e indica a quantidade de variabilidade espacial dos dados; Range (alcance) é a distância em que o semivariograma atinge o Sill, indicando até qual distância os pontos são correlacionados. O Nugget (efeito pepita) é a combinação de erros de amostragem e variações em pequenas escalas (CURRAN, 1988).

Figura 1 – Representação gráfica do semivariograma experimental, $\hat{\gamma}(h)$ – estimador da semivariância para cada distância h ; (σ^2) – Sill; (ϕ) – Range; (τ^2) – Nugget.



Fonte: Do autor (2020).

Balaguer et al. (2010) criaram um conjunto de índices extraídos do semivariograma usando imagens de alta resolução espacial. Uma vantagem do conjunto de recursos proposto, em oposição aos valores semivariância, é que são capazes de sintetizar as informações mais relevantes sobre a forma do semivariograma em poucos recursos. Eles identificam os pontos singulares e melhoram as informações contidas nas primeiras defasagens, onde correlação espacial em distâncias curtas é maior.

2.5 Algoritmos *machine learning*

A inteligência computacional é uma tecnologia inovadora que tem sido empregada em análises e modelagens de variáveis, baseada nas características da natureza, isto é, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação. A computação natural é subdividida em três categorias: a Vida Artificial, a Inteligência Computacional e os Sistemas Dinâmicos Não Lineares. Os algoritmos usados na Inteligência Computacional, ou os algoritmos inteligentes, procuram representar a natureza e o biológico, entre eles estão: Redes Neurais Artificiais (RNAs), Computação Evolutiva (CE), Sistema de Enxames (SE), Sistemas Imunes (SIA) e Sistemas *Fuzzy* (SF) (ENGELBRECHT, 2007).

O algoritmo de árvores de decisão *Random Forest* (RF) classifica ou faz previsão a partir de um conjunto de árvores criadas com o auxílio de um subconjunto de amostras de treinamento utilizando a substituição, cada árvore é produzida de forma independente e cada nó é dividido usando um número de recursos que é definido pelo usuário (BELGIU; DRĂGU, 2016; BREIMAN, 2001; LU; WENG, 2007). De acordo com Breiman (2001), dentre outras vantagens do classificador *Random Forest* está a alta acuracidade comparada aos algoritmos atuais, eficiência em grandes base de dados, robustez para outliers e ruídos e, finalmente capacidade de lidar com enormes quantidades de variáveis de entrada, sem eliminação.

As Redes Neurais Artificiais surgiram com o intuito de representar o cérebro humano através da modelagem matemática. As entradas dos neurônios (x_m) estão conectadas a seus pesos (w_m) que por sua vez representam as sinapses que podem ser positivas ou negativas. Cada entrada é multiplicada pelo seu peso, indicando a importância de cada entrada na saída. Posteriormente, ocorre o processamento de cada neurônio por uma soma ponderada dos valores de $x_i w_i$ recebidos, se o valor obtido ultrapassar o limiar do neurônio a função é ativada e gera a saída, dentre as funções de ativação estão a tangente, hiperbólica, logística, semilinear ou limiar (HAYKIN, 2001).

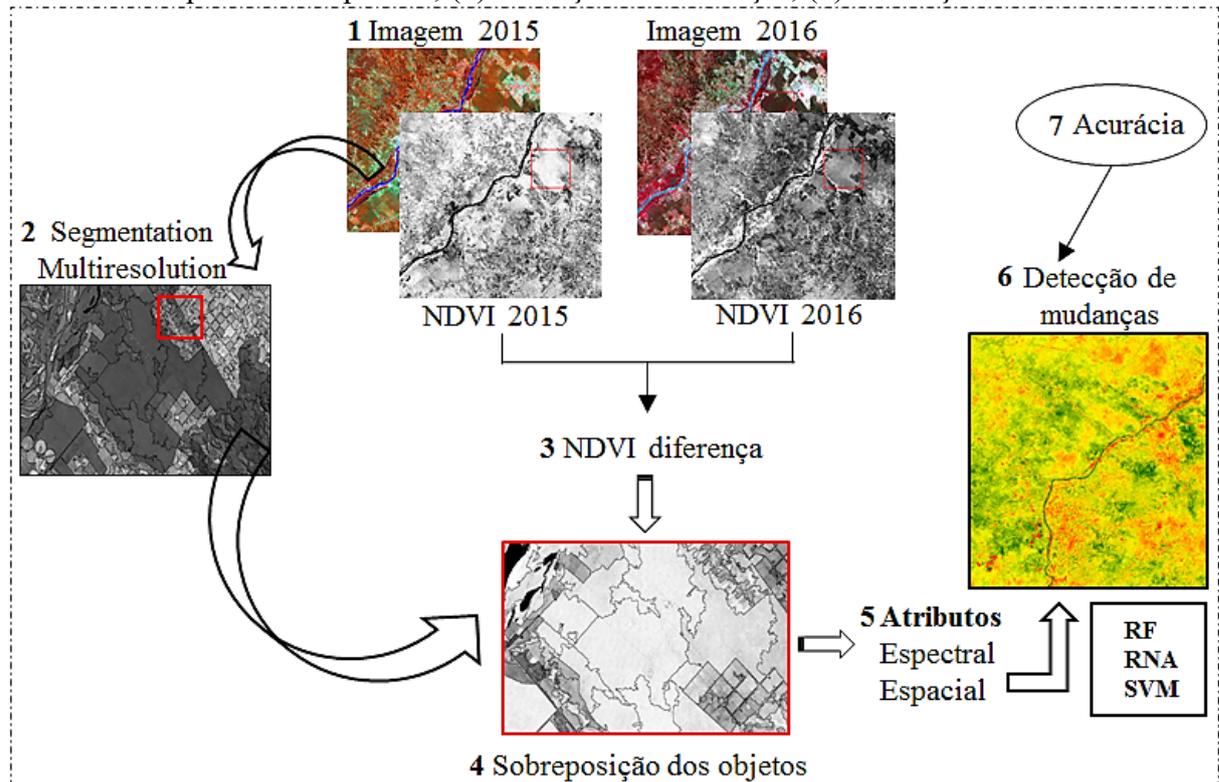
Dentre as principais características de uma RNA estão o mapeamento de entrada e saída de dados, a tolerância a falhas e ruídos, a adaptabilidade dos pesos, isto é, uma rede treinada pode ser retreinada ou projetada para modificar os seus pesos em tempo real, a capacidade de informar a confiança de uma decisão tomada, a aprendizagem a partir de exemplos e a generalização para dados incógnitos (HAYKIN, 2001).

O SVM é um conjunto de máquinas de vetor de aprendizado, citado entre os algoritmos de desempenho superiores em detecção de mudanças, principalmente na presença de classes heterogêneas e de baixa intensidade amostral (WU et al., 2015). As máquinas de suporte operam assumindo que cada entrada terá uma relação única com a variável resposta, dessa forma o conjunto de relações entre os preditores torna-se suficiente para identificar as regras de previsão da variável resposta em dados nunca antes vistos. De maneira simplificada, os SVM's projetam os dados de entrada num espaço de dimensão muito maior, permitindo que os dados possam ser linearmente separáveis. Alguns trabalhos utilizaram SVM com sucesso na resolução de problemas de classificação de vegetação (GARCÍA-GUTIÉRREZ et al., 2015; SILVEIRA et al., 2017c; WU et al., 2015).

3 METODOLOGIA

O estudo explorou ambas as informações, espaciais e espectrais, derivadas de imagens NDVI bi-temporais Landsat para desenvolver uma metodologia que seja precisa na detecção de mudanças em áreas com forte influência da fenologia da vegetação, garantindo eficiência na intensidade amostral. As mudanças no uso e cobertura da Terra foram detectadas numa abordagem baseada em objetos utilizando algoritmos de inteligência computacional, conforme demonstrado no fluxograma da Figura 2.

Figura 2 – Fluxograma da metodologia. (1) Aquisição e pré-processamento das imagens; (2) Segmentação multiresolução; (3) Cálculo da imagem diferença; (4) Sobreposição dos objetos na imagem diferença; (5) Geração dos atributos espectrais e espaciais; (6) Detecção de mudanças; (7) Avaliação da acurácia.



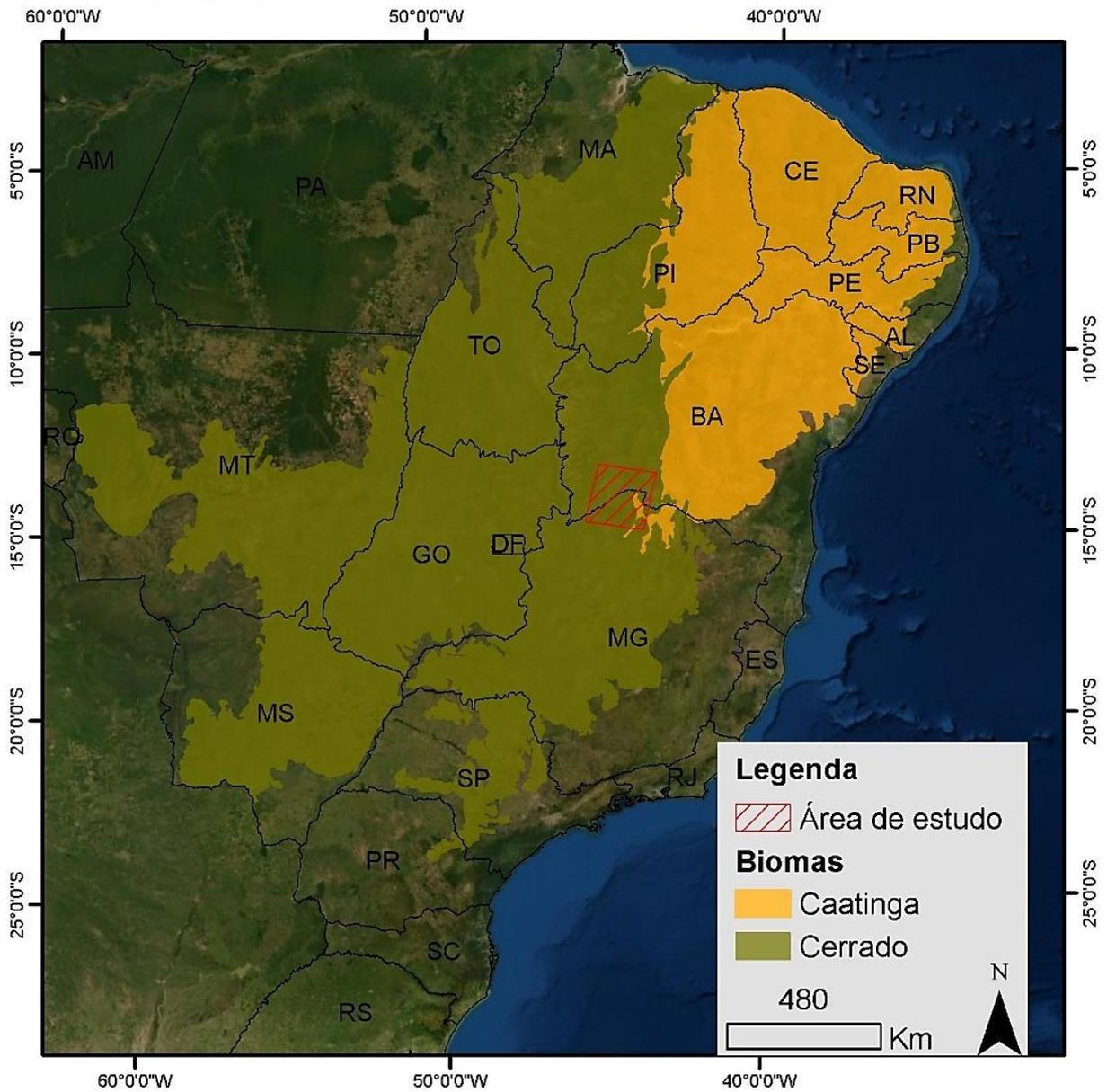
Fonte: Do autor (2020).

3.1 Área de estudo

A área de estudo está localizada no norte da *Bacia Hidrográfica dos Afluentes Mineiros do Médio São Francisco*, Brasil. Situada entre 43°00' a 46°00' oeste de longitude e 14°00' a 16°00' sul de latitude. Os tipos de vegetação predominantes são os cerrados (pastos abertos), cerrado sensu stricto (pastos abertos com arbustos esparsos), floresta decídua

(predominância de indivíduos decíduos cuja perda de folhagem atinge mais de 50%) e pântano de palmeiras (veredas) (CARVALHO; DUTRA, 2008). O clima da região é tropical úmido, com inverno seco e verão chuvoso (PEEL; FINLAYSON; MCMAHON, 2007) (FIGURA 3).

Figura 3 – Área de estudo. Localização da área de estudo no Brasil, no estado de Minas Gerais - MG.



3.2 Dados

Para a realização deste estudo, foram utilizadas imagens provenientes da missão *Earth Resources Technology Satellite* (ERTS), chamada Landsat, sensor *Operational Land Imager* (OLI) a bordo do satélite Landsat 8, com imagens pós-processadas a nível do subproduto *Landsat Surface Reflectance*. As imagens, fornecidas pelo *United States Geological Survey for Earth Observation and Science* (USGS/EROS) possuem as correções geométricas e radiométricas e o critério de seleção das imagens foi a presença de no máximo 20% de cobertura de nuvens em relação à área de estudo.

3.3 Segmentação multiresolução

A segmentação de imagens objetiva agrupar pixels espectralmente semelhantes e espacialmente adjacentes, formando objetos homogêneos, minimizando a variabilidade dentro do objeto em relação à variabilidade entre objetos (DESCLÉE; BOGAERT; DEFOURNY, 2006). O algoritmo utilizado neste processo é o de segmentação multirresolução, descrito por Baatz e Schäpe (2000) e implementado no software *eCognition*, descrito na Equação 2.

$$w_{sp} \sum_{nb} w_b \sigma_b + (1 - w_{sp}) \left[w_{cp} \frac{l}{\sqrt{np}} + (1 - w_{cp}) \frac{l}{lr} \right] \leq h_{sc} \quad (2)$$

Onde, há parâmetros que são definidos pelo usuário, como o w_{sp} que é o parâmetro espectral variando de 0 a 1, zero é o peso total para homogeneidade espectral e 1 para forma do objeto; w_{cp} é o parâmetro espacial que tem a mesma interpretação do w_{sp} , porém para quesitos de compacidade e suavidade dos objetos; e h_{sc} é o parâmetro de escala que controla o tamanho dos objetos. Já os parâmetros automáticos são o n_b que é o número de bandas espectrais, σ_b é a variância dos valores dos pixels dentro do objeto para a banda espectral b , l é o comprimento da borda do objeto, np é o número de pixels, e lr é o menor comprimento possível dado por um retângulo de pixels.

O parâmetro escala determina um limiar de quantos pixels vizinhos podem ser mesclados para formar um objeto na imagem, ou seja, a heterogeneidade máxima permitida para um objeto (MUI; HE; WENG, 2015), esse parâmetro foi definido pelo método de tentativa e erro através de uma análise visual da segmentação gerada. A definição do

parâmetro escala é o ponto mais crítico na segmentação multirresolução, influenciando diretamente na extração dos índices do semivariograma (SILVEIRA et al., 2018a).

Três opções diferentes podem ser consideradas na definição da unidade de análise: (a) sobreposição do objeto na imagem: são gerados objetos na segmentação de uma das imagens da série temporal e então é feita a comparação com outras por sobreposição simples; (b) comparação de objetos na imagem: os objetos são gerados pela segmentação de cada imagem nas séries temporal independentemente e (c) objeto de imagem multitemporal: os objetos de imagens são gerados através da segmentação da série temporal juntas (TEWKESBURY et al., 2015). Para este estudo foi aplicado a segmentação por sobreposição dos objetos na imagem, foi realizada a aplicação do algoritmo multiresolution segmentation na primeira imagem de análise e os objetos foram sobrepostos na imagem NDVI diferença.

3.4 Análise espectral

O comportamento espectral dos objetos geográficos na área de estudo foi analisado em todas as bandas espectrais dos sensores OLI (TABELA 1). Para análise espectral das classes de vegetação foram calculados a média e o desvio padrão espectral do NDVI de cada objeto no período analisado.

Tabela 1 – Características espectrais do sensor *Operational Land Imager* (OLI).

Banda	Nome	Comp. de onda (µm)	Sigla
1	Azul	0,45 – 0,52	B1
2	Verde	0,52 – 0,60	B2
3	Vermelho	0,63 – 0,69	B3
4	Infravermelho próximo	0,76 – 0,90	B4
5	Infravermelho de ondas curtas	1,55 – 1,75	B5
7	Infravermelho de ondas curtas	2,08 – 2,35	B7

Fonte: NASA (2010).

3.5 Análise espacial

A variabilidade espacial da cobertura do solo pode ser caracterizada a partir de índices extraídos do semivariograma experimental para cada objeto geográfico de estudo. Para gerar os índices propostos por Balaguer et al. (2010) foi utilizado o software FETEX 2.0 (RUIZ et al., 2011), uma ferramenta para análise de imagem. Os índices do semivariograma permitem uma descrição mais detalhada da forma do semivariograma experimental e dos padrões espaciais que fornecem informações usadas na classificação das imagens (SILVEIRA et al., 2017a).

Para extração dos índices é necessário que o usuário defina dois parâmetros: o número de lags que foi definido em 20 e lag distance (30), que são estabelecidos de acordo com a resolução espacial da imagem *Landsat*. Os índices calculados são definidos e categorizados conforme sua posição na curva do semivariograma experimental: parâmetros próximos a origem do semivariograma e parâmetros até o primeiro máximo (TABELA 2).

Tabela 2 – Índices do semivariograma descritos por Balaguer et al. (2010) agrupados em: (A) Parâmetros do semivariograma próximos à origem; (B) Parâmetros usando valores até o primeiro máximo.

Grupo	Descrição	Índices
A	Razão entre a variância total e a semivariância na primeira distância	$RVF = \frac{\text{variância}}{\gamma_i}$
A	Derivada próxima da origem	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
B	Primeiro valor máximo de distância	$FML = h_{\text{máx}_1}$
B	Média dos valores de semivariância até o primeiro máximo	$MFM = \frac{1}{\text{máx}_1} \sum \gamma_i$
B	Variância dos valores de semivariância	$VFM = \frac{1}{\text{máx}_1} (\gamma_i - \gamma)^2$
B	Razão entre a semivariância no primeiro máximo e a média do semivariograma até esse máximo	$RMM = \frac{\gamma_{\text{máx}_1}}{\gamma_{\text{máx}_1}^{\text{mean}}}$
B	Diferença de segunda ordem entre o primeiro intervalo e o primeiro máximo	$SDF = \gamma_{\text{máx}_1} - 2\gamma_{\frac{\text{máx}_1}{2}} + \gamma_2$

Fonte: Adaptado de Balaguer et al. (2010).

3.6 Detecção de mudanças

As amostras representativas de mudança (desmatamentos e queimadas) e não mudança (sazonalidade) na paisagem foram selecionadas manualmente, através da interpretação visual das imagens *Landsat 8*. A detecção de mudanças baseada em objetos foi realizada em três conjuntos de atributos: utilizando apenas atributos espectrais, utilizando apenas atributos espaciais e usando a combinação dos dois atributos.

Um conjunto de objetos foi amostrado na área de estudo, sendo que a intensidade amostral sofreu variações. Foram aplicados nos três conjuntos de atributos da detecção (espectral, espacial e ambos) quatro níveis de intensidade amostral: 850; 400; 200 e 100 amostras. Do total de objetos amostrados em cada nível de intensidade amostral 70% foram utilizados no treinamento e os outros 30% empregados na validação da detecção.

A etapa de detecção dos objetos que sofreram mudança foi baseada em aprendizagem de máquinas, empregando três tipos de classificadores robustos: *Neural Network Multilayer Perceptron* (MLP), *Support Vector Machine* (SVM) e *Random Forest* (RF). O software *Matlab R2011a* foi utilizado neste processo por conter uma vasta coleção de algoritmos de classificação, revelando facilidade na geração e comparação dos modelos mais adequados à solução do problema (BELGIU; DRĂGU, 2016).

Para calcular o desempenho dos classificadores nas etapas de treino e validação foram contabilizados os acertos da classe de mudança e os acertos para a classe não mudança, depois e o desempenho foi medido pela razão do número de acertos da classe mudança em relação ao número de amostras total e o número de acertos da classe não mudança em relação ao total, em porcentagem.

3.7 Classificadores de inteligência artificial

O algoritmo *Random Forest* faz previsões a partir de árvores de decisão e para gerar as árvores é necessário que o usuário defina o número de variáveis de entrada (*Mtry*) e o número de árvores de decisão (*Ntree*) (BREIMAN, 2001). O parâmetro *Mtry* foi estabelecido em 9, ou seja, todos os atributos espectrais e espaciais selecionados para o estudo e *Ntree* foi definido em 10. A MLP foi a rede empregada no estudo, sua arquitetura foi composta por uma camada de entrada (atributos), uma camada intermediária e uma camada de saída. O algoritmo de treinamento utilizado na *Neural Network Multilayer Perceptron* para o ajuste da detecção foi o *backpropagation* (RIEDMILLER; BRAUN, 1993), especificamente o *Bayesian*

regularization backpropagation (função do software *MatLab trainbr*). A função de ativação utilizada foi tangente hiperbólica, usando uma taxa de aprendizagem de $0,1 \times 10^{-10}$ e o número máximo de 30 épocas.

O critério de parada do processo de treinamento da rede foi a regularização bayesiana que combina a minimização do erro quadrático mínimo e a modificação da combinação linear, para que no final do treinamento, o resultado seja capaz de boa generalização. Na implementação do algoritmo *Support Vector Machine* foi usada a função de base radial (RBF) como a função Kernel para detecção das mudanças. Toda a detecção de mudanças usando inteligência artificial foi realizada no software *Matlab* R2011a, empregando todos os algoritmos em cada um dos conjuntos de atributos e para cada intensidade amostral.

3.8 Análise da acurácia da detecção

A análise de acurácia é um processo de validação aplicado a fim de se incrementar a qualidade e permitir a possibilidade de comparação entre técnicas, algoritmos ou intérpretes entre diferentes produtos (CONGALTON; GREEN, 2019). Para a comparação da respectiva detecção, os polígonos de validação previamente selecionados como de verdade de campo foram usados na elaboração de uma matriz de confusão, utilizada para obtenção da acurácia global, erros de omissão providos da acurácia do produtor, erros de inclusão providos da acurácia do usuário, e somatório dos erros.

Para efeito de validação foi utilizado outra medida descritiva, o índice Kappa, um coeficiente de concordâncias para escalas nominais, variando de zero onde a concordância é nula a um com a concordância perfeita (FOODY, 2002). Como forma de validar visualmente a detecção de mudanças realizada, a base foi confrontada com a conferição manual utilizando imagens de alta resolução espacial *Rapideye*.

3.9 Análise da abrangência da detecção

A análise da abrangência foi realizada considerando uma cena *Landsat* contígua (219/71) à cena da área de estudo (219/70). A imagem *Landsat* 8 OLI foi adquirida e processada conforme citado no item 3.2.

Para a aplicação da detecção de mudanças proposta para a imagem de abrangência foi implementado o algoritmo mais robusto em relação aos dados e a melhor intensidade amostral obtidos na primeira parte do estudo. Sendo que o algoritmo foi aplicado usando as mesmas

amostras de treinamento e os mesmos parâmetros do algoritmo treinado para a cena da área de estudo.

A detecção foi validada da mesma maneira citada anteriormente, via coleta de amostras de validação, análise de acurácia, índice Kappa e análise visual de imagens de alta resolução espacial.

4 RESULTADOS E DISCUSSÃO

4.1 Análise dos conjuntos de atributos

Para analisar qual tipo de atributo (espacial ou espectral) e ou a combinação deles melhor diferenciam as mudanças sazonais foram empregados nos três conjuntos de atributos: espectral, espacial e ambos a detecção usando os três classificadores.

O primeiro conjunto da detecção de mudanças empregou apenas atributos espectrais: média e desvio padrão espectral da imagem NDVI *Landsat* 8. Foi utilizado o Discriminante Linear de *Fisher* para selecionar os atributos mais relevantes na detecção (THEODORIS; KOUTROUMBAS, 2008), onde a média espectral correspondia a 90 % dos acertos.

No segundo conjunto foram empregados apenas atributos espaciais na detecção de mudanças, em que, baseado nas análises de *Fisher*, os melhores atributos espaciais foram: Razão entre a variância total e a semivariância na primeira distância (RVF), média do semivariograma (MFM) e Razão entre a semivariância no primeiro máximo e a média do semivariograma até esse máximo (RMM).

No terceiro conjunto de atributos foi empregado o método de junção do espectral e espacial para detecção. Os atributos selecionados pelo Discriminante Linear de *Fisher* foram: a média espectral, RVF, MFM, variância dos valores de semivariância (VFM) e RMM.

Nas detecções baseadas nos atributos espectrais, espaciais e na combinação deles foram implementados *Neural Network Multilayer Perceptron* (MLP), *Support Vector Machine* (SVM) e *Random Forest* (RF), considerando a primeira intensidade amostral testada no estudo de 850 amostras (TABELA 3).

Posteriormente as análises de desempenho do treinamento e validação foi testada a capacidade de generalização da detecção para toda a imagem *Landsat* 8, precisamente para o restante dos 28.340 objetos que não foram utilizados no treino e validação. E dessa forma, pode ser calculada a acuracidade global da detecção de mudanças (TABELA 4).

Tabela 3 – Desempenho dos classificadores nos conjuntos de atributos da detecção de mudanças; em que MD (classe mudança) e NM (classe não mudança).

Média espectral	MLP	SVM	RF
Desempenho Treinamento MD (%)	1,33	0,67	1,33
Desempenho Treinamento NM (%)	100	100	100
Desempenho Treinamento Geral (%)	50,67	50,33	50,67
Desempenho Validação MD (%)	1,56	0,78	1,56
Desempenho Validação NM (%)	100	100	100
Desempenho Validação Geral (%)	50,78	50,39	50,78
RVF MFM RMM	MLP	SVM	RF
Desempenho Treinamento MD (%)	66,67	51	95
Desempenho Treinamento NM (%)	66,33	75,67	99,67
Desempenho Treinamento Geral (%)	67	63,33	97,33
Desempenho Validação MD (%)	84,38	78,91	89,06
Desempenho Validação NM (%)	62,3	71,31	90,98
Desempenho Validação Geral (%)	73,34	75,11	90,02
Média RVF MFM VMF RMM	MLP	SVM	RF
Desempenho Treinamento MD (%)	46,33	44,33	99
Desempenho Treinamento NM (%)	74	73,67	97,67
Desempenho Treinamento Geral (%)	60,17	59	98,33
Desempenho Validação MD (%)	66,41	65,62	88,28
Desempenho Validação NM (%)	79,51	77,87	86,89
Desempenho Validação Geral (%)	72,96	71,75	87,58

Fonte: Do autor (2020).

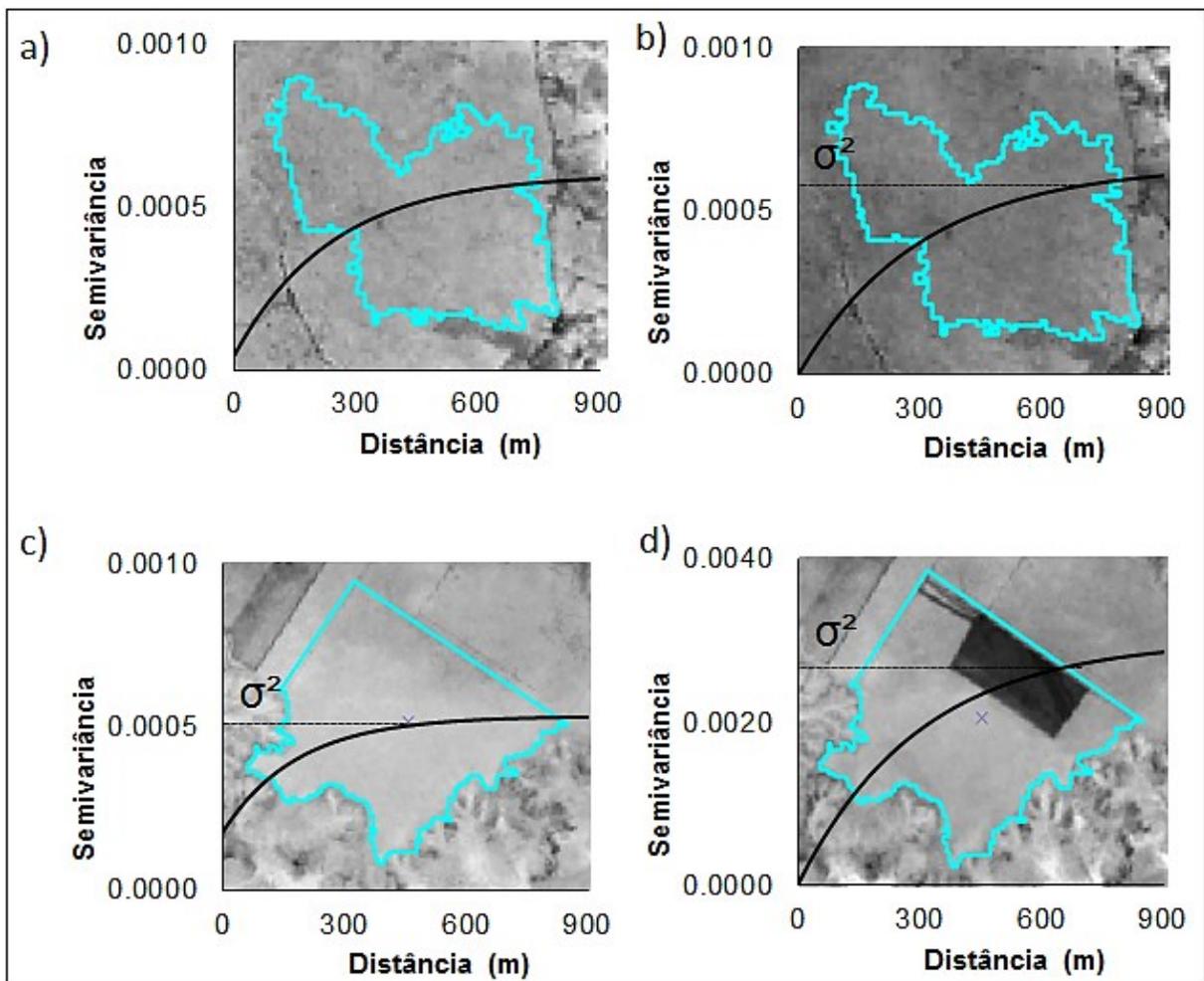
Tabela 4 – Capacidade de generalização dos classificadores nos conjuntos de atributos da detecção de mudanças na imagem *Landsat 8*; em que MD (classe mudança) e NM (classe não mudança).

Média espectral	MLP	SVM	RF
Erro de inclusão MD (%)	100	100	100
Erro de inclusão NM (%)	0	0	0
Erro de omissão MD (%)	0	0	0
Erro de omissão NM (%)	46,69	46,69	46,69
Acurácia global (%)	50,31	50,31	50,31
Índice kappa	0	0	0
RVF MFM RMM	MLP	SVM	RF
Erro de inclusão MD (%)	23,13	33,75	35
Erro de inclusão NM (%)	26,54	20,37	32,1
Erro de omissão MD (%)	25,9	23,74	33,33
Erro de omissão NM (%)	23,72	29,71	33,73
Acurácia global (%)	75,16	72,98	66,46
Índice kappa	0,5	0,46	0,33
Média RVF MFM VMF RMM	MLP	SVM	RF
Erro de inclusão MD (%)	43,13	42,5	42,5
Erro de inclusão NM (%)	19,14	20,37	29,63
Erro de omissão MD (%)	25,41	26,4	34,29
Erro de omissão NM (%)	34,5	34,52	37,36
Acurácia global (%)	68,94	68,63	63,98
Índice kappa	0,38	0,37	0,28

Fonte: Do autor (2020).

Diante das análises apresentadas foi possível constatar que os atributos espaciais derivados dos semivariogramas são capazes de detectar as mudanças no cerrado brasileiro com acuracidade. A partir da análise dos semivariogramas foi possível determinar dois padrões distintos: o primeiro onde a forma e a variabilidade permanecem constantes durante o período analisado; e um segundo padrão onde a forma e a variabilidade aumentam em áreas onde ocorreram desmatamentos e ou queimadas conforme apresentados na Figura 4.

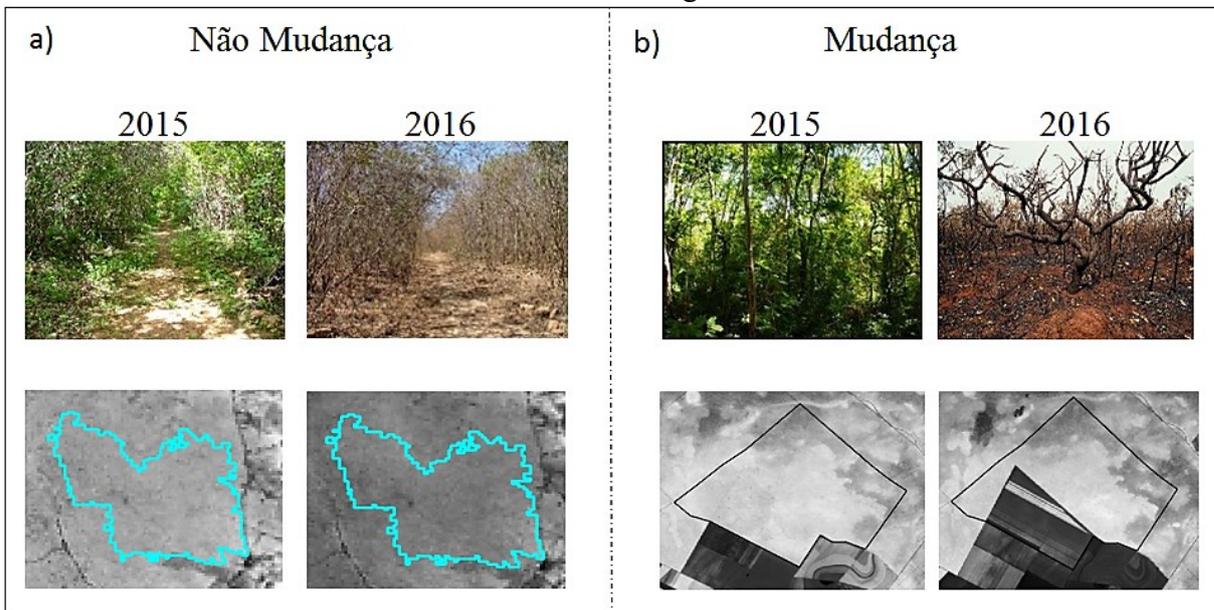
Figura 4 – Exemplos de valores do semivariograma extraído da imagem NDVI diferença (a) segmento de vegetação natural em 2015; (b) segmento com sazonalidade da vegetação em 2016 – valores de semivariância constantes; (c) segmento de vegetação natural em 2015 e (d) segmento com desmatamento em 2016 – aumento no valor de semivariância.



Fonte: Do autor (2020).

Ao analisarmos os atributos espaciais extraídos da imagem diferença, como esperado, apresentaram padrões distintos. A variabilidade espacial da imagem aumentou consideravelmente de áreas de mudança sazonal para áreas que sofreram desmatamentos e queimadas (FIGURA 5).

Figura 5 – Exemplos em (a) áreas afetadas pela sazonalidade que apresentaram padrão homogêneo nos atributos do semivariograma; (b) áreas de mudanças nas quais há um acréscimo nos atributos do semivariograma.



Fonte: Do autor (2020).

Resultados semelhantes foram encontrados por outros estudos. Acerbi Júnior et al. (2015), trabalhando em savanas brasileiras, analisaram parâmetros do semivariograma na detecção de mudanças e concluíram que os parâmetros de variabilidade do semivariograma aumentam na presença de desmatamentos e permanecem constantes em áreas que sofrem sazonalidade. Silveira et al. (2018b) demonstraram que a variabilidade espacial do NDVI, captada pelos atributos do semivariograma não é afetada pela sazonalidade da vegetação e, portanto, pode separar com precisão a antropização da sazonalidade do ambiente. Wu et al. (2015) em estudo de classificação do uso do solo comprovaram que os atributos do semivariograma são um complemento as características espectrais. Assim, os atributos espaciais do semivariograma detectam com precisão os desmatamentos e queimadas ocorridos no cerrado brasileiro, sem serem afetados pelo efeito sazonal.

4.2 Análise da intensidade amostral

A intensidade amostral do estudo sofreu variações, foram aplicados nps três conjuntos de atributos quatro níveis de intensidade amostral respectivamente: 850; 400; 200 e 100 amostras (TABELA 5).

Tabela 5 – Variação da intensidade amostral, total de amostras coletadas, número de amostras de treinamento e teste.

Intensidade Amostral	Total de amostras	Treinamento	Teste
Intensidade amostral 1	850	300	255
Intensidade amostral 2	400	140	120
Intensidade amostral 3	200	70	60
Intensidade amostral 4	100	35	30

Fonte: Do autor (2020).

A intensidade amostral é sempre muito discutida como ponto principal quando o trabalho envolve amostragem, no entanto, poucos estudos se dedicaram a analisar a intensidade amostral em imagens de satélite para de detecção de mudanças.

4.2.1 Análise do desempenho dos classificadores nas diferentes intensidades amostrais

A análise do desempenho dos classificadores procura demonstrar sua eficiência em relação à variação da intensidade amostral, ou seja, analisa a robustez do algoritmo em relação ao número de amostras disponibilizadas para seu treinamento. Na Tabela 6 contém os resultados da acurácia global por classificadores obtidos com a variação da intensidade amostral usando os atributos espaciais como dados de entrada. Pode-se observar que com a diminuição da intensidade amostral a acurácia apresentou uma variação ínfima. Com isso, presume-se que não é o número de amostras de treinamento do algoritmo que irá afetar a acurácia da detecção, mas sim a qualidade das amostras coletadas, ou seja, a homogeneidade de cada amostra coletada para o treinamento do classificador.

Tabela 6 – Acurácia da detecção para os três classificadores em cada uma das intensidades amostrais.

Intensidade Amostral	MLP (%)	SVM (%)	RF (%)
Intensidade amostral 1	75,16	72,98	66,46
Intensidade amostral 2	72,98	74,84	60,87
Intensidade amostral 3	72,36	74,84	58,7
Intensidade amostral 4	72,36	71,12	64,6

Fonte: Do autor (2020).

Van Niel, McVicar e Datt (2005), estudando a relação entre o tamanho das amostras de treinamento e a dimensionalidade dos dados em classificação de arroz na Austrália, demonstraram que o número de amostras de treinamento necessárias para um problema simples de classificação é muito menor que o definido pela regra geral (n -to- p), e portanto, a regra não é universal.

Estudos do efeito da intensidade amostral da classificação de genomas por imagens concluíram que a precisão do preditor genômico é determinada por uma interação da intensidade amostral e nível de complexidade do problema (POPOVICI et al., 2010). Resultados semelhantes ao apresentado nesse estudo em que, a homogeneidade das amostras de vegetação e desmatamentos/queimadas influencia diretamente na acurácia do classificador.

MLP apresentou os maiores valores de variação da intensidade amostral, assim como os maiores valores de acurácia dos resultados comparada aos demais algoritmos. O classificador SVM também não sofreu grandes variações de desempenho na acurácia devido à variação da intensidade amostral, entretanto, o RF além de não conseguir bom desempenho na generalização dos resultados apresentou instabilidade na variação da quantidade de amostras oferecidas ao treinamento e teste.

Ambos os algoritmos utilizados nesse estudo são considerados superiores e com boa capacidade de generalização (ENGELBRECHT, 2007), nesse caso específico de detecção de mudanças em áreas sazonais a MLP apresentou melhor desempenho na acurácia e maior estabilidade na variação da intensidade amostral.

4.3 Análise da acurácia

O algoritmo *Bayesian regularization backpropagation* da rede MLP gerou o mapa final da detecção (FIGURA 6) a partir da intensidade amostral 1 – 850 amostras. A matriz de confusão mostra os resultados em termos de precisão geral, do usuário e do produtor para cada uma das classes analisadas (TABELA 7).

Figura 6 – Mapa final em graduação de cores variando de verde (áreas de não mudança e sazonais) a vermelho (áreas de desmatamento e queimadas). Em (a) Polígono de uma área em 2015; (b) O mesmo polígono no ano de 2016 e (c) A segmentação do polígono desmatado.

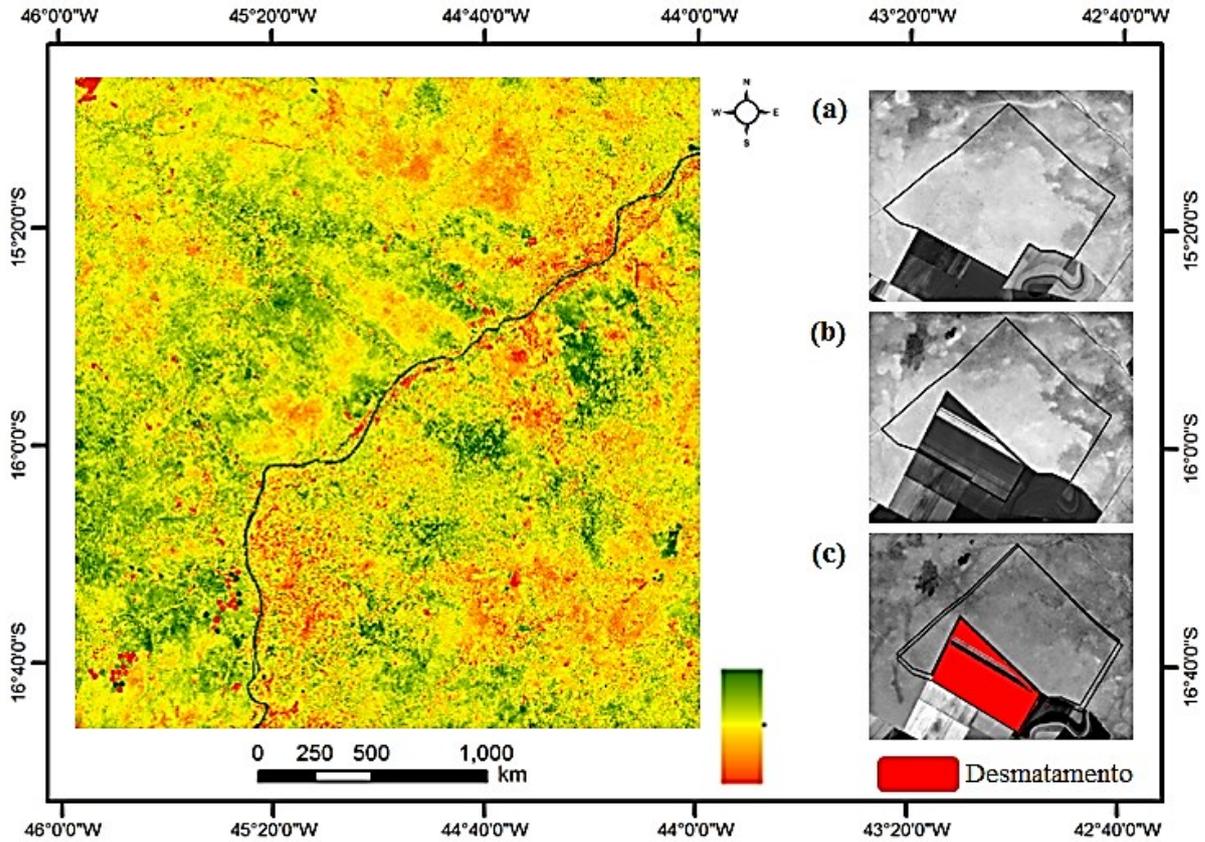


Tabela 7 – Acurácia da detecção para a área de estudo.

Matriz de Confusão	Mudança	Não Mudança
Acurácia do Produtor (%)	25,90	23,72
Acurácia do Usuário (%)	23,13	26,54
Acurácia Global (%)	75,16	
Índice Kappa	0,51	

Fonte: Do autor (2020).

A acurácia do modelo MPL aplicado foi de 73,34% para as amostras de validação (TABELA 3), a aplicação do modelo para o restante dos objetos da imagem *Landsat* garantiu a boa performance do classificador obtendo uma acurácia global de 75,16%.

Silveira et al. (2018a) classificando mudanças no uso do solo na Savana brasileira encontraram o resultado de 95% de acurácia global usando o parâmetro *Sill* do

semivariograma. O parâmetro de variabilidade global do semivariograma *Sill* é capaz de distinguir as classes de mudança e não mudança com precisão (SILVEIRA et al., 2018b) e os parâmetros derivados da curva do semivariograma também fornecem informações representativas da variação global da imagem NDVI.

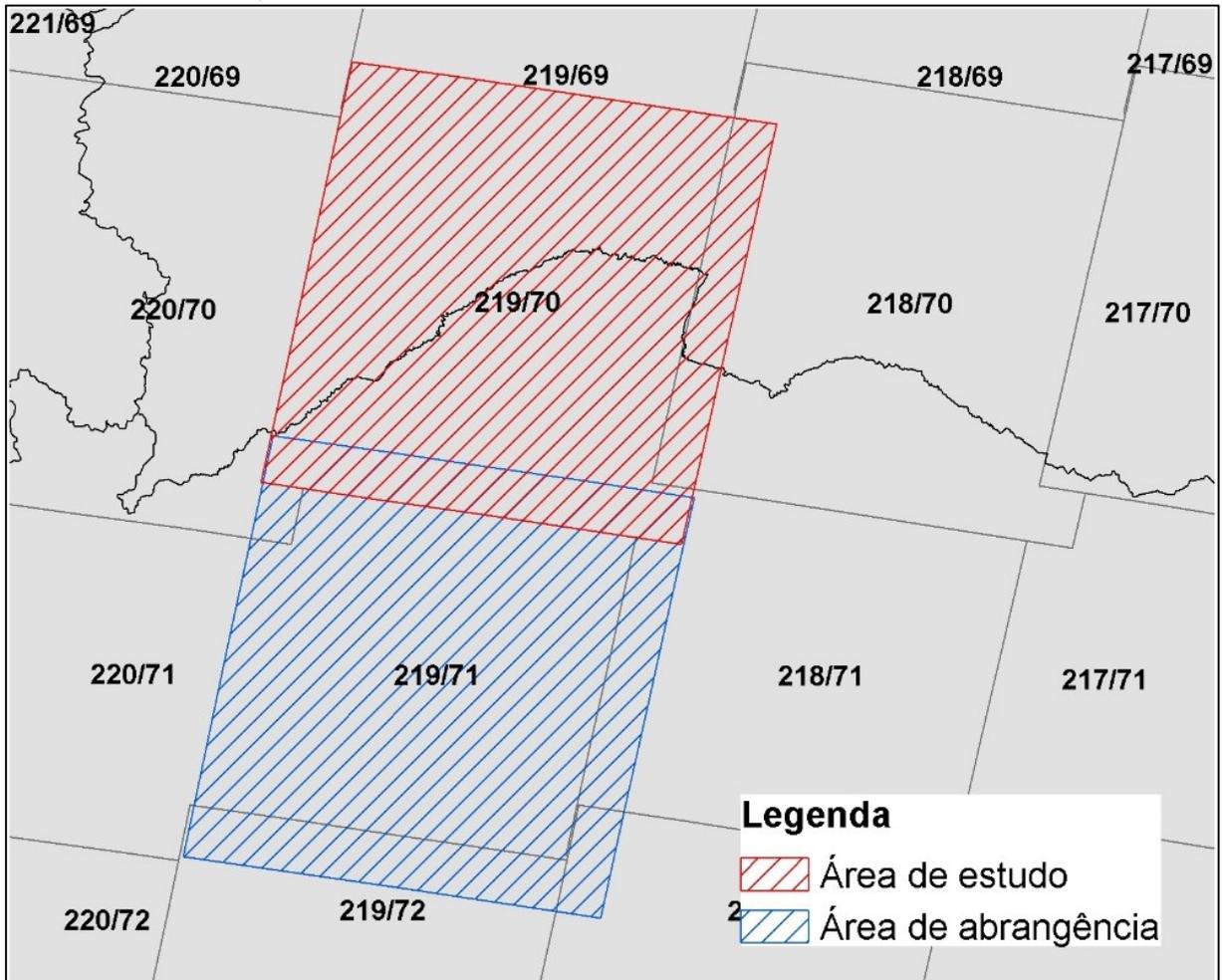
Pereira et al. (2019), em estudo de classificação na savana brasileira, testaram os atributos espectrais e espaciais na detecção de mudanças e encontraram uma acurácia global de 97,73% de generalização do classificador *Random Forest*, portanto, os resultados desse estudo são complementares.

4.4 Análise da abrangência da detecção

A metodologia proposta foi implementada numa cena *Landsat* 8 próxima a área de estudo (FIGURA 7). O algoritmo MLP foi empregado para a detecção de mudanças. Para análise do raio de abrangência foi utilizado os mesmos parâmetros do algoritmo treinado para a área de estudo, inclusive as mesmas amostras de treinamento e validação. Posteriormente, foram coletadas amostras de validação e em seguida gerada uma matriz de confusão (TABELA 8) apresentando os erros de omissão providos da acurácia do produtor, erros de inclusão providos da acurácia do usuário, e somatório dos erros.

A acurácia da detecção na área de abrangência não apresentou resultados semelhantes ao encontrado na área de estudo e isso se deve em parte a complexidade do problema da detecção apresentado. As áreas de Cerrado são áreas altamente sazonais, as imagens de satélite sofrem variação de coloração nas bandas em função da densidade do dossel e os atributos aqui propostos foram eficientes em captar as mudanças desconsiderando essa sazonalidade em uma mesma cena *Landsat*, no entanto, ao aplicar o algoritmo treinado em uma cena *Landsat* numa área contígua, com padrões sazonais diferentes, os resultados da detecção não são precisos. Portanto, a utilização da detecção de mudanças utilizando o algoritmo treinado com amostras em outra imagem está sujeito a erros e a mapas menos precisos.

Figura 7 – Área de abrangência. Localização da área de abrangência no grid de imagens *Landsat*, no estado de Minas Gerais - MG.



Fonte: Do autor (2020).

Tabela 8 – Acurácia da detecção para a área de abrangência.

Matriz de Confusão	Mudança	Não Mudança
Acurácia do Produtor (%)	36,27	52
Acurácia do Usuário (%)	47,27	40,74
Acurácia Global (%)		56,0
Índice Kappa		0,12

Fonte: Do autor (2020).

5 CONCLUSÃO

Os atributos espaciais fornecem bons resultados para a detecção de mudanças no Cerrado brasileiro. Os atributos espaciais derivados dos objetos na imagem NDVI não foram afetados pelo período da sazonalidade do ambiente, separando as classes de não mudança dos desmatamentos e queimadas ocorridos entre 2015 e 2016.

A homogeneidade das amostras foi considerada primordial na detecção de mudanças e o classificador mais robusto em relação à variação de intensidade amostral e com boa capacidade de generalização dos resultados foi a Rede *Multilayer Perceptron* (MLP).

A análise da abrangência determinou que a metodologia proposta sofre limitações quanto a aplicação do algoritmo treinado numa cena Landsat ser aplicado em outra para solução de problemas complexos sem a necessidade da coleta de novas amostras de treinamento.

REFERÊNCIAS

- ACERBI JÚNIOR, F. W. et al. Change Detection in Brazilian Savannas using Semivariograms derived from NDVI Images. **Ciências Agropecuárias**, Lavras, v. 39, n. 2, p. 103-109, 2015.
- AQUINO, C. M. S.; OLIVEIRA, J. G. B. Estudo da Dinâmica do Índice de Vegetação por Diferença Normalizada (NDVI) no núcleo de São Raimundo Nonato - PI. **GEOUSP – Espaço e Tempo**, São Paulo, n. 31, p. 157-168, 2012.
- BAATZ, M.; SCHÄPE, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: STROBL, J.; BLASCHKE, T.; GRIESEBNER, G. (Eds.). **Angewandte Geographische Informations Verarbeitung XII**. Heidelberg: Wichmann-Verlag, 2000. p. 12-23.
- BALAGUER, A. et al. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers & Geosciences**, Oxford, v. 36, p. 231-240, 2010.
- BAYMA, A. P.; SANO, E. E. Séries Temporais de Índices de Vegetação (NDVI E EVI) do sensor Modis para detecção de desmatamentos no Bioma Cerrado. **BCG – Boletim de Ciências Geodésicas**, Curitiba, v. 21, n. 4, p. 797-813, 2015.
- BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 114, p. 24-31, 2016.
- BLASCHKE, T. Object based image analysis for remote sensing. **Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 65, p. 2-16, 2010.
- BREIMAN, L. Random Forests. **Machine Learning**, Dordrecht, v. 45, p. 5-32, 2001.
- CARVALHO, L. M. T.; DUTRA, G. C. Fundamentos de sensoriamento remoto. In: CARVALHO, L. M. T. (Ed.). **Sistemas de informações geográficas e sensoriamento remoto dos recursos florestais**. Lavras: UFLA, 2008.
- CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 3. ed. New York: CRC Press, 2019. 346 p.
- COSTANTINI, M. L. et al. NDVI spatial pattern and the potential fragility of mixed forested áreas in volcanic lake watersheds. **Forest Ecology and Management**, Amsterdam, v. 285, p. 133-141, 2012.
- CURRAN, P. J. The Semivariogram in Remote Sensing: An Introduction. **Remote Sensing of Environment**, New York, v. 24, p. 493-507, 1988.
- DESCLÉE, B.; BOGAERT, P.; DEFOURNY, P. Forest change detection by statistical object-based method. **Remote Sensing of Environment**, New York, v. 102, p. 1-11, 2006.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 p. Tese (Doutorado em Agronomia) – Universidade de São Paulo, Piracicaba, 2000.

ENGELBRECHT, A. P. **Computacional intelligence: an introduction**. Chichester: Wiley, 2007. 628 p.

FERREIRA, G. H. S. **Aplicação do algoritmo para detecção de mudanças em dados contínuos do sensor Modis na Bacia Hidrográfica do Alto Xingu-MT como auxílio à gestão territorial**. 2015. 35 p. Monografia (Bacharelado em Geografia) – Universidade de Brasília, Brasília, 2015.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, New York, v. 80, p. 185-201, 2002.

FRANÇOSO, R. et al. Habitat loss the effectiveness of protected áreas in the Cerrado biodiversity hotspot. **Natureza & Conservação**, Rio de Janeiro, v. 13, n. 1, p. 35-40, 2015.

GALEANA-PIZAÑA, J. M. et al. Modeling the spatial distribution of above-ground carbon in Mexican coniferous forests using remote sensing and a geostatistical approach. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 30, p. 179-189, 2014.

GARCÍA-GUTIÉRREZ, J. et al. An evolutionary-weighted majority voting and support vector machines applied to contextual classification of LiDAR and imagery data fusion. **Neurocomputing**, London, v. 163, n. 2, p. 17-24, 2015.

GENELETTI, D.; GORTE, B. G. H. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. **International Journal of Remote Sensing**, Oxon, v. 24, n. 6, p. 1273-1286, 2003.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de Imagens Digitais**. São Paulo: Edgard Blücher, 2000. 528 p.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2001. 900 p.

HEYDARI, S. S.; MOUNTRAKIS, G. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. **Remote Sensing of Environment**, New York, v. 204, p. 648-658, 2018.

HILL, M. J. et al. Relationships between vegetation indices, fractional cover retrievals and the structure and composition of Brazilian Cerrado natural vegetation. **International Journal of Remote Sensing**, Oxon, v. 38, n. 3, p. 874-905, 2016.

HOEKSTRA, J. M. et al. Confronting a biome crisis: global disparities of habitat loss and protection. **Ecology Letters**, Hoboken, v. 8, p. 23-29, 2005.

HUSSAIN, M. et al. Change detection from remotely sensed images: From pixel-based to object-based approaches. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 80, p. 91-106, 2013.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, 1989. 561 p.

KLINK, C. A.; MACHADO, R. B. Conservation of the Brazilian Cerrado. **Conservation Biology**, Hoboken, v. 19, n. 3, p. 707-713, 2005.

LU, D.; WENG, Q. A survey of image classification methods and techniques for improving classification performance. **International Journal of Remote Sensing**, Oxon, v. 28, n. 5, p. 823-870, 2007.

LU, M. et al. Land cover change detection by integrating object-based data blending model of Landsat and MODIS. **Remote Sensing of Environment**, New York, v. 184, p. 374-386, 2016.

MELLO, J. M. **Geoestatística aplicada ao Inventário Florestal**. 2004. 110 p. Tese. (Doutorado em Recursos Florestais) – Universidade de São Paulo, Piracicaba, 2004.

MMA. Ministério do Meio Ambiente. **Monitoramento do Desmatamento nos Biomas Brasileiros Por Satélite: Bioma Cerrado 2009-2010**. 2011. Disponível em: <https://www.mma.gov.br/estruturas/sbf_chm_rbbio/_arquivos/relatoriofinal_cerrado_2010_f inal_72_1.pdf>. Acesso em: 25 jan. 2020.

_____. **Monitoramento do Desmatamento nos Biomas Brasileiros Por Satélite: Cerrado 2010-2011**. 2015. Disponível em: <<https://webcache.googleusercontent.com/search?q=cache:GcOubhwua-gJ:https://www.mma.gov.br/publicacoes/biomas/category/62-cerrado.html%3Fdownload%3D1138:monitoramento-do-desmatamento-dos-biomas-brasileiros-por-sat%25C3%25A9lite-cerrado-cerrado-2010-2011+%&cd=2&hl=pt-BR&ct=clnk&gl=br>>. Acesso em: 25 jan. 2020.

MUI, A.; HE, Y.; WENG, Q. An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 109, p. 30-46, 2015.

NASA. **The Landsat program**. 2010. Disponível em: <<http://landsat.gsfc.nasa.gov/>>. Acesso em: 11 nov. 2019.

NAVULUR, K. **Multispectral image analysis using the object-oriented paradigm**. Davis: CRC, 2006. 184 p.

NOGUEIRA, C. H. **Análise de variância com dependência espacial sob uma abordagem geostatística**. 2013. 124 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

OLIVEIRA-FILHO, A.; RATTER, J. A. Vegetation physiognomies and woody flora of the Cerrado Biome. In: OLIVEIRA, S. P.; MARQUIS, R. J. (Eds.). **The Cerrados of Brazil: ecology and natural history of a neotropical savanna**. New York: Columbia University Press, 2002. p. 91-120.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences Discussions**, European Geosciences Union, v. 4, n. 2, p. 439-473, 2007.

PEREIRA, E. A. et al. Spatial and spectral remote sensing features to detect deforestation in Brazilian Savannas. **Advances in Forestry Science**, Cuiabá, v. 6, n. 4, p. 775-782, 2019.

POPOVICI, V. et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. **Breast Cancer Research**, London, v. 12, n. 1, p. R5, 2010.

RIEDMILLER, M.; BRAUN, H. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. In: IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1., 1993, San Francisco. **Anais...** San Francisco: IEEE, 1993. p. 586-591.

RUIZ, L. A. et al. A feature extraction software tool for agricultural object-based image analysis. **Computers and Electronics in Agriculture**, Oxford, v. 76, p. 284-296, 2011.

SCHWIEDER, M. et al. Mapping Brazilian savanna vegetation gradients with Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 52, p. 361-370, 2016.

SHALABY, A.; TATEISHI, R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. **Applied Geography**, Oxon, v. 27, p. 28-41, 2007.

SILVEIRA, E. M. O. et al. Characterizing Landscape Spatial Heterogeneity using Semivariogram Parameters Derived from NDVI Images. **Cerne**, Lavras, v. 23, n. 4, p. 413-422, 2017a.

_____. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Florida, v. 11, n. 3, p. 036004, 2017b.

_____. Object-based change detection using semivariogram indices derived from NDVI images: The environmental disaster in Mariana, Brazil. **Ciência e Agrotecnologia**, Lavras, v. 41, n. 5, p. 554-564, 2017c.

_____. Using Spatial Features to Reduce the Impact of Seasonality for Detecting Tropical Forest Changes from Landsat Time Series. **Remote Sensing**, Basel, v. 10, n. 6, p. 808, 2018a.

_____. Object-based land-cover change detection applied to Brazilian seasonal savannas using geostatistical features. **International Journal of Remote Sensing**, Oxon, v. 39, n. 8, p. 2597-2619, 2018b.

TEWKESBURY, A. P. et al. A critical synthesis of remotely sensed optical image change detection techniques. **Remote Sensing of Environment**, New York, v. 160, p. 1-14, 2015.

THEODORIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. San Diego: Academic Press, 2008. 961 p.

VAN NIEL, T. G.; MCVICAR, T. R.; DATT, B. On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. **Remote Sensing of Environment**, New York, v. 98, p. 468-480, 2005.

VIANA, H. et al. Estimation of crown biomass of Pinus pinaster stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. **Ecological Modelling**, Amsterdam, v. 226, p. 22-35, 2012.

WU, X. et al. Evaluation of semivariogram features for object-based image classification. **Geo-spatial Information Science**, Wuhan, v. 18, n. 4, p. 159-170, 2015.

YAN, G. et al. Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China. **International Journal of Remote Sensing**, Oxon, v. 27, n. 18, p. 4039-4055, 2006.

ZHENG, C.; WANG, N.; CUI, J. Hyperspectral Image Classification With Small Training Sample Size Using Superpixel-Guided Training Sample Enlargement. **IEEE Transactions on Geoscience and Remote Sensing**, Piscataway, v. 57, n. 10, p. 7307-7316, 2019.