
Metodologia de análise de imagens baseada em objetos geográficos (GEOBIA) utilizando RPAS (drone) com sensor RGB

Estrabão
2021(2):41–85
©The Author(s) 2021
DOI: 10.53455/re.v2i1.5



Vinicius Paiva Gonçalves¹

Abstract

O presente trabalho apresenta um método para o mapeamento de vegetação, por um processo de classificação por regiões geográficas, denominado GEOBIA (Geographic Object-Based Image Analysis) considerado adequado para classificar imagens de muito alta resolução (very high resolution – VHR). É possível executar o procedimento com qualquer equipamento que disponha de um sensor RGB de boa qualidade e permita execução de aplicativos para plano de voo. O método foi desenvolvido com base em softwares de código aberto (open source) para evitar custos com licenças, em todas as etapas, desde a captação das imagens, elaboração de produtos cartográficos, processamento da classificação por regiões e conclusão mediante cálculos de áreas. O estudo foi aplicado em quatro áreas de interesse, todas na região da Grande Florianópolis-SC, contendo porções do ecossistema de Formações Pioneiras - Vegetação com Influência Marinha, também denominadas áreas de restinga, cujo principal alvo da classificação foi o mapeamento das áreas invadidas por Pinus sp. O método demonstrou útil para classificação de imagens em geral, podendo ser utilizado no manejo de outras espécies vegetais exóticas, ou até em outras aplicações ambientais.

Keywords

Drone, mapeamento, alta resolução, Pinus sp

Introdução

A partir de 2015 os equipamentos popularmente conhecidos como Drones ficaram mais acessíveis e fáceis de pilotar, embarcando sensores RGB de alta qualidade, disponibilizados na modalidade “prontos para voo”, ou seja, dispensando qualquer tipo de conhecimento técnico em eletrônica ou automação por parte do operador. No mercado agro esses equipamentos passaram a ser utilizados no diagnóstico, e planejamento de ações de manejo em cultivos de diversas espécies, o que demonstra viabilidade de utilização em contexto associado ao monitoramento e controle de espécies vegetação.

Nesse sentido, o presente trabalho apresenta um método para o mapeamento de vegetação, por um processo de classificação por regiões geográficas, denominado GEOBIA (Geographic Object-Based Image Analysis) considerado adequado para classificar imagens de muito alta resolução (very high resolution – VHR). É possível executar o procedimento com qualquer equipamento que disponha de um sensor RGB de boa qualidade e permita execução de aplicativos para plano de voo.

¹ Instituto do Patrimônio Histórico e Artístico Nacional, Florianópolis, Santa Catarina, Brasil

Corresponding author:

Vinicius Paiva Gonçalves, Instituto do Patrimônio Histórico e Artístico Nacional, Florianópolis, Santa Catarina, Brasil
Email: viniciuspvg@gmail.com

O método foi desenvolvido com base em softwares de código aberto (open source) para evitar custos com licenças, em todas as etapas, desde a captação das imagens, elaboração de produtos cartográficos, processamento da classificação por regiões e conclusão mediante cálculos de áreas.

Materiais e métodos

Áreas de estudo Para exemplificação o presente método foi aplicado a quatro áreas de interesse – AOI, conforme Figura 1, para compor o estudo, todas na região da Grande Florianópolis- SC, contendo porções do ecossistema de Formações Pioneiras - Vegetação com Influência Marinha, segundo (Ibge, 2012), também denominadas áreas de restinga, cujo principal alvo da classificação foi o mapeamento das áreas invadidas por *Pinus sp.*

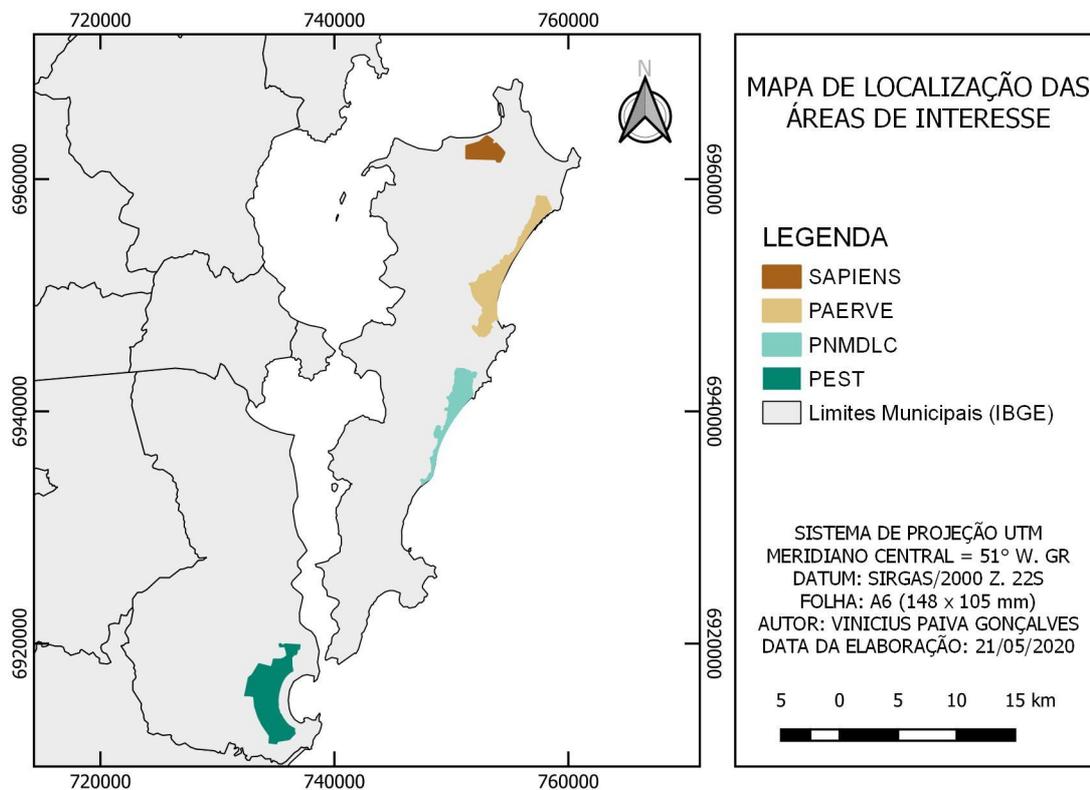


Figure 1. Mapa de localização das áreas de interesse.

Em cada área de interesse os voos foram executados para cobrir 300x300m totalizando 9 hectares por voo.

Coleta e tratamento dos dados

Equipamentos e softwares

As áreas de voo são definidas conforme necessidade. Recomenda-se utilizar o software QGIS (QGIS.org 2021), para gerar arquivos vetoriais em formato kml com as dimensões exatas, conforme localização desejada, utilizando

como referência espacial os mosaicos de imagens de satélite disponíveis no plugin QuickMapServices. Essa etapa deve levar em consideração a proximidade dos alvos em relação aos acessos e local adequado para decolagem.

O formato kml foi escolhido porque é aceito pela plataforma DroneDeploy em seu aplicativo utilizado para criação e execução do plano de voo, com base na localização fornecida, e pode ser utilizado em aplicativos de smartphone para localização em campo. Os voos realizados para aplicação dessa metodologia foram executados com equipamento da Figura 2, multirrotor modelo Inspire 1 (T600) fabricado pela empresa DJI, equipado com uma câmera modelo X3 (FC350) com 12.4Megapixels.



Figure 2. Equipamento utilizado para coleta das imagens

Planejamento e execução dos voos O planejamento dos voos deve ser realizado pela plataforma DroneDeploy que, embora proprietária, possibilita a execução dessas tarefas de forma livre e gratuita, tornando uma ótima alternativa a esse propósito, já que possui interface de gráfica amigável e aceita importação da área em formato vetorial (kml). Embora não seja obrigatório, recomenda-se que a definição da área de voo seja feita previamente, em ambiente SIG (QGIS) para ajustes finos em relação à melhor cobertura dos alvos e otimização das baterias. A Figura 3 apresenta o ambiente da plataforma para planejamento do voo. Em verde o trajeto planejado.

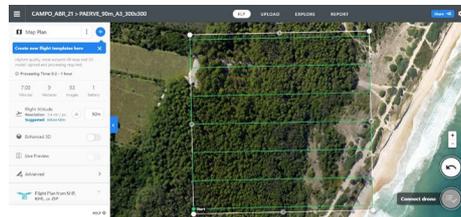


Figure 3. Plataforma utilizada para planejamento de voo.

Podem ser utilizados os parâmetros padrão sugeridos pela plataforma para o equipamento utilizado, que no caso indicou sobreposição frontal de 75% e lateral de 65%, para uma altura de voo definida em 90m. Para o equipamento essa configuração permite a realização de todo o voo com apenas uma bateria do tipo TB48, com o equipamento retornando ao ponto de decolagem com mais de 30% de carga restante, como margem de segurança na operação. Com essas características definidas para esse equipamento, cada voo resulta na coleta de aproximadamente 150 imagens.

Elaboração dos produtos cartográficos

A etapa de elaboração de produtos cartográficos pode ser executada com a plataforma open source OpenDroneMap – ODM (OpenDroneMap Authors 2020). Essa plataforma utiliza modernos algoritmos de fotogrametria e permite a construção dos produtos cartográficos base para a execução do trabalho, que são o ortomosaico, o modelo digital de superfície – MDS, e o modelo digital do terreno (MDT). Sugere-se que o processamento seja realizado com as configurações padrão do software para geração desses três produtos, com Ground Sample Distance – GSD definido no valor de 5cm.

Valores inferiores são mais custosos para as próximas etapas de processamento, portanto deve ser avaliado se o hardware irá suportar o processamento.

O ortomosaico gerado a partir do ODM possui quatro bandas, as bandas vermelho, verde e azul (RGB) e uma banda alfa com valor de 255 bytes, ou seja, completamente transparente que serve para evitar que imagens de formato irregular não apareçam com bordas pretas em softwares de SIG. Mas, essa banda não é útil para classificação, motivo pelo qual recomenda-se sua remoção pelo algoritmo “Reorganizar bandas” do GDAL, implementado no QGIS.

Para formar o raster que será classificado, é necessária a adição de uma banda formada por um índice de vegetação calculado a partir das bandas espectrais. Em geral, para trabalhos com vegetação é comum a utilização do Red Normalized Difference Vegetation Index, Calibrated NDVI - CDVI, conforme utilizado por (Luca et al., 2019), pois esse responde muito bem à biomassa. Porém, para aplicá-lo, seria necessária captação do espectro do infravermelho próximo, banda NIR, indisponível no sensor utilizado. Por esse motivo optou-se por utilizar o índice Visible Atmospherically Resistant Index Green – VARI, um dos índices que utiliza apenas as bandas RGB.

A criação do raster contendo o VARI, pode ser realizada na calculadora raster do QGIS para realização da álgebra de mapas com as bandas vermelho, verde e azul do ortomosaico, de acordo com o disposto no site Index Database (Henrich, Krauss, Götze, & Sandow, 2021), calculado a partir da equação:

$$VVVVVVVV = \frac{GGGGGGGG - VVGGRR}{GGGGGGGG + VVGGRR - BBBBRRGG}$$

A álgebra para geração do índice VARI resulta em um raster com valores flutuantes (float), em uma faixa de valores muito diferente das bandas RGB. Com base no método proposto por Luca et al. (2019), para reduzir o efeito de potenciais outliers na posterior segmentação, essa banda foi discretizada e normalizada para uma faixa comum de valores, variando de 0 a 255, compatível com as bandas RGB. Esse procedimento foi realizado pela Caixa de ferramentas de processamento do QGIS< utilizando os algoritmos Raster Normalization do SAGA GIS e Converter do GDAL, o primeiro para normalizar os valores e o segundo para converter para 8 bits (byte).

Outro raster necessário ao processo, também gerado a partir do QGIS foi o modelo digital de elevação – MDE, obtido pela diferença entre MDS e MDT, ou seja, subtraindo-se do MDS os valores de MDT. Esse procedimento também realizado com SAGA GIS a partir da Caixa de Ferramentas do QGIS, origina o raster com valores de elevação. Após ser normalizado, o MDE, precisa ser amostrado para 5cm, com o algoritmo Resampling do SAGA, pois o procedimento inicial resulta em um raster com tamanho de pixel incompatível com as outras camadas. Para concluir, o raster do MDE também precisa ser discretizado para 8 bits (byte).

GEOBIA

O fluxo da classificação por regiões aqui descrito foi baseado naquele descrito por Luca et al. (2019), com alguns ajustes. Foram utilizados três softwares livres, sendo o principal deles o Orfeo Toolbox – OTB, disponível na Caixa de ferramentas de processamento do QGIS, a partir do qual já seria possível proceder todo o fluxo. Os demais, GEODMA, um plugin do software de SIG Terraview desenvolvido pelo INPE, e o Weka, da Universidade de Waikato, Nova Zelândia, são opcionais, com propósito de refinamento da classificação, visando melhoria na acurácia.

Segmentação

O processo de segmentação é realizado a partir do algoritmo LargeScaleMeanShift do OTB, que executa quatro processos sequenciais, segundo documentação (OTB Development Team 2021), sendo eles MeanShiftSmoothing, LSMSSegmentation, LSMSSmallRegionsMerging por fim LSMVectorization. O primeiro é um algoritmo de suavização de imagem que preserva bordas, o segundo produz uma imagem rotulada onde os pixels vizinhos são agrupados considerando o raio de alcance definido, o terceiro realiza a fusão dos segmentos próximos cujo tamanho em pixels seja muito pequeno, inferior ao tamanho mínimo definido, e o último realiza a vetorização da imagem segmentada, tratada na etapa anterior, gerando um polígono vetorial para cada segmento da imagem (objeto).

O OTB permite utilizar o LargeScaleMeanShift, que realiza todos os processos automaticamente, ou então a execução de cada processo como uma etapa distinta das demais, sendo essa uma decisão que compete ao usuário. Nos testes realizados o LargeScaleMeanShift apresentou bons resultados com todos os processos, motivo pelo qual optou-se por utilizá-lo. A entrada desse algoritmo são as bandas empilhadas (layer stacking) e a saída resulta no arquivo vetorial contendo a segmentação da imagem.

Importante ressaltar que o algoritmo possui dois parâmetros principais que precisam ser informados, “spatial radius(-spatialr)” ou raio espacial, ou seja, a distância máxima para construir o entorno do objeto, a partir da média dos pixels analisados, e range radius (-ranger)” ou raio de alcance, definido como o limite da distância da assinatura espectral entre as feições, expresso em unidades de radiometria, tendo como base a diferença espectral ou distância euclidiana entre os valores de pixels da assinatura espectral (OTB Development Team 2021).

Para a definição dos parâmetros spatial radius e range radius, cujos valores padrão são 5 e 15 respectivamente, sugere-se testar valores sucessivos variando de 5 a 25, com passo de 5, por exemplo, para selecionar visualmente o melhor resultado de segmentação. Para um processamento mais rápido, recomenda-se realizar esses testes em amostras menores, por exemplo, áreas medindo 50x50m, selecionadas no interior das áreas de estudo. Para o equipamento utilizado, os valores mais adequados foram respectivamente 25 e 15 porque resultaram em melhor contorno dos objetos, segundo avaliação visual, conforme Figura 4.

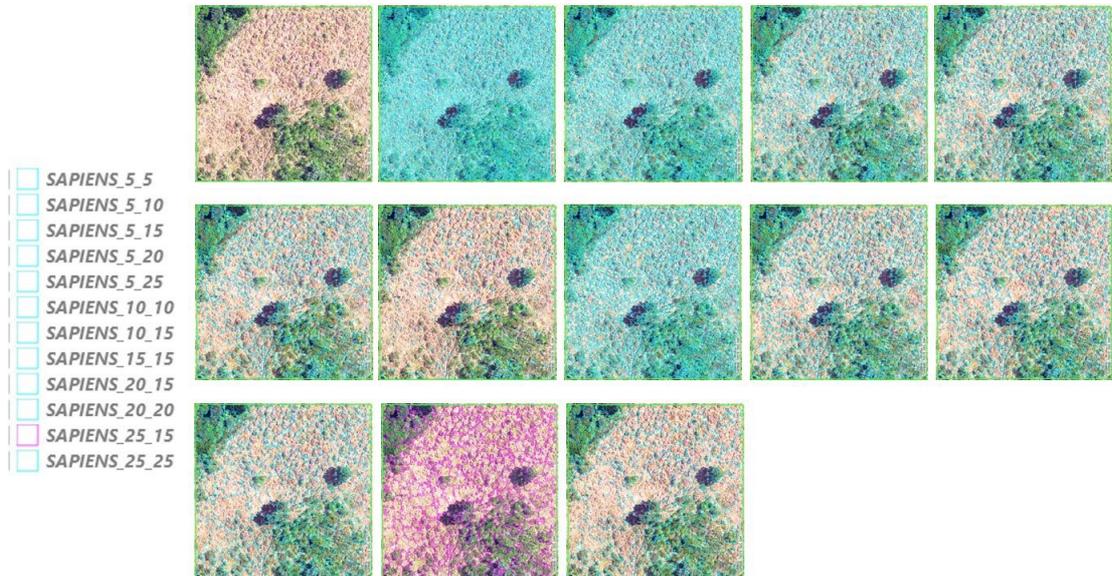


Figure 4. Testes para definição dos parâmetros de segmentação realizados na área do *Sapiens* Parque

A definição dos valores ideais para os parâmetros foi definida visualmente a partir dos resultados dos testes, conforme indicação por [Liau \(2014\)](#) apresentada na Figura 5.

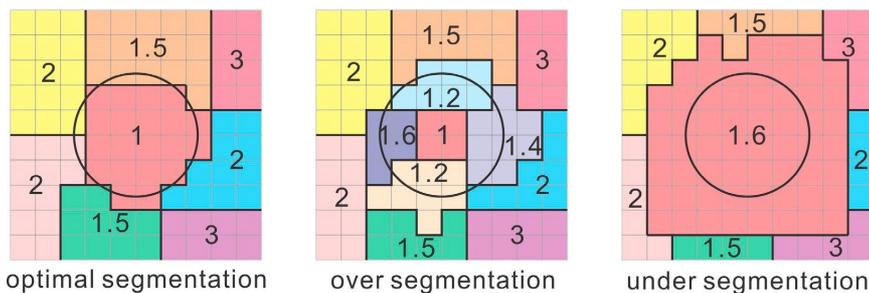


Figure 5. Exemplo de segmentação ideal em comparação a super-segmentação (muitos pequenos polígonos), e sub-segmentação (onde os polígonos são maiores do que as segmentações ideais).

Círculo negro representa o objeto e as regiões coloridas, os resultados da segmentação.

Além do resultado vetorial, ou seja, a geração de segmentos representativos dos objetos da imagem, o algoritmo *LargeScaleMeansShift* também identifica cada segmento com um rótulo (label) que fica registrado na tabela de atributos, o que permite identificação de cada segmento, também referido como objeto. Além do rótulo, o algoritmo ainda computa algumas estatísticas, o número de pixels, a média e a variância.

Extração de atributos

A extração de atributos consiste no cálculo das estatísticas associadas aos segmentos ou objetos. Através do OTB, na Caixa de ferramentas de processamento do QGIS, foi utilizado o algoritmo *ZonalStatistics*, que calcula as estatísticas zonais número de pixels, valor mínimo, valor máximo, valor médio, desvio padrão e variância, para cada segmento, para cada banda, com exceção da contagem dos pixels, cujo valor é calculado para o segmento e independe da banda.

O algoritmo calcula as estatísticas a partir dos valores do raster com as cinco bandas empilhadas (layerstack), tendo como referência a delimitação dos segmentos da camada vetorial obtida na segmentação pelo *ZonalStatistics* na etapa anterior. Por isso, é fundamental que os segmentos representem adequadamente os objetos da imagem, pois é a partir desse procedimento que as estatísticas calculadas serão registradas na tabela de atributos de um novo arquivo vetorial, correspondendo diretamente aos segmentos que as originaram.

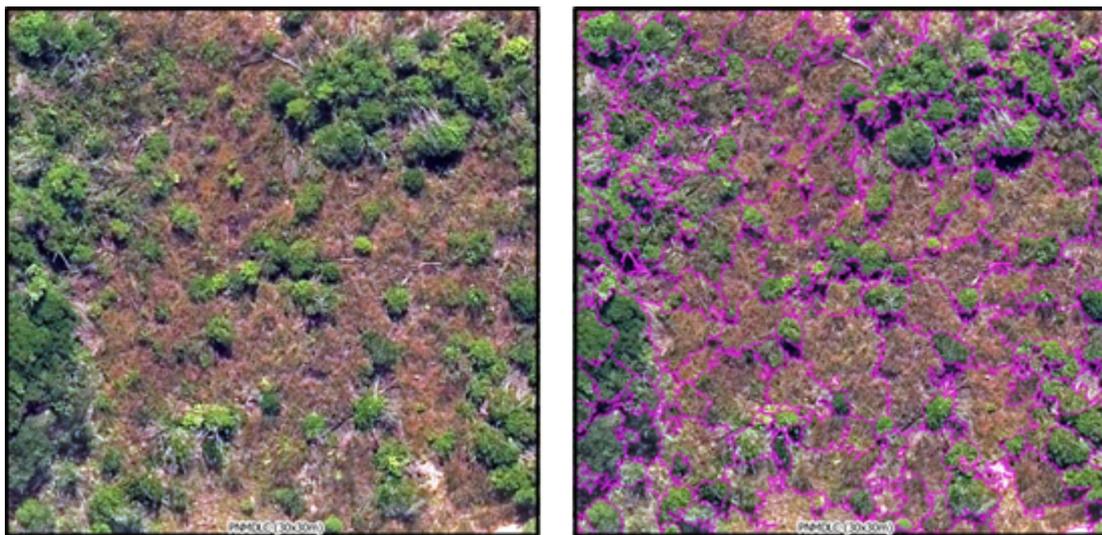


Figure 6. Exemplo do resultado da segmentação na área do PNMDLC utilizando os valores de parâmetros raio espacial 25 e raio de alcance 15, selecionados por avaliação visual.

Como resultado, *ZonalStatistics* computa, média, desvio padrão, mínimo máximo, para cada banda (OTB Development Team 2021), e ainda a contagem de pixels no segmento. A Tabela 1 apresenta os campos computados pelo QGIS/OTB *ZonalStatistics*, em adição àqueles computados anteriormente durante a segmentação pelo OTB *LargeScaleMeanShift*. Cabe ressaltar que a contagem de pixels e média são métricas calculadas em redundância, o que é importante observar durante a seleção de atributos.

No entanto a variabilidade de informações passível de extração para cada segmento não se restringe apenas às estatísticas zonais. Em função disso, recomenda-se o uso do plugin *GeoDMA* do software *TerraView*, desenvolvido pelo Instituto Nacional de Pesquisas Espaciais – INPE. Além das métricas zonais, o *GeoDMA* calcula uma ampla variedade de métricas espectrais e também de forma, conforme descrito em (INPE 2021). Cabe, no entanto, ressaltar que o cálculo dessas estatísticas adicionais não constitui etapa obrigatória no fluxo da *GEOBIA*, porém, com sua utilização permite explorar possibilidades de melhoria na acurácia da classificação.

Table 1. Campos computados pelos algoritmos LargeScaleMeanShift e ZonalStatistics.

CAMPO	ATRIBUTO	ALGORITMO	DESCRIÇÃO
label	Não é atributo da segmentação	LargeScaleMean-shift	Rotulagem dos segmentos, não é atributo para classificação
nbPixels	Número de Pixels	LargeScaleMean-shift	Número de pixels em cada segmento. Idêntico a contagem de pixels
meanB0	Média	LargeScaleMean-shift	Valor médio do nível de cinza dos pixels dentro do objeto
varB0	Variância	LargeScaleMean-shift	Valor de variância do nível de cinza dos pixels dentro do objeto
count	Contagem de pixels	ZonalStatistics	Número de pixels em cada segmento. Idêntico ao Número de Pixels
mean_0	Média	ZonalStatistics	Valor médio do nível de cinza dos pixels dentro do objeto
stdev_0	Desvio Padrão	ZonalStatistics	Valor do desvio padrão do nível de cinza dos pixels dentro do objeto
min_0	Mínimo	ZonalStatistics	Valor máximo do nível de cinza dos pixels dentro do objeto
max_0	Máximo	ZonalStatistics	Valor máximo do nível de cinza dos pixels dentro do objeto

Em negrito estão os que são computados para cada uma das bandas do raster utilizado no processo, nesse caso a banda 0. Fonte OTB Development Team (2021), adaptado.

Após a instalação no TerraView, o plugin GeoDMA apresenta uma interface bastante intuitiva, devendo ser utilizada a função Feature Extraction para extração dos atributos da imagem, também referida como geração das estatísticas ou métricas. A camada raster deve ser a mesma utilizada no OTB (layer stacking) e a camada vetorial o output do algoritmo ZonalStatistics do OTB, mantendo as demais configurações

conforme valores padrão. Ao término, o arquivo resultante do GeoDMA apresentará todas as estatísticas computadas por ambos os softwares. A Tabela 2 (ANEXO) apresenta as métricas calculadas pelo Terraview/GeoDMA.

O vetor resultante da extração de atributos realizada pelo GeoDMA deve ser levado ao QGIS para continuidade do processo. Entretanto, é importante verificar se a codificação do arquivo se mantém. Para evitar possíveis problemas envolvendo codificação, caso necessário o arquivo poderá ser exportado novamente no QGIS, com a codificação correta, a exemplo do UTF8 que preserva os caracteres particulares da língua

portuguesa como de acentuação, dentre outros. Apesar de na tabela ainda não constarem campos do tipo string (textos), recomenda-se que esse cuidado já seja observado.

Outra observação é que as métricas calculadas no TerraView/GeoDMA são comuns ao QGIS/OTB, portanto, a tabela de atributos apresentará alguns dados redundantes, o que deve ser tratado na etapa de seleção de atributos.

Para preparar os arquivos vetoriais para as próximas etapas, é preciso criar dois campos que serão preenchidos ao longo do processo, um numérico, do tipo integer para números inteiros, com comprimento de 10 denominado como “classe” e outro do tipo string para texto, com comprimento sugerido 100, denominado “tipologia”. Esse procedimento pode ser realizado em lote, a partir de um modelo criado com o modelador do QGIS.

Definição das classes

A definição das classes, tipologias ou feições para o processo de classificação supervisionada pressupõe uma avaliação visual das características da área de estudo. É sempre recomendável que a área seja previamente reconhecida in loco, porque isso auxilia a fotointerpretação. Esse reconhecimento pode ser executado de duas maneiras, visitando-se as áreas por solo ou a partir de imagens aéreas captadas por foto filmagem, utilizando o próprio equipamento para realizar sobrevoos panorâmicos com a câmera filmando a paisagem com a câmera entre 45 e 60 graus de inclinação em relação ao nadir. Caso a opção seja visitar a área por solo, recomenda-se a realização de registros fotográficos com geotag, a partir de smartfone equipado com GNSS de navegação. Isso pode ser realizado por diversos aplicativos, com é o caso do Oruxmaps (Figura 8), que embora não seja gratuito, apresenta um baixo custo de aquisição da licença.

Só é viável realizar sobrevoos para o reconhecimento por foto filmagem se o equipamento for um multirrotor. Cabe destacar que esse método apresenta grande economia de tempo e, se devidamente executado, auxilia muito na diferenciação das tipologias de vegetação, ou outras feições que possam ser alvo da classificação, com a vantagem de poder ser executada praticamente de maneira conjunta com o sobrevoos que irá captar as imagens para classificação, em poucos minutos.

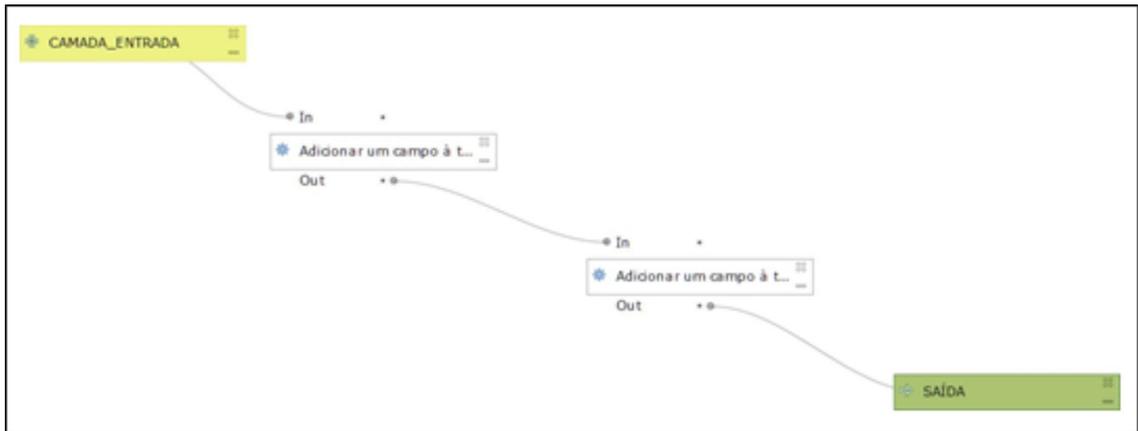


Figure 7. Modelo desenvolvido para acrescentar os campos "classe", numérico, e "tipologia", string, às tabelas de atributos.

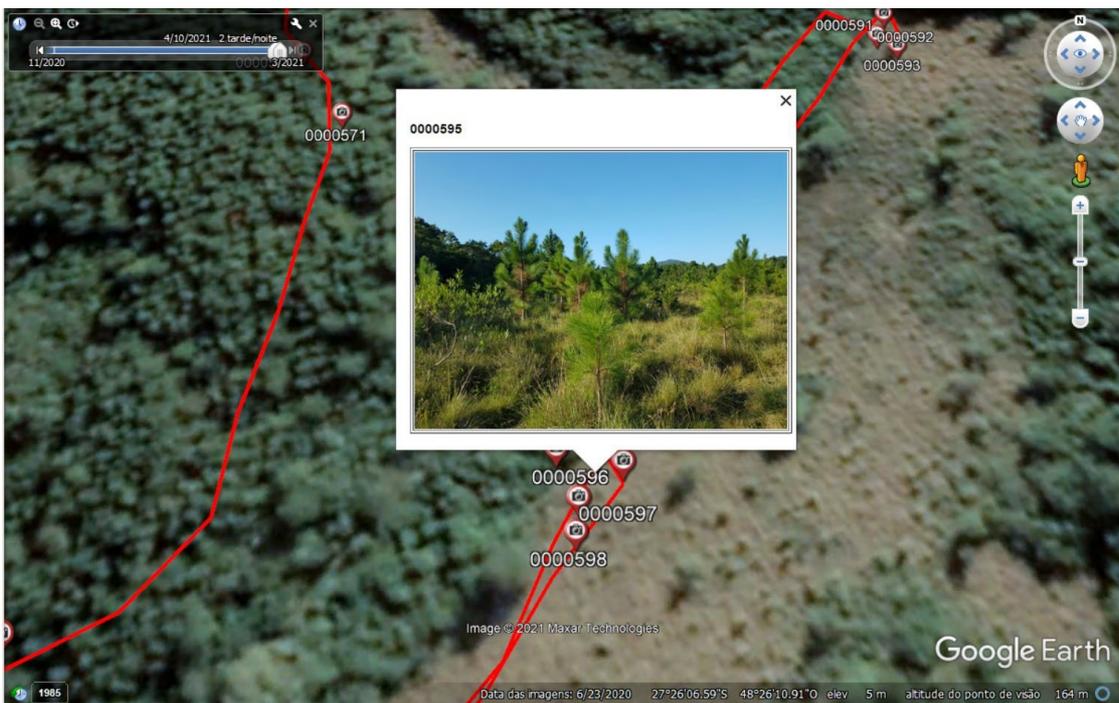


Figure 8. Fotografias tomadas em campo, registradas com geotag a partir de smartfone com o aplicativo Oruxmaps.

Para demonstrar esse processo, a Figura 9 apresenta uma região do ortomosaico com destaque em área de vegetação que poderia gerar dúvidas durante o processo de definição das classes ou de seleção das amostras para treinamento, em um trabalho que buscou identificar áreas invadidas por *Pinus sp.*



Figure 9. Área do mapeada na região do PNMDLC onde o retângulo destaca vegetação que poderia gerar dúvidas na definição de classes e seleção de amostras, enquanto a linha curva representa o trajeto percorrida durante o sobrevoo para foto filmagem.

A Figura 10 Apresenta um frame extraído da filmagem aérea realizada para servir com material auxiliar no processo de fotointerpretação.

Seguindo com o método, as classes numéricas e respectivas tipologias precisam ser definidas conforme a realidade de cada área de estudo. É importante também padronizar as cores para cada classe/tipologia, utilizando-se a codificação hexadecimal segundo demonstrado na Tabela 3 que apresenta classes definidas para quatro áreas distintas, em um estudo cujo alvo principal foi a classe 1, *Pinus sp.*

Seleção de amostras (treinamento e validação)

A seleção das amostras pressupõe a indicação manual de um número mínimo de segmentos representativos para cada uma das classes. Recomenda-se utilizar como referência o quantitativo indicado por Luca et al. (2019), que em seu processo selecionou para treinamento uma quantidade de segmentos que representassem ao menos 2% do total de segmentos da imagem, distribuídos de maneira uniforme entre as classes. Com base na imagem segmentada é possível calcular os quantitativos mínimos de segmentos necessários, de acordo com o número total de segmentos em cada área.

Uma forma de realizar o processo de indicação das amostras é através de “seleção por localização”. Para isso, é necessário indicar a localização dos segmentos a serem selecionados, o que pode ser feito pela distribuição de pequenos polígonos em formato regular, dimensionados de acordo com o objeto que se deseja indicar. Para um processo mais ágil, recomenda-se gerar os polígonos de forma padronizada para todas as classes, em diversas dimensões, conforme apresentado na Figura 11.

A indicação dos segmentos é realizada a partir da distribuição dos “quadrados”, facilmente “copiados e colados” sobre os segmentos através da ferramenta “mover e copiar”, nativa do QGIS. Os segmentos maiores, nesse caso medindo 2m de lado, podem ser utilizados nas áreas mais homogêneas, sobrepostos às bordas dos segmentos para abranger mais de um da mesma tipologia. Já os menores, medindo meio metro, ou até menos, são usados na parte interna dos segmentos, quando for necessário selecionar apenas um, conforme como na Figura 12.

Nesse processo o auxílio verificação de campo ou foto filmagem para fotointerpretação é fundamental. Figura 13 demonstra a dificuldade na fotoidentificação dos indivíduos alvo (*Pinus sp.*) de menor porte, quando próximos a espécies nativas de hábito arbustivo, com dimensões similares.



Figure 10. Frame extraído da filmagem utilizado para facilitar a identificação da área invadida



Figure 11. Vetores quadrados criados para selecionar os segmentos para cada classe a partir da ferramenta "mover e copiar" nativa do QGIS.

Table 2. Classes, tipologias e código de cores definidos para classificação das quatro áreas de estudo.

SAPIENS		
Classe	Tipologia	Código da cor
1	1-Pinus sp.	#fe0000
2	2-Restinga Arbórea	#011fff
3	3-Vegetação Herbácea	#e0eb09
4	4-Banhados	#01ff1a
5	5-Sombras	#030e00 ou #ffffff
PAERVE		
Classe	Tipologia	Código da cor
1	1-Pinus sp	#fe0000
2	2-Eucaliptus sp.	#6915e8
3	3-Casuarina sp.	#eb5af2
4	4-Restinga Herbácea	#b4eb09
5	5-Restinga Arbustiva	#29d297
6	6-Sombras	#030e00 ou #ffffff
7	7-Solo exposto	#cbce9a
8	8-Água	#3c92d5
9	9-Antrópico (Carros/objetos/decolagem/pessoas)	#ff7158
PNMDLC		
Classe	Tipologia	Código da cor
1	1-Pinus sp.	#fe0000
2	2-Restinga Herbácea	#b4eb09
3	3-Restinga Arbustivo-arbórea	#29bc97
4	4-Restinga Arbórea	#011fff
5	5-Sombras	#030e00 ou #ffffff
6	6-Solo exposto	#cbce9a
7	7-Vegetação Herbácea	#e0eb09
PEST		
Classe	Tipologia	Código da cor
1	1-Pinus sp.	#fe0000
2	2-Restinga Herbácea	#b4eb09
3	3-Restinga Arbustivo-arbórea	#29bc97
4	4-Restinga Arbórea	#215ddd
5	5-Sombras	#030e00 ou #ffffff
6	6-Solo exposto	#cbce9a
7	7-Troncos e galhos	#585827
8	8-Banhados	#01ff1a

Table 3. Quantitativos mínimo de amostras necessárias, de acordo com o número de segmentos obtidos em cada área de interesse (AOI)

AOI	Total de segmentos	2% do total	Classes	Segmentos/classe
SAPIENS	45.445	909	5	182
PAERVE	59.589	1.192	9	132
PNMDLC	47.531	951	6	158
PEST	119.190	2.384	7	341

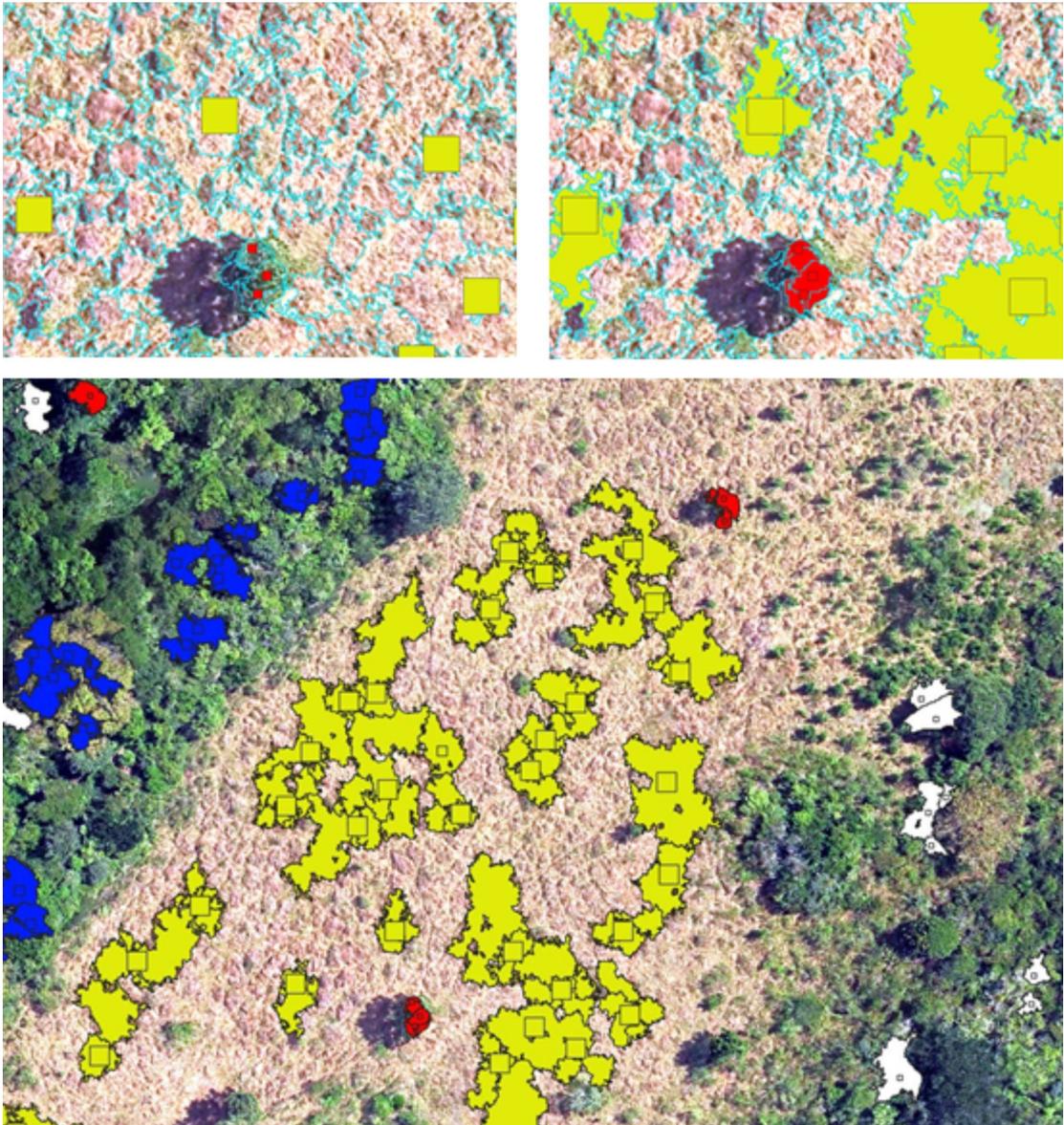


Figure 12. Exemplo de distribuição de polígonos para seleção de segmentos por localização na área do *Sapiens* Parque.



Figure 13. Parte do ortomosaico com retângulo que destaca indivíduo jovem de *Pinus* sp., em área de difícil identificação por inspeção visual, a partir do ortomosaico, por estar localizado próximo a arbustos de espécies nativas na área do PEST.

Já a Figura 14 mostra que, apesar de diminuta, a planta pode ser reconhecida pelo seu formato típico (hábito) quando vista a partir de um ângulo de visão que se aproxima daquele que o pesquisador está habituado em campo.



Figure 14. Ampliação de frame obtido a partir da filmagem realizada em sobrevoo na área do PEST.

Após a distribuição de todos os quadrados para indicar as amostras, a seleção dos segmentos (Figura 15) foi realizada através do algoritmo “Seleção por localização”, disponível na Caixa de Ferramentas de processamento do QGIS para indicar os que apresentarem intersecção com os quadrados. Importante verificar se o número de objetos selecionados foi superior a 2% do total, uniformemente distribuídos entre as classes, de acordo com a previsão indicada na Tabela 4.

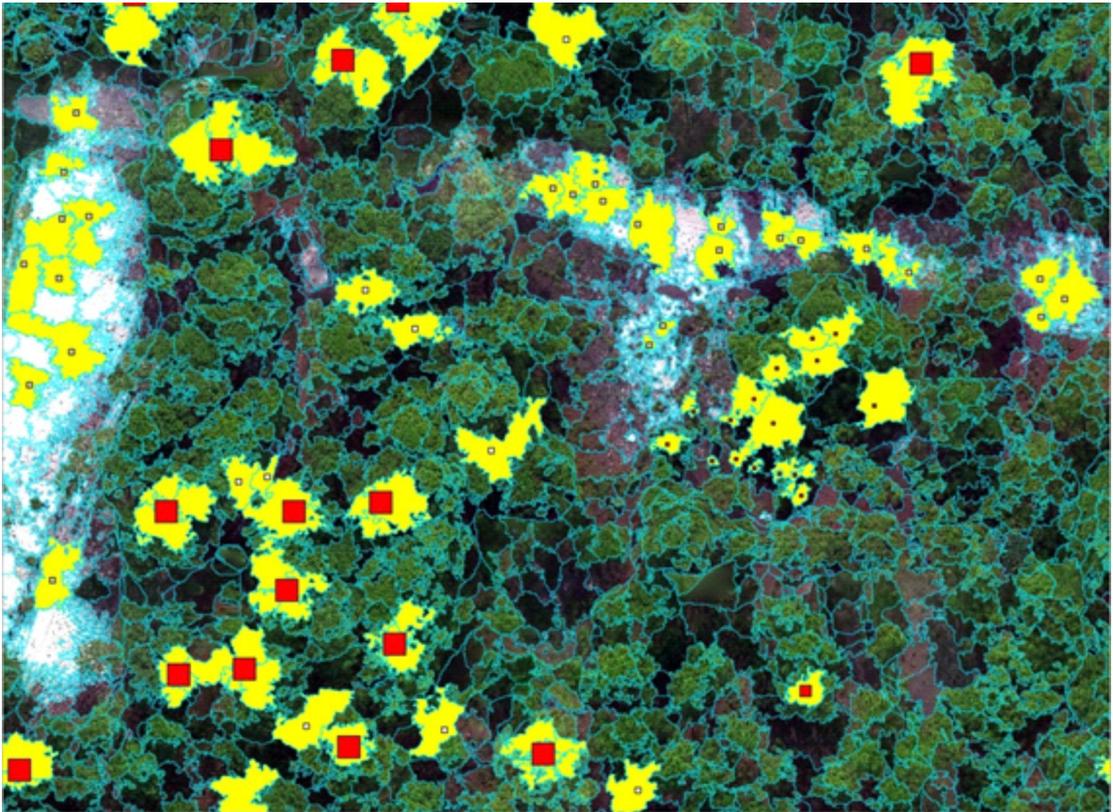


Figure 15. Seleção de segmentos que interseccionam com os quadrados em área do PAERVE, utilizando a ferramenta de “Seleção por localização” do QGIS.

Essa seleção precisa ser salva como uma nova camada vetorial de polígonos, contendo as amostras de treinamento e validação com todos os atributos calculados para cada segmento. Os campos “classe” e “tipologia” anteriormente criados, precisam ser preenchidos com a calculadora de campos, primeiramente classe por classe, utilizando novamente a seleção por localização e o recurso “atualizar apenas feições selecionadas” da calculadora de campo, e depois a tipologia, usando a expressão SQL “CASE WHEN”, conforme a tabela a seguir:

Isso resulta em arquivos vetoriais contendo os polígonos a serem utilizados para treinamento e validação, devidamente classificados conforme Figura 16.

Por fim, o total de segmentos selecionados em cada área deverá ser dividido entre treinamento e validação. Recomenda-se dividir em 70% para treinamento e 30% para validação, gerando dois arquivos vetoriais para cada área de estudo. Esse procedimento é realizado no QGIS, a partir do algoritmo “Seleção aleatória dentro de subconjuntos” disponível na Caixa de Ferramentas de processamento, indicando o campo “classe” como subconjunto e o percentual de 70%. A seleção deve ser salva como vetor de “treinamento” e depois, usando-se o recurso inverter seleção disponível na tabela de atributos seleção deve ser “invertida” para salvar os 30% restantes como vetor de “validação”.

Seleção de atributos A seleção de atributos, ou features, é uma etapa opcional, porém é importante na busca de melhoria na acurácia da classificação. Trata-se de um procedimento comum em processos de aprendizagem de máquina (machine learning) a partir do qual se busca descobrir quais são os atributos mais relevantes para classificação. Além disso, alguns atributos calculados

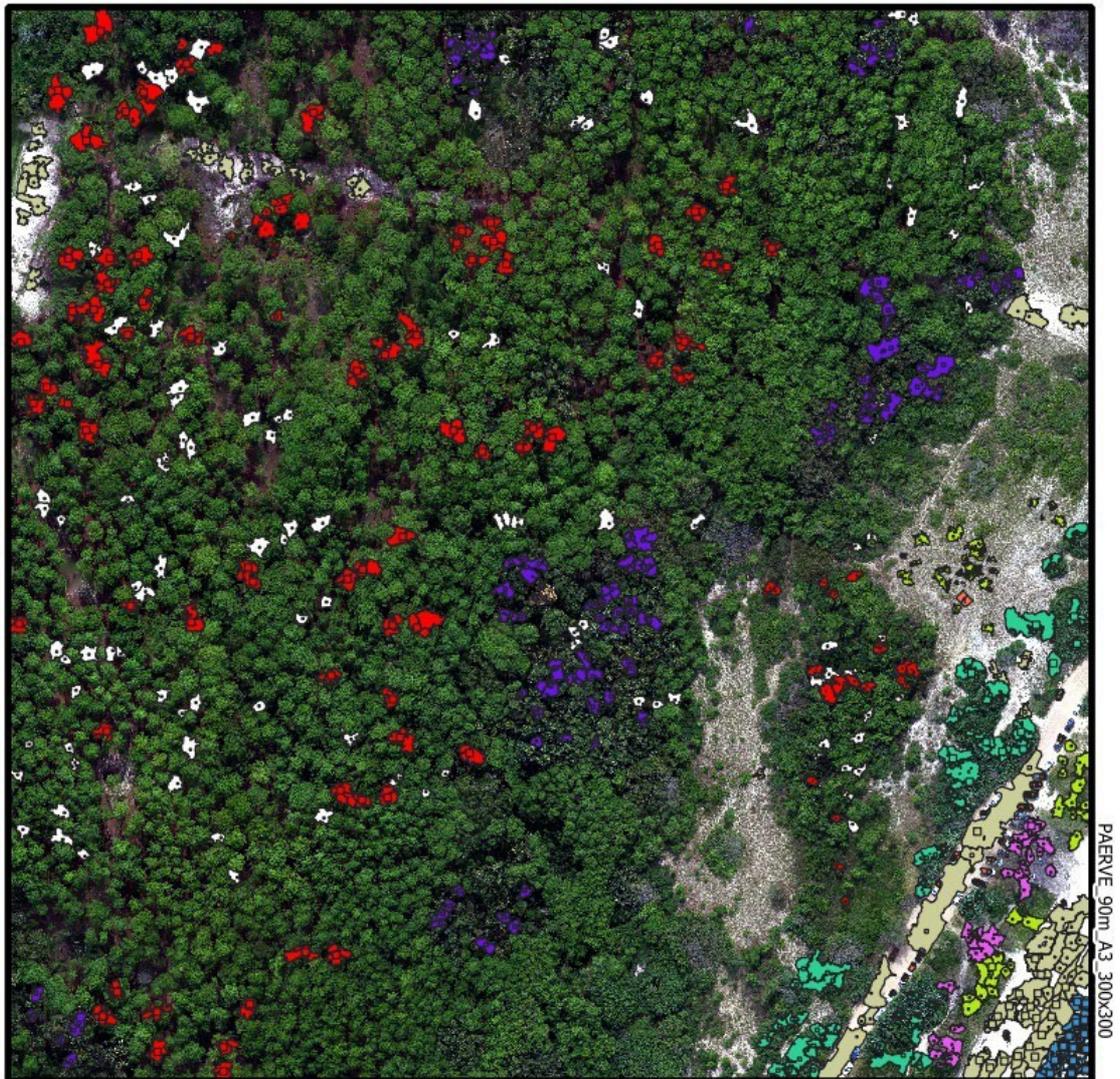


Figure 16. Amostras selecionadas para treinamento e validação na área do PAERVE

Table 4. Exemplos de expressões SQL para preenchimento do campo tipologia a partir das classes, usando a expressão "CASE WHEN".

ÁREA	EXPRESSÃO SQL
SAPIENS	CASE WHEN "classe" = 1 THEN '1- Pinus sp.' WHEN "classe" = 2 THEN '2- Restinga Arbórea' WHEN "classe" = 3 THEN '3-Vegetação Herbácea' WHEN "classe" = 4 THEN '4-Banhados' WHEN "classe" = 5 THEN '5-Sombras' END
PAERVE	CASE WHEN "classe" = 1 THEN '1- Pinus sp.' WHEN "classe" = 2 THEN '2- Eucaliyptus sp.' WHEN "classe" = 3 THEN '3-Casuarina sp.' WHEN "classe" = 4 THEN '4-Restinga Herbácea' WHEN "classe" = 5 THEN '5-Restinga Arbustiva' WHEN "classe" = 6 THEN '6-Sombras' WHEN "classe" = 7 THEN '7-Solo exposto'
PNMDLC	CASE WHEN "classe" = 1 THEN '1- Pinus sp.' WHEN "classe" = 2 THEN '2- Restinga Herbácea' WHEN "classe" = 3 THEN '3-Restinga Arbustivo-arbórea' WHEN "classe" = 4 THEN '4-Restinga Arbórea' WHEN "classe" = 5 THEN '5- Sombras' WHEN "classe" = 6 THEN '6- Solo exposto' WHEN "classe" = 7 THEN '7-Vegetação Herbácea' END
PEST	CASE WHEN "classe" = 1 THEN '1- Pinus sp.' WHEN "classe" = 2 THEN '2- Restinga Herbácea' WHEN "classe" = 3 THEN '3-Restinga Arbustivo-arbórea' WHEN "classe" = 4 THEN '4-Restinga Arbórea' WHEN "classe" = 5 THEN '5- Sombras' WHEN "classe" = 6 THEN '6- Solo exposto' WHEN "classe" = 7 THEN '7-Troncos e galhos' WHEN "classe" = 8 THEN '8-Banhados' END

em duplicidade pelos softwares fazem com que seja necessário um tratamento para evitar redundâncias que posam confundir o classificador.

As tabelas de atributos dos arquivos vetoriais contendo as amostras selecionadas para treinamento devem ser exportadas para o formato de planilhas eletrônicas para criação dos data sets utilizados na seleção de atributos. Para evitar que modelo fique muito "ajustado" aos dados fornecidos, perdendo capacidade de generalizar para novos dados, o chamado overfitting, recomenda-se que não sejam utilizados dados de validação para seleção de atributos.

Para que a seleção de atributos seja realizada de maneira abrangente, é necessário gerar três planilhas para cada área de interesse, uma contendo atributos calculados com QGIS/OTB, outra com os gerados apenas pelo TerraView/GeoDMA e a terceira contendo ambas. Dos atributos gerados pelo OTB indica-se a exclusão do Número de Pixels (nbPixels) e as médias (meanB0, meanB1, meanB2, meanB3 e meanB4) calculadas pelo LargeScaleMeanshift, uma vez que cálculo do ZonalStatistics foi idêntico para esses campos. Já para os atributos comuns entre os dois softwares (OTB e GeoDMA) notou-se que resultam pequenas diferenças em 2 a 3% dos valores.

Considerando apenas o GeoDMA, não é necessária remoção de atributos em função de duplicidade. Entretanto, alguns atributos podem apresentaram valores não numéricos (not a number – NaN), que fatalmente prejudicam a classificação. Por isso, recomenda-se a remoção de tais atributos, o que em geral ocorre com Kurtose (KURT_BX) e assimetria (SKEW_BX) associados às bandas 3 (VARI) e 4 (MDE). Para verificar essa necessidade de exclusão, basta fazer uma busca pelo termo "nan" na planilha eletrônica e excluir a coluna que possui esses valores não numéricos.

Após os devidos tratamentos, as planilhas devem ser importadas para o software Weka (Frank, Hall, & Witten, 2016), desenvolvido pela Universidade de Waikato, Hamilton, Nova Zelândia.

Para abrir os arquivos das planilhas salvas no formato Microsoft Excel (.xls ou .xlsx) e realizar a seleção de atributos no Weka, é necessário instalar os pacotes , WekaExcel e os da categoria Attribute Selection, que habilitam essas funções. Eles estão disponíveis na guia Tools, Package manager.

A seleção de atributos é realizada no aplicativo Explorer, na aba Select attributes, utilizando o método "wrapper", conforme explicado no vídeo tutorial do curso de utilização do Weka (Wekamooc, 2014). Após abertura da planilha, Figura 17, os campos numéricos "label" e "classe" devem ser removidos porque não podem ser utilizados nesse processo. Esse procedimento não altera o arquivo original da planilha, servindo apenas para indicar ao software quais campos não devem ser considerados na seleção dos atributos.

O processo de seleção de atributos é realizado na aba Select attributes Figura 18, considerando as seguintes configurações:

- Na aba Select Attributes, selecionar método "WrapperSubsetEval";
- Selecionar o campo tipologia para ser utilizado como "classe"
- Clicar no nome "WrapperSubsetEval" para abrir as configurações.

Clicar em Choose para no campo classifier selecionar J48 que fica na aba trees e alterar folds de 5 para 10, e treshhold para "-1"

- Click OK para aceitar a configuração.

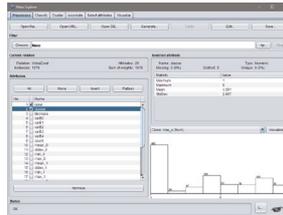


Figure 17. Ambiente da aplicação Explorer do Weka com a planilha PAERVE aberta para remoção dos campos label e classe.

- Mudar o Search Method para BestFirst, duplo clique para abrir opções, direction para Bi-Direcional e, Search Termination para 5
- Selecionar opção Cross-validation (folds 10, seed 1).
- Clicar em “Start” para iniciar o processo;
- Ao final, clicar com o direito no resultado (Results List) e salvar o log como formato de texto (.txt) para depois abrir no Excel e ordenar as features por importância.
- Repetir o procedimento para cada planilha

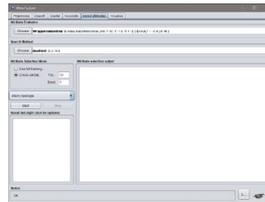


Figure 18. Aba utilizada para seleção de atributos no Weka.

Importante ressaltar que que alterações eventualmente realizadas na ordem das linhas ou colunas da planilha resultam em alterações aleatórias na seleção de atributos.

Por esse motivo, as planilhas contendo os data sets devem ser mantidas íntegras para a seleção de atributos, sem ajustes além daqueles já descritos como necessários.

A seleção de atributos pelo Weka resulta em um log que deve ser exportado em formato de texto (.txt) contendo os atributos e sua importância em percentual. Os logs devem ser convertidos em planilhas eletrônica para ordenamento e efetiva seleção das listas de atributos a serem utilizados na classificação.

Já com os logs convertidos em planilhas eletrônicas, recomenda-se ordenar os atributos em ordem de importância (folds) para selecionar com base nos percentuais, conforme Figura 19.

Importante ressaltar que para executar a classificação no OTB, a lista dos parâmetros deverá ser informada em uma lista, entre aspas duplas, utilizando espaço como separador. Conforme consta na Figura 19. Essa lista precisa ser elaborada com base no ordenamento fornecido pelo log do Weka.

No caso dos dados utilizados para realização desse experimento foram testadas treze combinações de atributos selecionados, conforme Tabela 6, para cada classificador, em cada área.

Cabe observar que a definição dos percentuais que serão testados é livre. Nesse caso utilizou-se como critério de seleção os percentuais comuns a todas as áreas de estudo, considerando cada planilha.

Treinamento e validação

A etapa de treinamento é realizada com o algoritmo TrainVectorClassifier disponível na Caixa de Ferramentas de processamento do OTB/QGIS. Os inputs do processo são os vetores contendo os polígonos selecionados para treinamento (70% do total), e validação (30% do total), indicando-se em “Field names for training features” os campos selecionados para treinamento na etapa anterior, e em “Field containing the class integer label for supervision” o campo numérico “classe”. O campo “User defined input centroids” precisa ser preenchido com “0” para que o processamento se realize.

A	B	C	D	E	F	G	H
41	numbê	of	folds	(%)	attribu	atr	
42		10(100	%)	24	mean_4	100 "mean_4"
43		9(90	%)	7	count	
44		9(90	%)	8	mean_0	90 "mean_4 count mean_0"
45		8(80	%)	22	min_3	80% "mean_4 count mean_0 min_3"
46		6(60	%)	16	mean_2	60% "mean_4 count mean_0 min_3 mean_2"
47		5(50	%)	26	min_4	50% "mean_4 count mean_0 min_3 mean_2 min_4"
48		4(40	%)	6	varB4	
49		4(40	%)	12	mean_1	
50		4(40	%)	19	max_2	40% "mean_4 count mean_0 min_3 mean_2 min_4 varB4 mean_1 max_2"
51		3(30	%)	5	varB3	
52		3(30	%)	17	stdev_2	
53		3(30	%)	18	min_2	
54		3(30	%)	23	max_3	30% "mean_4 count mean_0 min_3 mean_2 min_4 varB4 mean_1 max_2 varB3 stdev_2 min_2 max_3"
55		2(20	%)	2	varB0	
56		2(20	%)	4	varB2	
57		2(20	%)	9	stdev_0	
58		2(20	%)	10	min_0	
59		2(20	%)	11	max_0	
60		2(20	%)	15	max_1	
61		2(20	%)	20	mean_3	
62		2(20	%)	25	stdev_4	
63		2(20	%)	27	max_4	20% "mean_4 count mean_0 min_3 mean_2 min_4 varB4 mean_1 max_2 varB3 stdev_2 min_2 max_3 varB0 varB2 stdev_0 min_0 max_0 max_1 mean_3 stdev_4 max_4"
64		1(10	%)	3	varB1	
65		1(10	%)	14	min_1	
66		1(10	%)	21	stdev_3	10% "mean_4 count mean_0 min_3 mean_2 min_4 varB4 mean_1 max_2 varB3 stdev_2 min_2 max_3 varB0 varB2 stdev_0 min_0 max_0 max_1 mean_3 stdev_4 max_4 varB1 min_1 stdev_3"
67		0(0	%)	13	stdev_1	all "mean_4 count mean_0 min_3 mean_2 min_4 varB4 mean_1 max_2 varB3 stdev_2 min_2 max_3 varB0 varB2 stdev_0 min_0 max_0 max_1 mean_3 stdev_4 max_4 varB1 min_1 stdev_3 stdev_1"

Figure 19. Log do Weka para os parâmetros computados pelo OTB para a área do SAPIESN, convertido em formato de planilha para ordenamento dos atributos de acordo com a importância percentual (folds).

Table 6. Critérios para seleção das listas de atributos para classificação de acordo com cada planilha.

CRITÉRIO DE SELEÇÃO (% mín. de importância)	Planilhas OTB	Planilhas GeoDMA	Planilhas OTB e GeoDMA
0% (todos os atributos)	X	X	X
10%	-	X	X
20%	X	X	X
30%	X	X	X
40%	X	-	-
50%	X	-	-

Dentre os algoritmos para classificação disponíveis no QGIS, nessa aplicação recomenda-se utilizar dois, Random Forest (RF) e Support Vector Machine (SVM), que segundo OTB Development Team (2021) são baseadas nas bibliotecas OpenCV Machine Learning (2.3.1 e posteriores) e LibSVM.

A partir das combinações de atributos selecionados pelo Weka, nesse caso as treze indicadas na Tabela 6, é necessário que sejam testadas para cada classificador, em cada área. Em função disso, recomenda-se que sejam realizados processamentos em lote no QGIS, que permite definir os parâmetros para realização dos treinamentos em lote.

Esse algoritmo gera, obrigatoriamente, um arquivo .txt contendo o modelo produzido pelo classificador e opcionalmente permite salvar, no mesmo formato, a matriz de confusão produzida com base nos segmentos do arquivo vetorial indicado para validação. Também gera um log de grande utilidade que, dentre as informações inerentes ao processamento, também indica o classificador utilizado, uma cópia da matriz de confusão e as métricas de acurácia da classificação Precision, Recall, Fscore e Kappa Index (global). Com isso é possível já identificar se o processamento foi bem-sucedido, tempo decorrido, bem como ter uma ideia da qualidade da classificação obtida.

Avaliação da acurácia

A avaliação de acurácia pode ser realizada a partir dos índices o índice Kappa Cohen, Cohen (1960) apud P Pantaleão and Scofield (2009), que permite avaliar a classificação como um todo, e pelo F-score (Sokolova, Japkowicz, & Szpakowicz, 2006), que permite avaliar cada classe. Também é possível computar as médias de Acurácia Global, Acurácia do usuário ou precision e Acurácia do

produtor ou recall, segundo Story e Congalton (1986) e Fung e LeDrew(1988) apud (Pantaleão & Scofield, 2009), apud (Pantaleão & Scofield, 2009).

Essas medidas podem ser implementadas em planilhas eletrônicas segundo fórmulas apresentadas na Tabela 7.

Table 7. Fórmulas das medidas de acurácia

Nome	Fórmula	Objetivo
Acurácia do usuário (precision)	$au_i = x_{ii}/x_{i+}$ Em que: x_{ii} = objetos classificados corretamente para uma classe i ; x_{i+} = total de objetos classificados para uma classe i	Reflete os erros de comissão que indicam a probabilidade de um objeto classificado em uma determinada classe realmente pertencer a essa classe (LILLESAND et al. 2004 apud Sothe (2015) Sothe (2015)).
Acurácia do produtor (recall)	$ap_i = x_{ii}/x_{+i}$ Em que: x_{ii} = objetos classificados corretamente para uma classe i ; x_{+i} = total de objetos de referência para uma classe i	Reflete os erros de omissão, ou seja, a probabilidade de um objeto ser excluído (não classificado) da classe a que ele pertence (CONGALTON e GREEN 1999; LILLESAND et al. 2004, apud Sothe (2015)). Também referido como “sensibilidade”, segundo Sokolova et al. (2006).
Erro de Comissão (inclusão)	$p_{i+} = (x_{i+} - x_{ii})/x_{i+}$	Probabilidade de ocorrência de erro de Comissão (inclusão).
Erro de Omissão	$p_{+i} = (x_{+i} - x_{ii})/x_{+i}$	Probabilidade de ocorrência de erro de omissão.
Kappa condicional (usuário)	$kcu_i = (au_i - p_{i+})/(1 - p_{i+})$ Em que: au_i = acurácia do usuário; p_{i+} = erro de comissão	Kappa condicional permite avaliar a acurácia de uma determinada classe (ROSENFELD e FITZPATRICK- LINS, 1986, apud Sothe (2015)).
Kappa condicional (produtor)	$kcp_i = (ap_i - p_{+i})/(1 - p_{+i})$ Em que: ap_i = acurácia do produtor; p_{+i} = erro de omissão	O índice Kappa condicional do usuário avalia os erros de comissão. O Kappa condicional do produtor avalia os erros de omissão.
Acurácia Global	$Ag = x_i/n$ Em que: x_i = total de objetos classificados corretamente; n = total de objetos considerados na análise	Medida utilizada na avaliação global da classificação.
Probabilidade estimada		Utilizada para compor o índice Kappa Cohen.
Kappa Cohen	$\kappa = 1 - \frac{p_{i+} + p_{+i} - p_{ii}}{1 - p_{ii}}$	Kappa utilizado para a avaliação global da classificação. Se diferencia da acurácia global por incorporar os elementos fora da diagonal principal (CONGALTON e GREEN 1999, apud Sothe (2015) Sothe (2015)).
F-score	$F = 2 * \frac{precision * recall}{precision + recall}$	Medida parametrizada com pesos iguais para precision e recall permitindo comparar o desempenho dos modelos para cada classe, sendo a medida mais frequentemente utilizada em machine learning com dados desbalanceados entre precision e recall (He & Ma, 2013).

F-score e índice Kappa são automaticamente fornecidos no log de processamento do QGIS para o algoritmo TrainVectorClassifier. Assim, é possível identificar os processamentos que obtiveram valores mais elevados para Kappa global e F-score da classe 1 Pinus sp., referente ao alvo de interesse no estudo.

Identificando-se os conjuntos de atributos que obtiveram valores mais elevados nessa etapa, é possível identificar selecionar os modelos promissores pra a classificação, e também realizar computar as outras métricas a partir da matriz de confusão fornecida pelo QGIS.

Classificação

A classificação propriamente pode ser aplicada a todos os arquivos de modelo gerados pelo QGIS na etapa de treinamento. Para selecionar apenas as melhores classificações, recomenda-se utilizar nessa etapa apenas os modelos que obtiveram os maiores valores do índice Kappa, para cada classificador, em cada área, e para os modelos que resultaram no maior Fscore para a classe alvo, nesse caso a classe 1, referente à tipologia 1-Pinus sp. Isso porque, nem sempre os maiores índices Kappa irão coincidir com os maiores F-score para a classe alvo.

A classificação é realizada com o algoritmo VectorClassifier disponível no pacote OTB a partir da Caixa de Ferramentas de processamento do QGIS. Os inputs são o vetor da segmentação gerado pelo ZonalStatistics já acrescido das métricas calculadas pelo GeoDMA. O campo “Field names to be calculated” foi preenchido com os mesmos campos selecionados para treinamento e o output resultou arquivos vetoriais do tipo Shapefile (.shp).

Estimativas de área

Após a conclusão da classificação, é possível computar as áreas ocupadas por cada classe, individualizadas por segmentos, ou então em sua totalização para cada classe. Para tornar os arquivos vetoriais da classificação mais leves, basta exportá-los para novas camadas vetoriais apenas com os campos “label, classe e tipologia”, o que irá descartar todos os outros utilizados no processo de classificação.

Feito isso, para calcular as áreas de cada segmento, utilizar o algoritmo nativo do QGIS “Corrigir geometrias” e depois calcular a área em m² utilizando a calculadora de campos. O procedimento resulta em arquivo vetorial parte simples, ou seja, contendo cada segmento individualizado na tabela de atributos. Também pode ser criado um modelo do QGIS para execução dessa tarefa em lote, conforme Figura 20 (Anexo).

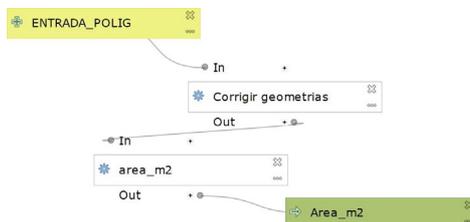


Figure 20. Modelo para calcular de cada segmento em m².

Já para gerar arquivos vetoriais multiparte, com os polígonos de cada classe dissolvidos e as respectivas áreas em m2 e hectares devidamente calculadas, é necessário utilizar o algoritmo Dissolver indicando-se o campo “classe”, para que os polígonos possam ser “mesclados” por classes antes de calcular as áreas. Importante ressaltar que antes de cada procedimento, para evitar problemas de processamento devido a erros de topologia deve ser executado o algoritmo Corrigir geometrias. Por fim, deve ser descartado o campo Label que passa a não mais fazer sentido após a dissolução dos polígonos. Como esse processo envolve mais etapas, recomenda-se a criação de um modelo do QGIS conforme Figura 21 (Anexo).

A proposta metodológica pode ser sintetizada na figura 22 (Anexo).

Considerações finais

Acredita-se que o método empregado possa contribuir, tanto na interpretação de fenômenos, quanto no planejamento das ações necessárias ao manejo e controle envolvendo áreas de vegetação.

Além do mapeamento da espécie alvo nesse trabalho, o método também se demonstrou útil para classificação de imagens em geral, podendo ser utilizado no manejo de outras espécies vegetais exóticas, ou até em outras aplicações ambientais.

Apesar dos resultados alcançados nesse trabalho, é provável que classificações mais acuradas possam ser obtidas mediante ajustes na técnica empregada. Nesse sentido, sugere-se investigar a possibilidade de utilização de outros índices de vegetação em substituição, ou em conjunto com VARI, bem como a utilização de algoritmos para extração de atributos texturais pelo OTB e outras técnicas de seleção de atributos que possam ser programadas sob medida em scripts utilizando a linguagem Python, por exemplo.

Em relação à aplicação da técnica em áreas maiores, sugere-se investigar possibilidades de escalabilidade, visto que a capacidade computacional pode representar um gargalo. Uma alternativa pode ser a divisão em subáreas com dimensões adequadas, bem como uma divisão do fluxo de segmentação em etapas, podendo-se estudar eliminação da primeira, que é opcional, conforme proposto por Luca et al. (2019). Outra possibilidade é o desenvolvimento do modelo em uma área restrita que contemple suficientemente todas as classes, e sua aplicação para classificação de toda a cena segmentada.

References

- Bechara FC. 2003. Restauração Ecológica De Restingas Contaminadas Por Pinus No Parque Florestal Do Rio Vermelho, Florianópolis, Sc. Univ Fed St Catarina [Internet]. [accessed 2021 Feb 6] 108(3):136. <https://repositorio.ufsc.br/bitstream/handle/123456789/86536/190967.pdf?sequence=1>
- Convenção sobre Diversidade Biológica. 2010. Panorama da Biodiversidade Global. [place unknown]. <http://link.springer.com/10.1007/s00740-010-0229-z%0Ahttps://www.cbd.int/doc/gbo/gbo2/cbd-gbo2-po.pdf>
- Dechoum M de S, Giehl ELH, Sühs RB, Silveira TCL, Ziller SR. 2019. Citizen engagement in the management of non-native invasive pines: Does it make a difference? *Biol Invasions*. 21(1):175–188.
- Frank E, Hall MA, Witten IH. 2016. The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques” [Internet]. Fourth Edi. Hamilton, New Zeland: Morgan Kaufmann; [accessed 2021 Jun 26]. https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf
- He H, Ma Y. 2013. Imbalanced learning: foundations, algorithms, and applications. <https://doi.org/10.1002/9781118646106>
- Henrich V, Krauss G, Götze C, Sandow C. 2021. Index DataBase: A database for remote sensing indices [Internet]. [accessed 2021 Jun 20]. <https://www.indexdatabase.de/>
- IBGE. 2012. Manual Técnico da Vegetação Brasileira. 2nd ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; [accessed 2021 Feb 6]. <http://www.terrabrasil.org.br/ecotecadigital/pdf/manual-tecnico-da-vegetacao-brasileira.pdf>
- INPE. 2021. GeoDMA Features [Internet]. [accessed 2021 Jun 24]. http://wiki.dpi.inpe.br/doku.php?id=geodma_2:features
- Liau YT. 2014. Hierarchical segmentation framework for identifying natural vegetation: A case study of the Tehachapi Mountains, California. *Remote Sens* [Internet]. 6(8):7276–7302. <https://www.mdpi.com/2072-4292/6/8/7276>
- De Luca G, Silva JMN, Cerasoli S, Araújo J, Campos J, Di Fazio S, Modica G. 2019. Object-based land cover classification of cork oak woodlands using UAV imagery and Orfeo Toolbox. *Remote Sens*. 11(10).
- OpenDroneMap Authors. 2020. ODM – A command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images [Internet]. [accessed 2021 Jun 20]. <https://github.com/OpenDroneMap/ODM>
- OTB Development Team. 2021. OTB Cookbook Documentation [Internet]. [accessed 2021 Jun 20]. <https://www.orfeo-toolbox.org/CookBook/>
- Pantaleão E, Scofield GB. 2009. Comparação entre medidas de acurácia de classificação para imagens do satélite ALOS. XIV Simpósio Bras Sensoriamento Remoto [Internet]. [accessed 2021 Mar 28]:7039–7046. <http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr%04080/2008/11.17.20.26/doc/7039-7046.pdf>
- QGIS.org. 2021. QGIS Geographic Information System [Internet]. [accessed 2021 Jun 20]. <http://www.qgis.org>
- Sobrinho M da S, Cavalcante A de MB, Duarte A de S, de Sousa GDS. 2019. Modeling the potential distribution of *Mangifera indica* L. Under future climate scenarios in the caatinga biome. *Rev Bras Meteorol*. 34(3):351–358.

Sokolova M, Japkowicz N, Szpakowicz S. 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. AAAI Work - Tech Rep [Internet]. [accessed 2021 Feb 27] WS-06-06(c):24–29. <https://www.aaai.org/Papers/Workshops/2006/WS-06-06/WS06-06-006.pdf>

Sothe C. 2015. Classificação do estágio sucessional da vegetação em áreas de floresta ombrófila mista empregando análise baseada em objeto e ortoimagens [Internet]. [accessed 2021 Mar 28]:249. https://www.udesc.br/arquivos/cav/id_cpmenu/1482/CAMILE_SOTHE_dissertacao_1_5683968824675_1482.pdf

WekaMOOC. 2014. More Data Mining with Weka (4.1: Attribute selection using the “wrapper” method) [Internet]. [accessed 2021 Jun 20]. <https://www.youtube.com/watch?v=Pf9yKjSiVnw>

Ziller SR, Dechoum M de S, Duarte Silveira RA, Marques da Rosa H, Mello Oliveira BC, Zenni RD, Motta MS, Filipe da Silva L. 2020. A priority-setting scheme for the management of invasive non-native species in protected areas. *NeoBiota*. 62(October):591–606.

Anexo

Em negrito os que são computados para cada uma das bandas do raster utilizado no processo, nesse caso a banda 0. Fonte: INPE(2021), adaptado.

Table 7. Campos computados pela função Feature Extraction do plugin GeoDMA implementado no software Terraview do INPE.

CAMPO	ATRIBUTO	ALGORITMO	DESCRIÇÃO
AMPL_B0	Amplitude	Feature Extraction	Define a amplitude dos pixels dentro do objeto. A amplitude significa o valor máximo do pixel menos o valor mínimo do pixel.
COUNT_B0	Contagem de pixels	Feature Extraction	Define o número total de pixels dentro do objeto, incluindo pixels com valores fictícios.
KURT_B0	Curtose	Feature Extraction	Retorna o valor de curtose para todos os pixels válidos (não fictícios) dentro do objeto.
MAX_VAL_B0	Máximo valor	Feature Extraction	Calcula o valor máximo do nível de cinza (não fictício) dentro do objeto.
MEAN_B0	Média	Feature Extraction	Calcula o valor médio de todos os N pixels dentro do objeto.
MEDIAN_B0	Mediana	Feature Extraction	Calcula a mediana de todos os N pixels dentro do objeto.
MIN_VAL_B0	Mínimo valor	Feature Extraction	Calcula o valor mínimo do nível de cinza (não fictício) dentro do objeto.
MODE_B0	Moda	Feature Extraction	Retorna o valor de nível de cinza de maior ocorrência (moda) para todos os N pixels (não fictícios) dentro do objeto. Quando o objeto é

multimodal, o primeiro valor
é assumido.

Continued on next page

Table 7 continued

N_MOD_B0	Número de modas	Feature Extraction	Retorna o número de modas do objeto.
SKEW_B0	Assimetria (skewness)	Feature Extraction	Retorna o valor de assimetria para todos os pixels válidos (não fictícios) dentro do objeto.
STDDEV_B0	Desvio Padrão	Feature Extraction	Retorna o desvio padrão de todos os N pixels (não fictícios) dentro do objeto.
SUM_B0	Soma	Feature Extraction	Define o número total de pixels dentro do objeto sem valores fictícios.
VLDCNT_B0	Contagem de pixels válidos (sem valores fictícios)	Feature Extraction	Define o número total de pixels dentro do objeto sem valores fictícios.
VARCOEF_B0	Coefficiente de variação para os pixels válidos	Feature Extraction	Retorna o coeficiente de variação dos valores para todos os pixels válidos (não fictícios) dentro do objeto
VAR_B0	Variância	Feature Extraction	Retorna a variação de todos os N pixels (não fictícios) dentro do objeto.
CONTSE_B0	Contraste	Feature Extraction	Retorna uma medida do contraste de intensidade entre um pixel e seu vizinho sudeste sobre o objeto. O contraste é 0 para um objeto constante. Também é conhecido como Sum of Squares Variance
DISSE_B0	Dissimilaridade	Feature Extraction	Mede o quão diferentes os elementos da Matriz de Coocorrência de Nível de Cinza - GLCM são uns dos outros e é alto quando a região local tem alto contraste.
ENERGSE_B0	Energia	Feature Extraction	Ele retorna a raiz quadrada do segundo momento angular, calculada pela soma dos elementos quadrados na Matriz de Coocorrência de Nível de Cinza - GLCM. A energia é 1 para uma imagem constante.

Continued on next page

Table 7 continued

ENTRSE_B0	Entropia	Feature Extraction	Mede a desordem em uma imagem. Quando a imagem não é uniforme, muitos elementos da Matriz de Coocorrência de Nível de Cinza - GLCM têm valores pequenos, resultando em grande entropia.
HOMOGSE_B0	Homogeneidade	Feature Extraction	Assume valores mais altos para diferenças menores na Matriz de Coocorrência de Nível de Cinza - GLCM. Também chamado de momento de diferença inversa. A homogeneidade é 1 para um GLCM diagonal.
BRATIO_B0	Taxa de contribuição	Feature Extraction	Descreve a contribuição de determinada banda para a região.
P_AREA	Área do polígono (segmento)	Feature Extraction	Retorna a área do objeto, medida na unidade de medida do Sistema de Referência Espacial atual.
P_PERIM	Perímetro do polígono (segmento)	Feature Extraction	Retorna o perímetro do objeto, medido na unidade de medida do Sistema de Referência Espacial atual.
P_FRACDIM	Dimensão fractal do polígono (segmento)	Feature Extraction	Retorna a dimensão fractal de um objeto.
P_PERARAT	Proporção entre perímetro e área do segmento	Feature Extraction	Calcula a proporção entre o perímetro e a área de um objeto.
P_COMPAC	Compacidade do segmento	Feature Extraction	Retorna a compacidade do objeto.
PBOX_AREA	Área da caixa delimitadora do segmento	Feature Extraction	Retorna a área da caixa delimitadora de um objeto, medida na unidade de medida do Sistema de Referência Espacial atual.
PBOX_PERIM	Perímetro da caixa delimitadora do segmento	Feature Extraction	Retorna o perímetro da caixa delimitadora de um objeto, medido na unidade de medida do Sistema de Referência Espacial atual.

Continued on next page

Table 7 continued

PBOX_LEN	Altura da caixa delimitadora do segmento	Feature Extraction	É a altura da caixa delimitadora do objeto, medida na unidade de medida do Sistema de Referência Espacial atual.
PBOX_WIDTH	Largura da caixa delimitadora do segmento	Feature Extraction	Retorna a largura da caixa delimitadora do objeto, medida na unidade de medida do Sistema de Referência Espacial atual.
POL_ANGLE	Ângulo principal do objeto (segmento)	Feature Extraction	Representa o ângulo principal de um objeto. É obtido calculando a elipse circunscrita mínima e o ângulo do
PELLIP_FIT	Proporção entre área do objeto e área da elipse mínima circunscrita	Feature Extraction	maior raio da elipse se ajusta ao ângulo do objeto. Encontra a elipse circunscrita mínima ao objeto e retorna a proporção entre a área do objeto e a área da elipse.
PGYRATIUS	Distância média entre cada vértice do polígono e seu centroide	Feature Extraction	Este recurso é igual à distância média entre cada vértice do polígono e seu centroide. Quanto mais semelhante a um círculo for o objeto, mais provavelmente o centróide estará dentro dele e, portanto, esse recurso estará mais próximo de 0.
POLRADIUS	Raio do polígono	Feature Extraction	Retorna o raio do polígono. Corresponde à distância máxima entre o centroide do polígono e seus vértices.
PCIRCLE	Relação entre área do segmento e o menor círculo interno	Feature Extraction	Relaciona as áreas do objeto e o menor círculo circunscrito ao redor do objeto. Na equação, R é a distância máxima entre o centróide e todos os vértices.
PSAHPEIDX	Relação entre área do segmento e o menor círculo interno	Feature Extraction	Este recurso corresponde à razão entre o perímetro do polígono e a raiz quadrada da área do polígono.
PDENSITY	Proporção entre área e raio do polígono	Feature Extraction	Este recurso corresponde à proporção entre a área do polígono e o raio do polígono.

Continued on next page

Table 7 continued

PRECTFIT

Proporção entre área do segmento e de seu mínimo retângulo externo

Feature Extraction

Este recurso ajusta um retângulo mínimo fora do objeto e calcula a proporção entre sua área e a área deste retângulo. Quanto mais próximo 1 estiver esse recurso, mais semelhante a um retângulo.

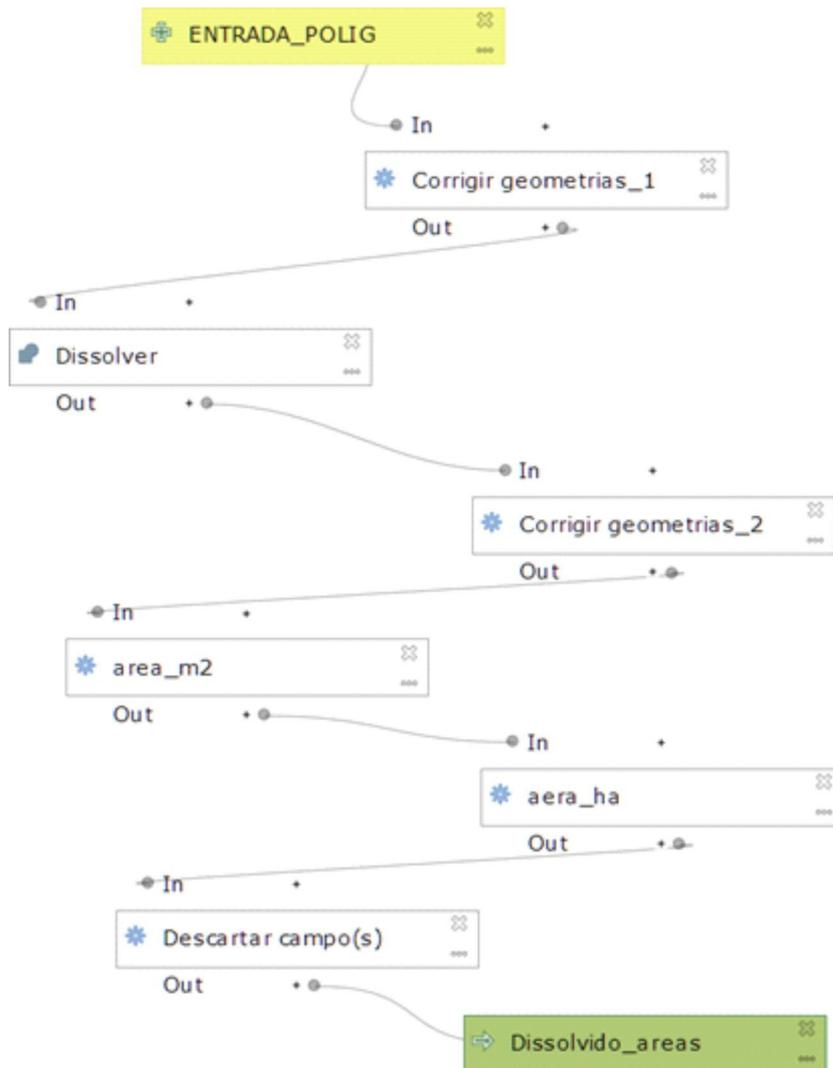


Figure 21. Modelo do QGIS para calcular áreas totais para cada classe

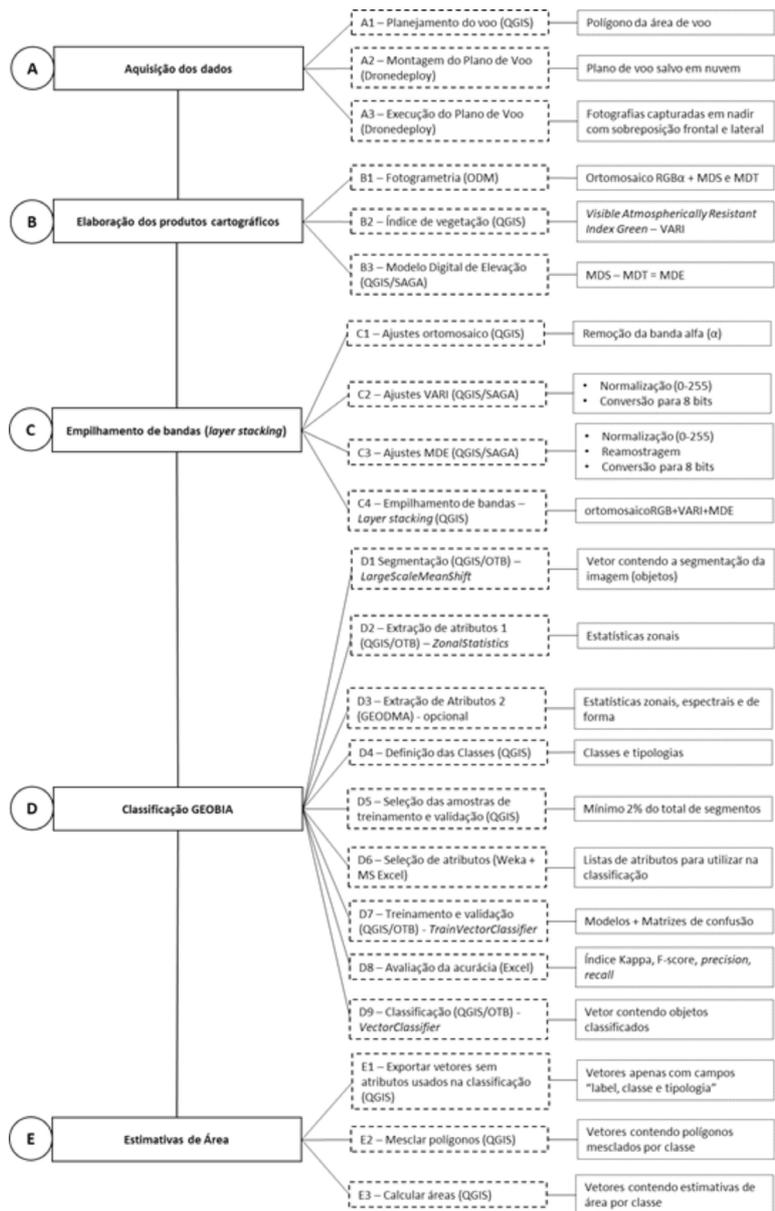


Figure 22. Fluxograma geral das etapas propostas na metodologia