

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

EVELAINE BERGER RIBEIRO

**CLASSIFICAÇÃO DE IMAGEM ORBITAL RAPIDEYE UTILIZANDO BANCO DE
DADOS NOSQL E MÉTODO GEOBIA**

PONTA GROSSA
2017

EVELAINE BERGER RIBEIRO

CLASSIFICAÇÃO DE IMAGEM ORBITAL RAPIDEYE UTILIZANDO BANCO DE DADOS NOSQL E MÉTODO GEOBIA

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada, curso de Mestrado em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para obtenção do título de Mestre.

Orientadora: Prof.^a Dr.^a Selma Regina Aranha Ribeiro

PONTA GROSSA
2017

Ficha Catalográfica
Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG

R484 Ribeiro, Evelaine Berger
 Classificação de imagem orbital
 Rapideye utilizando banco de dados NoSQL e
 método GEOBIA/ Evelaine Berger Ribeiro.
 Ponta Grossa, 2017.
 92f.

 Dissertação (Mestrado em Computação
 Aplicada - Área de Concentração:
 Computação para Tecnologias em
 Agricultura), Universidade Estadual de
 Ponta Grossa.

 Orientadora: Prof^a Dr^a Selma Regina
 Aranha Ribeiro.

 1.Classificação digital. 2.GEOBIA.
 3.NoSQL. I.Ribeiro, Selma Regina Aranha.
 II. Universidade Estadual de Ponta Grossa.
 Mestrado em Computação Aplicada. III. T.


CDD: 630.2

TERMO DE APROVAÇÃO


Evelaine Berger Ribeiro

“CLASSIFICAÇÃO DE IMAGEM ORBITAL RAPIDEYE UTILIZANDO BANCO DE DADOS NOSQL E MÉTODO GEOBIA”.

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:


Dr.ª Alaine Margarete Guimarães
UEPG


Dr.ª Maria Salete Marcon Gomes
UEPG


Dr. Luiz/Gilberto Bertotti
UNICENTRO/Guarapuava - PR

Ponta Grossa, 20 de junho de 2017.

AGRADECIMENTOS

A Deus por ter me concedido o dom da vida, saúde e forças para nunca desistir dos meus ideais.

Aos meus pais, João Luiz Ribeiro e Roseni Berger Ribeiro, que dedicaram suas vidas para cuidar da nossa família, por priorizarem minha educação oferecendo a oportunidade de estudar. Saibam que tudo o que sou hoje, eu devo a vocês.

À minha irmã Josiane por estar ao meu lado em todos os momentos, pela amizade e também o incentivo aos estudos. Saiba que cresci me espelhando em você.

Ao meu noivo Fabio Almeida pelo companheirismo, por todo o apoio e incentivo que me ofereceu e por sempre ter acreditado no meu potencial.

Aos familiares, meu cunhado Francismar, o qual considero como irmão, afilhado João Miguel, meus sogros Jussara e Walfrido, tios Nilton e Marcia, primos Gabriel e Rafael, por estarem ao meu lado, pelo incentivo e compreensão nos momentos em que não pude estar presente.

À minha orientadora Prof^a. Dr^a. Selma Regina Aranha Ribeiro pela paciência, amizade, incentivo e por todos os ensinamentos e contribuições para a realização deste trabalho.

Aos amigos, especialmente a Jaqueline pela amizade, parceria, conversas e conselhos durante esse tempo.

À Universidade Estadual de Ponta Grossa e ao Programa de Pós-Graduação pela oportunidade e estrutura cedida para os estudos.

A todos que de forma direta ou indireta contribuíram para o desenvolvimento deste trabalho.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito.
Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.
(Martin Luther King)

RESUMO

Com as informações adquiridas das imagens capturadas pelo Sensoriamento Remoto e das técnicas disponíveis nos Sistemas de Informação Geográfica pode-se gerar mapeamentos temáticos para uso e cobertura do solo. Para isso, é realizada a classificação de imagens para definir as classes de interesse. Essa classificação pode ser feita *pixel a pixel* ou por regiões. Em imagens de alta resolução, como a Rapideye, é indicada a classificação por regiões. Esse método considera as informações do *pixel* e sua vizinhança, agrupando *pixels* com características semelhantes, formando as regiões. Portanto, recomenda-se aplicar o método da segmentação pela GEOBIA, que segmenta a imagem em regiões, visando extrair características espaciais, espectrais e de textura. Como resultado desse método, têm-se o vetor de regiões e o banco de dados relacional com os atributos (espaciais, espectrais e de textura). O presente trabalho teve como objetivo obter a classificação do uso e cobertura do solo da imagem Rapideye com banco de dados NoSQL orientado a grafos para análise dos atributos extraídos mediante a GEOBIA. A metodologia desenvolvida utilizou a Análise Multivariada para analisar os atributos resultantes da segmentação. Por meio do dendrograma foi possível a separação dos grupos de atributos (espaciais, espectrais e de textura), que foram utilizados para as consultas de busca por agrupamentos de regiões com características semelhantes no grafo formado pelo banco de dados NoSQL. As regiões foram classificadas de acordo com as classes de interesse definidas no processo de fotointerpretação, gerando a imagem classificada. Para validar o resultado, realizou-se a classificação da imagem da área de estudo pelos algoritmos Distância Mínima, Máxima Verossimilhança e KNN e a matriz de confusão. O algoritmo KNN apresentou melhor classificação, com índice kappa de 0,77 e então foi utilizada para comparação com a imagem classificada pelo banco de dados NoSQL, por meio da tabulação cruzada. O cruzamento dos dados mostrou que a imagem classificada pelo banco de dados NoSQL obteve resultados positivos. Conclui-se que a pesquisa alcançou os objetivos propostos apresentando resultados satisfatórios para o método desenvolvido para classificação do uso e cobertura do solo.

Palavras-chave: Classificação Digital, GEOBIA, NoSQL.

ABSTRACT

The information from images captured by Remote Sensing and the techniques available in the Geographic Information Systems, it is possible to generate thematic mappings for use and land cover. For this, the classification of images is realized to define interest classes. This classification can be done pixel by pixel or by regions. In high resolution images, such as Rapideye, classification by region is indicated. This method considers the information of the pixel and its neighborhood, grouping pixels with similar characteristics create the regions. Therefore, it is recommended to apply the GEOBIA segmentation method, which segments the image in regions to extract spatial, spectral and texture characteristics. As a result of this method, have the region vector and the relational database with the attributes (spatial, spectral and texture). The objective of this work was to obtain the classification of the use and coverage of the soil of the Rapideye image using the NoSQL database oriented to graphs to analyze the attributes extracted through GEOBIA. The developed methodology used the Multivariate Analysis to analyze the attributes resulting from the segmentation. The dendrogram it was possible to separate the groups of attributes (spatial, spectral and texture), which were used for the search queries by groupings of regions with similar characteristics in the graph formed by the NoSQL database. The regions were classified according to the interest classes defined in the photointerpretation process, generating the classified image. To validate the result, the image area of the study area was classified by the Minimum Distance, Maximum Likelihood and KNN algorithms and the confusion matrix. The KNN algorithm presented better classification, with a kappa index of 0.77 and was then used for comparison with the image classified by the NoSQL database, through cross tabulation. The cross-validation of the data showed that the image classified by the NoSQL database obtained positive results. It was concluded that the research reached the proposed objectives presenting satisfactory results for the method developed for classification of land use and land cover.

Keywords: Digital Classification, GEOBIA, NoSQL.

LISTA DE FIGURAS

Figura 1 - Imagem Rapideye da área de estudo.	32
Figura 2 - Representação das bandas espectrais na escala cinza da imagem Rapideye da área de estudo.	34
Figura 3 - Composições coloridas das bandas espectrais da imagem Rapideye.....	35
Figura 4 - Diagrama de blocos para a metodologia proposta.....	36
Figura 5 - Diagrama de blocos: inserção dos dados no banco de dados NoSQL.	44
Figura 6 - Imagens da área de estudo apresentada pela verdadeira cor, pela composição das bandas 4R5G3B e a banda artificial NDVI.....	49
Figura 7 - Resultado das amostras coletadas na fotointerpretação para as 8 classes de interesse utilizadas nas classificações.	50
Figura 8 - Imagem resultante da classificação supervisionada pelo algoritmo de Distância Mínima, com 8 classes de uso e ocupação do solo.....	52
Figura 9 - Resultados dos testes de segmentação com o algoritmo de escala Edge e algoritmo de fusão FLSA, com os vetores de segmentos gerados durante o processo.....	54
Figura 10 - Apresentação das regiões e do banco de dados espacial gerados a partir da segmentação mediante GEOBIA.....	56
Figura 11 - Dendrograma dos 46 atributos provenientes da GEOBIA, as distâncias entre os grupos e com a linha de corte formando 4 grupos.	57
Figura 12 - Tela do programa para inserir os dados no banco de dados Neo4j.....	59
Figura 13 - Resultado da <i>cypher query</i> para consulta por classificações.....	59
Figura 14 - Resultado da <i>cypher query</i> para consulta por regiões.	60
Figura 15 - Resultado da <i>cypher query</i> para consulta dos atributos de determinada região.	60
Figura 16 - Resultado da <i>cypher query</i> para consulta dos relacionamentos entre atributos de determinada região com as classificações.	61
Figura 17 - Apresentação de todos os dados resultantes da GEOBIA armazenados no banco de dados Neo4j, com as classificações (baixo, médio e alto) e os relacionamentos entre dados.....	62
Figura 18 - Tela do programa para consultas das regiões com atributos semelhantes.	63
Figura 19 - Imagem classificada pelo agrupamento dos atributos pelo banco de dados Neo4j.	66
Figura 20 - Comparação das classificações supervisionadas com a classificação por meio do agrupamento de atributos no banco de dados NoSQL orientado a grafo.	67
Figura 21 - Resultados do cálculo da exatidão global e do índice kappa calculados por meio da matriz de confusão.	70
Figura 22 - Resultados dos cálculos de erros e omissões gerados pela matriz de confusão.....	71

LISTA DE QUADROS

Quadro 1 - Descrição dos atributos espaciais com suas respectivas fórmulas.	28
Quadro 2 - Descrição dos atributos espectrais.....	29
Quadro 3 - Descrição dos atributos de textura.	30
Quadro 4 - Composições para definição das classes da fotointerpretação.....	39
Quadro 5 - Resultados das consultas realizadas no banco de dados NoSQL por meio do agrupamento dos atributos da GEOBIA.	64

LISTA DE TABELAS

Tabela 1 - Características e especificações dos sensores Rapideye.....	19
Tabela 2 - Bandas espectrais com respectivos valores de faixa espectral da imagem Rapideye.	34
Tabela 3 - Valores de referência do índice kappa.	48
Tabela 4 - Testes de valores dos índices de escala e fusão para o algoritmo de segmentação FLSA.	53
Tabela 5 - Matrizes de Confusão geradas a partir das classificações supervisionadas, em metros.....	69
Tabela 6 - Resultado da tabulação cruzada entre a imagem classificada pelo algoritmo KNN e pelo banco de dados NoSQL orientado a grafos, em m ²	72

LISTA DE SIGLAS

AOO	Análise Orientada a Objetos
FLSA	<i>Full Lambda Schedule Algorithm</i>
FNEA	<i>Fractal Net Evolution Approach</i>
GEOBIA	Análise de Imagens Baseada em Objetos Geográficos
KNN	<i>K Neighbour Nearest</i>
MaxVer	Máxima Verossimilhança
MinDist	Distância Mínima
ML	<i>Merge Level</i>
NDVI	Índice de Vegetação por Diferença Normalizada
NDWI	Índice de Diferença Normalizada da Água
NoSQL	<i>Not Only SQL</i>
OBIA	Análise de Objetos Baseados em Imagem
PDI	Processamento de Imagem Digital
REIS	<i>Rapideye Earth Imaging System</i>
SAR	<i>Syntetic Aperture Radar</i>
SGBD	Sistemas Gerenciadores de Banco de Dados
SIG	Sistemas de Informação Geográfica
SL	<i>Scale Level</i>
SR	Sensoriamento Remoto
SVM	<i>Support Vector Machine</i>
VT	Verdade Terrestre

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	OBJETIVO GERAL.....	14
1.2	OBJETIVOS ESPECÍFICOS	14
2	REVISÃO DE LITERATURA	15
2.1	BANCO DE DADOS E SIG.....	15
2.2	BANCO DE DADOS NÃO RELACIONAL.....	16
2.2.1	Banco Não Relacional Orientado a Grafos.....	17
2.3	SENSORIAMENTO REMOTO E RECONHECIMENTO DE PADRÕES	18
2.4	PROCESSAMENTO DE IMAGENS DIGITAIS	22
2.4.1	Análise de imagens baseada em objetos geográficos.....	27
2.5	ANÁLISE MULTIVARIADA	31
3	METODOLOGIA.....	32
3.1	ÁREA DE ESTUDO.....	32
3.2	MATERIAL E MÉTODOS	33
3.3	DEFINIÇÃO DAS CLASSES	38
3.4	SEGMENTAÇÃO POR MEIO DA GEOBIA	41
3.5	APLICAÇÃO DA ANÁLISE MULTIVARIADA.....	42
3.6	BANCO DE DADOS NOSQL.....	43
3.7	ANÁLISE DOS RESULTADOS	47
3.8	VERDADE TERRESTRE.....	47
4	RESULTADOS E DISCUSSÃO	49
4.1	FOTOINTERPRETAÇÃO	49
4.2	GEOBIA.....	53
4.3	AGRUPAMENTO DE ATRIBUTOS	57
4.4	BANCO DE DADOS NOSQL NEO4J	58
4.5	ANÁLISE DAS CLASSIFICAÇÕES	63
4.6	VALIDAÇÃO DOS DADOS.....	68
5	CONCLUSÃO.....	74
5.1	TRABALHOS FUTUROS.....	75
	REFERÊNCIAS.....	76
	APÊNDICE A – CONSULTA DOS ATRIBUTOS NO BANCO DE DADOS NEO4J...81	

1 INTRODUÇÃO

O processo de classificação para o uso e ocupação do solo inclui elementos complexos para a identificação e diferenciação das diversas paisagens apresentadas. A evolução da tecnologia permite o uso de sensores remotos nos equipamentos agrícolas, tornando as técnicas mais adequadas para as atividades no campo, incluindo o manejo do solo, gerenciamento de culturas, aplicação de insumos, entre outras. Esses sensores remotos estão cada vez mais sofisticados e são capazes de capturar imagens com alto potencial para estudos e análises das superfícies, por exemplo, distinguir diferentes usos e coberturas do solo, tais como vegetação natural, culturas, corpos d' água, reflorestamento e áreas urbanas.

Com o uso das técnicas disponibilizadas nos Sistemas de Informação Geográfica (SIG) e as informações adquiridas das imagens capturadas por meio do Sensoriamento Remoto (SR), é possível criar mapas temáticos adequados aos seus propósitos com precisão exigida quanto à escala de representação.

A classificação em imagens orbitais de baixa resolução espacial, o SR utiliza o *pixel* como elemento mínimo de imagem para a classificação digital, entretanto, para imagens de alta resolução espacial é possível realizar a classificação, porém os resultados podem ser diferente do esperado quanto a separação das classes de uso e cobertura do solo, devido à complexidade e quantidade de dados que estas imagens fornecem.

A técnica de Análise de Imagens Baseada em Objetos Geográficos (GEOBIA) é a metodologia adequada para estudos em imagens de alta resolução. Essa técnica segmenta a imagem, mediante análise do *pixel* e sua vizinhança em função dos atributos espaciais (forma e tamanho), espectrais (cor e textura) e topologia (adjacência e contingência), possibilitando a classificação do uso e ocupação do solo mediante nível de abstração de regiões e não do *pixel*.

Como resultado da segmentação por meio da GEOBIA tem-se a imagem segmentada e banco de dados georreferenciado, que contém o vetor com as regiões geradas e seus respectivos atributos (espaciais, espectrais e textura) extraídos dessa região. Esse banco de dados relacional pode ser importado e processado em Sistemas Gerenciadores de Banco de Dados (SGBD) e também em SIGs.

A quantidade de dados extraídos e armazenados no banco de dados varia conforme o número de regiões geradas, pois cada região possui seus atributos relacionados, assim, quanto maior a quantidade de dados maior a demanda de armazenamento e manipulação dos mesmos dentro do banco de dados. O sistema de banco de dados relacional trabalha de forma eficiente com esses dados em SIG, porém há alternativas de sistemas de armazenamento que se destacam nas pesquisas atuais, como por exemplo, a utilização de bancos de dados *Not Only SQL* (NoSQL).

Bancos de dados NoSQL possuem diferentes estruturas para armazenar e organizar os dados, adequando os dados à aplicação desejada. O banco de dados NoSQL orientado à grafos utiliza da estrutura baseada em relacionamentos entre dados, facilitando às aplicações que necessitam de buscas complexas entre os dados armazenados.

O uso de NoSQL e SIG visam atender a demanda do volume de dados gerados, desde os dados gerados por meio de sensores remotos até processamentos que dependem de buscas mais complexas dentro do banco de dados, como a classificação digital de imagens.

Portanto, esta pesquisa teve como objetivo obter a classificação de uso e cobertura do solo em imagem orbital Rapideye, coletada por meio de SR e segmentada pelo método GEOBIA, utilizando o banco de dados NoSQL orientado a grafos para análise dos atributos espaciais, espectrais e textura. A validação do resultado foi realizada com a classificação por meio dos algoritmos de Distância Mínima (MinDist), Máxima Verossimilhança (MaxVer) e classificação por regiões pelo algoritmo Vizinho mais próximo (KNN), por meio da matriz de confusão e tabulação cruzada dos dados.

1.1 OBJETIVO GERAL

O objetivo geral desta pesquisa é classificar o uso e cobertura do solo em imagem orbital de alta resolução Rapideye, mediante GEOBIA e banco de dados NoSQL orientado a grafos.

1.2 OBJETIVOS ESPECÍFICOS

- Realizar levantamento do estado da arte das metodologias existentes em GEOBIA e banco de dados NoSQL;
- Segmentar a imagem orbital extraindo sua topologia e seus atributos, mediante GEOBIA;
- Selecionar ou reduzir a dimensionalidade dos atributos por Análise Multivariada, para agrupamento dos dados;
- Classificar o uso e cobertura do solo por meio dos métodos de classificação *pixel a pixel* pelo algoritmo de Distância Mínima (MinDist), Máxima Verossimilhança (MaxVer), classificação por regiões pelo algoritmo vizinho mais próximo (KNN) e por agrupamentos no banco de dados NoSQL orientado a grafos.

2 REVISÃO DE LITERATURA

2.1 BANCO DE DADOS E SIG

Os sistemas de bancos de dados são empregados para armazenamento e gerenciamento de dados nos mais diversos tipos de aplicação, em várias áreas do conhecimento.

Nas últimas décadas, os bancos de dados relacionais são capazes de trabalhar de maneira eficiente com dados geográficos. Com essa integração, os usuários de SIG passaram a utilizar o banco relacional, pois a maior parte das ferramentas passou a acessar dados armazenados (QUEIROZ; MONTEIRO; CÂMARA, 2013). Com as aplicações em SIG, o uso de banco de dados relacionais tem a mesma função, ou seja, armazenar dados geográficos para possibilitar a manipulação dos mesmos dentro dos sistemas de informação geográfica.

O conceito de SIG é utilizado em sistemas que tratam computacionalmente os dados geográficos, armazena e manipula atributos descritivos, juntamente com as geometrias dos tipos de dados geográficos (CASANOVA *et al.*, 2005). O SIG tem múltiplos usos, como por exemplo, mapeamento de áreas de risco e confecção de mapas temáticos (BRITO *et al.*, 2012).

Câmara *et al.* (2005) relaciona SIG com diferentes áreas alternativas, tais como cartografia, computação nas nuvens, *webgis*, *gismobile*, ou seja, há diversificação na oferta de SIG com tecnologias complementares.

Com a junção de banco de dados relacional, os SIGs oferecem outras funções, tais como a capacidade de importação, armazenamento, tratamento, integração, processamento, recuperação, transformação, manipulação, modelagem, atualização e exibição de informações digitais georreferenciadas, mediante bases de dados (BRITO *et al.*, 2012).

Os sistemas de banco de dados relacionais, após a década de 90, incluíram os tipos de dados, operadores e mecanismos de indexação espaciais. Algumas das empresas desenvolvedoras dos sistemas de bancos de dados relacionais disponibilizaram suporte espacial para o SIG. Entre elas, surgiu o complemento para o banco de dados Oracle, conhecido como Oracle *Spatial* (ORACLE, 2016), a IBM desenvolveu o DB2 *Spatial Extender* (IBM, 2002) e a Microsoft introduziu suporte espacial ao SQL Server (MICROSOFT, 2008). A comunidade de software livre

desenvolveu a extensão PostGIS, para o banco de dados PostgreSQL (POSTGIS, 2005).

Com as novas tecnologias computacionais e novos recursos para aquisição de dados, a quantidade de dados gerados diariamente aumentou em diversos domínios de aplicações, por exemplo, redes sociais, rede de sensores e *Web 2.0* (DATE, 2004). Assim, surgiram desafios para os bancos de dados relacionais manterem o armazenamento e manipulação dos dados, juntamente com o processamento das consultas realizadas.

Devido a isso, desenvolveram-se tecnologias de banco de dados com modelos e recursos diferenciados dos bancos relacionais existentes. De acordo com Queiroz; Monteiro; Câmara (2013) esse assunto, ainda, é pouco explorado no universo dos SIGs. Esse novo conceito de banco de dados é denominado banco de dados não relacional (NoSQL), tratado na seção 2.2.

2.2 BANCO DE DADOS NÃO RELACIONAL

Diferentemente dos bancos de dados relacionais, que constituem conjuntos organizados de dados, relacionados entre si e criados com objetivos para atender às necessidades dos usuários, os bancos de dados não relacionais, denominados *Not only SQL* (NoSQL), que significa “Não Apenas SQL” (POKORNY, 2013), vieram contribuir com aplicações nas diversas áreas, com volume de dados crescendo constantemente.

Segundo Almeida e Brito (2012), as ferramentas que usam o modelo de banco de dados relacional exigem que a estrutura dos dados seja pré-definida, o que necessita de conhecimento técnico e específico. Os bancos de dados NoSQL permitem diversas opções de estruturas de dados para que os dados sejam armazenados e manipulados, de acordo com a necessidade da aplicação. Os NoSQL atendem aos requisitos de alta escalabilidade, necessários para gerenciar grandes quantidades de dados, garantindo a disponibilidade dos mesmos (LÓSCIO; OLIVEIRA; PONTES, 2011).

Os bancos de dados NoSQL Neo4j (HUNGER, 2010), Apache CouchDB (ANDERSON; LEHNARDT; SLATER, 2010), MongoDB (CHODOROW; DIROLF,

2010) e Apache Cassandra (LAKSHMAN; MALIK, 2010) fornecem modelos de dados, linguagens de consulta e interfaces de programação específicos.

Queiroz, Monteiro e Câmara (2013), pesquisadores sobre banco de dados, membros das comunidades científicas, principalmente de Astronomia e Sensoriamento Remoto, buscam criar sistemas adequados às suas comunidades. Nas pesquisas incluem-se os novos modelos de dados e de linguagens de consultas, para responder a necessidade de armazenamento, processamento e análise do aumento da quantidade de dados, provenientes da evolução dos equipamentos de coleta de dados, como, telescópios, sensores a bordo de satélites, Geo-Sensores e GPS.

A importância do uso dos bancos NoSQL é atender a crescente demanda de dados gerados. O fato de possuir estrutura diferenciada faz que adapte o sistema de informação ao banco de dados que melhor represente os dados.

Os dados geoespaciais são criados, armazenados e utilizados em grande quantidade. O grande volume de dados gerados por meios dos satélites e serviços de localização, torna-se desafio para os bancos de dados comumente usados. Já os modelos NoSQL apresentam soluções eficientes para essa finalidade (LIZARDO; MORO; DAVIS JR, 2014).

2.2.1 Banco Não Relacional Orientado a Grafos

Os bancos de dados não relacionais orientados a grafos são baseados na teoria dos grafos, resultando no NoSQL orientado a grafos. Com estrutura de dados armazenados em forma de grafo, cada dado inserido armazena-se como nó e os relacionamentos como arestas. Tanto os nós como as arestas possuem propriedades, denominados atributos ou relacionamentos, possibilitando a navegabilidade entre os dados (ABREU *et al.*, 2012).

O percurso de busca em grafo facilita as consultas e manipulações dos dados. A busca desejada percorre apenas os nós relacionados, definindo a profundidade da busca realizada, reduzindo o tempo de processamento e conseqüentemente o tempo de resposta (MATSUSHITA & NGUESSAN, 2011).

A principal fonte geradora de dados e informações que alimentam banco de dados geográficos integrados em SIG são os dados provenientes do sensoriamento

remoto (BACANI; SAKAMOTO; LUCHIARI, 2015). A partir desses dados é possível realizar diversas análises dentro do contexto desejado, por exemplo, reconhecer padrões em imagens digitais.

2.3 SENSORIAMENTO REMOTO E RECONHECIMENTO DE PADRÕES

O Sensoriamento Remoto (SR) consiste na utilização de sensores distantes do objeto, ou seja, sem contato com o objeto, para coleta de dados de determinada área, realizado por câmeras, escâneres, lasers, dispositivos lineares e/ou matriciais localizados em plataformas, como os satélites (JENSEN, 2009).

Imagens capturadas por satélites possuem características resolutivas determinantes para extrair informações da superfície e identificar padrões na imagem, essas características variam de acordo com o tipo de sensor utilizado. Os sensores de satélites Rapideye assumiram grande importância para os usuários de SR e aumentaram a demanda de imagens Rapideye, devido a alta disponibilidade de dados associada às características geométrica, radiométrica e espectral das imagens (SILVA; EDUARDO; FAZAN, 2013).

O sistema de satélites com sensores Rapideye é composto por cinco satélites equipados com sensores idênticos e localizados na mesma órbita (PLANET, 2017). As características de imagens de satélites com sensores Rapideye são apresentadas na Tabela 1.

Tabela 1 - Características e especificações dos sensores Rapideye.

Características	Informação										
Número de satélites	5										
Órbita	Heliossíncrona com 630 km de altitude										
Passagem pelo Equador	±11:00 h em hora local										
Tipo do sensor	Imageador multiespectral pushbroom										
Bandas espectrais / Faixa do espectro (nm)	<table> <tr> <td>Azul</td> <td>440 – 510 nm</td> </tr> <tr> <td>Verde</td> <td>520 – 590 nm</td> </tr> <tr> <td>Vermelho</td> <td>630 – 685 nm</td> </tr> <tr> <td><i>Red Edge</i></td> <td>690 – 730 nm</td> </tr> <tr> <td>Infravermelho próximo</td> <td>760 – 850 nm</td> </tr> </table>	Azul	440 – 510 nm	Verde	520 – 590 nm	Vermelho	630 – 685 nm	<i>Red Edge</i>	690 – 730 nm	Infravermelho próximo	760 – 850 nm
Azul	440 – 510 nm										
Verde	520 – 590 nm										
Vermelho	630 – 685 nm										
<i>Red Edge</i>	690 – 730 nm										
Infravermelho próximo	760 – 850 nm										
Espaçamento de <i>pixel</i> (nadir)	6,5 m										
Tamanho da imagem	77 km										
Bits de quantização	12-bit										
Datum horizontal	WGS84										

Fonte: Adaptado de Planet (2017).

Os satélites Rapideye possibilitam melhores resoluções espaciais e temporais, foram os primeiros satélites a fornecerem imagens na região da borda do vermelho, conhecida como *red edge* (690-730nm), sensível a clorofila, ou seja, propícia para a determinação fotossintética da vegetação, por meio dos sensores REIS (Rapideye *Earth Imaging System*) (SOUZA *et al.*, 2015). O uso dessa banda espectral melhora os resultados da classificação do uso e ocupação do solo (SCHUSTER; FÖRSTER; KLEINSCHMIT, 2012), (SOUZA *et al.*, 2015) e (COSTA *et al.*, 2016).

Schuster, Förster e Kleinschmit (2012) compararam o potencial da banda *red edge* com objetivo de melhorar as classificações de uso do solo. Os algoritmos de classificação supervisionada usados foram *Vector Machine* e Máxima Verossimilhança. Obtiveram resultados positivos e indicaram que a inclusão da informação dessa banda aumenta a precisão da classificação. Os efeitos observados foram para classes de vegetação em paisagens abertas e vegetação do arbusto.

Souza *et al.* (2015) utilizaram a banda *red edge* do sensor REIS na confecção de índices de vegetação, para realizar a discriminação de diferentes tipos de cobertura vegetal. No resultado, a banda mostrou-se mais sensível para diferenciar espécies de plantas do que a banda do Infravermelho Próximo (NIR). As diferentes combinações de bandas, juntamente com os índices de vegetação, para auxiliaram para distinguir e mapear tipos específicos de vegetações.

Costa *et al.* (2016) analisaram as características das imagens Rapideye e avaliaram sua qualidade quanto a ortorretificação e mapeamento de uso do solo. As imagens passaram por correção atmosférica e foram avaliadas pela resposta espectral a diferentes alvos, acurácia da imagem e exatidão da classificação supervisionada. Como resultado a correção atmosférica possibilitou melhor avaliação da resposta espectral. Já a classificação supervisionada quantificou áreas e percentagem de cada classe de uso do solo de forma consistente.

As imagens capturadas possibilitam a classificação de imagens, que em SR é utilizada para criação de mapas temáticos. Mapear as áreas da superfície terrestre é associar cada *pixel* ou região da imagem a determinada classe que descreve algum tipo de cobertura e uso do solo, tais como agricultura, reflorestamento, vegetação, tipos de solos, área urbana, dentre outros, com a intenção de representar a distribuição geográfica de temas, por exemplo, uso e cobertura do solo, vegetação, relevo, fatores de erosão, capacidade de uso (RIBEIRO; BAHR; CENTENO, 2002).

Classificação de imagens é o processo de extração de informações de imagens com a finalidade de reconhecer padrões. Esse processo pode ser realizado por dois métodos, conhecidos como classificação supervisionada e não supervisionada. A classificação não supervisionada requer do analista apenas alguns parâmetros tais como número de classes e interações. O algoritmo permite agrupar *pixels* com características espectrais similares.

A classificação supervisionada consiste na classificação controlada pelo analista da área a ser classificada. Baseando-se na fotointerpretação o analista coleta amostras (*pixels*) na imagem e determina as classes para treinar o classificador, ou seja, categoriza o *pixel* especificando para o algoritmo os valores dos atributos espectrais de cada tipo de cobertura do solo de interesse na imagem.

Cada *pixel* no conjunto de dados não utilizado para treinamento é comparado com esses valores e então classificado com o nome da classe mais similar.

Esse método pode ser realizado de duas maneiras, *pixel a pixel* ou orientada ao objeto. A classificação *pixel a pixel* analisa apenas as características de cada *pixel* e os categoriza de acordo com métodos estatísticos, já a classificação orientada ao objeto analisa o *pixel* e sua vizinha, formando os objetos ou regiões da imagem, que são o conjunto de *pixels* com características semelhantes.

Os autores Soares, Santos e Silva (2016) compararam os métodos de classificação de imagens de SR, objetivando obter mapeamentos capazes de proporcionar informações detalhadas quanto ao uso do solo. Utilizaram imagens dos satélites *Landsat 5 TM* e *Rapideye*, os métodos de classificação supervisionada e não supervisionada. Os resultados apresentados foram satisfatórios, pois possibilitou determinar a classificação supervisionada mais próxima da validação dos dados.

Os algoritmos classificadores que se baseiam na distância, como por exemplo, os métodos da Distância Mínima (MinDist), paralelepípedo e distância de Mahalanobis utilizam principalmente valores médios espectrais de classes distintas, ignorando valores de variância. Classificadores como o método de Máxima Verossimilhança (MaxVer) incorporam tanto a média como a variância do conjunto de dados coletados para a regra de decisão de classificação (NERY *et al.*, 2013).

Os autores Cattani *et al.* (2013) compararam quatro métodos de classificação supervisionada e analisaram o potencial individual de cada classificador para mapeamento de culturas agrícolas a partir de imagens *Rapideye*. Os classificadores usados foram MinDist, Distância de Mahalanobis, MaxVer e Mapeador de Ângulo Espectral. Ambos classificadores apresentaram variações significativas quanto ao mapeamento de uso e ocupação do solo. Porém o classificador MaxVer apresentou excelentes resultados.

Acharya, Yang e Lee (2016) também compararam os métodos MinDist, Distância de Mahalanobis, MaxVer e *Support Vector Machine* (SVM) para determinar o uso e cobertura do solo, utilizando imagem de alta resolução *KOMPSAT-3A*. Para obter resultados mais precisos na classificação, calcularam o Índice de Vegetação por Diferença Normalizada (NDVI) e Índice de Diferença Normalizada da Água (NDWI). Os classificadores de Distância de Mahalanobis e SVM apresentaram

resultados semelhantes e precisos para o uso do solo. Diferentemente, MinDist e MaxVer resultaram em classificação mais pobre, porém destacando melhor a classe pertencente à vegetação.

Na classificação de imagens é importante analisar as características da imagem e realizar o processamento de imagens, com a intenção de obter resultados significativos e determinantes para o reconhecimento dos padrões.

2.4 PROCESSAMENTO DE IMAGENS DIGITAIS

O Processamento de Imagens Digitais (PDI) é o conjunto de técnicas para capturar, representar e transformar imagens com o auxílio da computação. Essas técnicas permitem extrair e identificar informações das imagens, e também melhorar a qualidade visual de aspectos estruturais, o que facilita a percepção humana e a interpretação automática por meio de máquinas (GONZALEZ; WOODS, 2010).

Moreira (2007) cita que os primeiros sistemas desenvolvidos contavam com a finalidade de análise de dados coletados por sensores, basicamente para agilizar as tarefas manuais de interpretação visual e a possibilidade de introduzir outros tipos de informações para cruzá-las com padrões espectrais, facilitando a identificação.

O PDI envolve algumas etapas, que tem como objetivo encontrar resultados para determinados problemas, como por exemplo, a classificação de imagens. Iniciando pela etapa da aquisição de imagens por meio dos recursos do SR, as imagens digitais são capturadas em várias regiões do espectro eletromagnético. Depois de adquiridas, são inseridas no banco de dados dentro dos SIGs, formando a base de conhecimento para a realização demais etapas presentes no PDI. A interpretação visual das imagens, ou fotointerpretação, permite reconhecer objetos, representar e descrever as classes presentes na imagem.

A fotointerpretação consiste na extração de informações de alvos da superfície terrestre, baseados nas respostas espectrais observadas nas imagens (GONZALEZ; WOODS, 2010). Nesse processo o analista observa e interpreta os objetos presentes na imagem, considerando as cores, textura, formato e conhecimento prévio para delimitar e categorizar as áreas referentes a cada uso do solo. Esse tipo de classificação possui alta confiabilidade, mas se torna inviável para a análise de áreas grandes e heterogêneas.

A interpretação dos dados de SR via computador chama-se análise quantitativa, pela função da habilidade em identificar *pixels* e suas propriedades. A implementação de técnicas de análise quantitativa nos sistemas disponíveis complementam os processos interativos de processamento de imagens. Alguns casos de processos de análise quantitativos são baseados em classificação de imagens, outros são realces de imagens, pré-processamento como as correções radiométricas, atmosféricas e geométricas, assim como a segmentação de imagens.

Os autores Santos et al. (2013) realizaram a comparação entre classificação por fotointerpretação e por orientada a objetos. Com imagem de alta resolução IKONOS avaliaram a capacidade de confiabilidade dessas imagens. Ambas as classificações apresentaram resultados confiáveis para análises de uso do solo, porém a classificação por fotointerpretação requer conhecimento do analista e necessita de tempo para executar o processo. Já a classificação orientada a objeto apresentou maior eficiência nos resultados pelo fato de separar diferentes classes em menor tempo de execução.

Para obter o realce de imagens destacam-se os índices de vegetação, que são operações entre as bandas do espectro eletromagnético.

O Índice de Vegetação por Diferença Normalizada (NDVI) é o índice de vegetação baseado entre razão entre bandas e utiliza a diferença da reflectância entre a faixa do NIR e a reflectância da faixa do visível. Na pesquisa o NDVI é o índice de vegetação escolhido para ser utilizado, pois é bastante usado nas pesquisas de mapeamento temático, à sua importância para o monitoramento de mudanças no desenvolvimento da vegetação e a redução de ruídos presentes em múltiplas bandas das imagens (JENSEN, 2009).

A próxima etapa do PDI é a realização da segmentação, que consiste em subdividir a imagem em regiões similares, analisando seus atributos. Esse processo é feito por meio de algoritmos baseados nas propriedades básicas de valores de níveis digitais: descontinuidade e similaridade (GONZALEZ; WOODS, 2010). Algoritmos de segmentação possibilitam encontrar diferenças entre dois ou mais objetos presentes na imagem (AGUIAR *et al.*, 2007). Algumas das técnicas de segmentação mais utilizadas em PDI serão brevemente descritas a seguir.

- **Detecção de descontinuidades**

Descontinuidade é baseada em particionar a imagem conforme as mudanças bruscas nos níveis digitais. Utiliza-se essa técnica quando o objetivo é detectar bordas ou pontos isolados e é importante para que haja diferenciação entre duas ou mais regiões numa imagem, para posterior análise (AGUIAR *et al.*, 2007).

- **Detecção de bordas**

As bordas ou contornos são definidos como fronteira entre regiões cujos valores digitais são diferentes. Porém as bordas não perfeitamente caracterizadas sem que haja tratamento posterior na imagem. Isto é devido aos ruídos, não uniformes da iluminação que introduzem descontinuidades na borda (WIGGERS, 2014).

- **Detecção de similaridades**

A detecção de similaridades verifica o interior das regiões, ou seja, a similaridade entre os *pixels* vizinhos identifica a região, as regiões devem ter o grau de homogeneidade relativo as suas propriedades (WIGGERS, 2014). A característica observada nas imagens digitais é o valor do *pixel*. Para detectar as similaridades entre *pixels* pode ser utilizada alguma das técnicas a seguir descritas:

- **Limiarização**

O princípio da limiarização é separar o primeiro plano do fundo, com o objetivo de marcar os *pixels* que pertencem às verdadeiras regiões do primeiro plano e as regiões de fundo com intensidade diferente (FACON, 2002). Essa técnica é matematicamente descrita como técnica do PDI no qual a imagem de entrada $f(x,y)$ de N níveis digitais produz saída $g(x,y)$, dita limiarizada, onde quantidade de tons de cinza é menor que N . Considerada a técnica mais simples é a do particionamento do histograma da imagem por determinado valor, ou limiar, T .

- **Crescimento de regiões**

O crescimento de regiões é a técnica de agrupamento de *pixels* ou sub-regiões em regiões maiores. De acordo com Gonzalez e Woods (2010), essa técnica consiste em partir de um ou vários *pixels* considerados uniformes, analisar os *pixels*

vizinhos e acrescentar os vizinhos que forem semelhantes ao grupo de *pixels* iniciais, de modo que a região cresça formando processo iterativo. O crescimento de regiões por agregação de *pixels* com propriedades semelhantes, exemplo brilho e cor.

- **Divisão e fusão**

A técnica de divisão e fusão consiste em subdividir a imagem em conjunto de regiões distintas e fundir as regiões na tentativa de satisfazer as condições de segmentação. O início do processo é subdividir a imagem por meio da estrutura *quadrees*, que permite dividir e agrupar partes da imagem.

A segmentação é o processo de agregação dos níveis digitais e regiões. Cada região será descrita espectralmente como a média dos níveis digitais ou o valor mínimo deles, que aborda somente os dados estatísticos. Porém os métodos *Fractal Net Evolution Approach* (FNEA) e *Full Lambda Schedule Algorithm* (FLSA) utilizam além dos valores espectrais, os atributos de forma, textura e a topologia nos bancos de dados relacionais. A técnica escolhida para ser utilizada nesta pesquisa é a FLSA.

- ***Fractal Net Evolution Approach* (FNEA)**

A imagem de SR é considerada de natureza fractal, sendo a segmentação realizada na imagem de maneira fina e grosseira, originando diferentes escalas de segmentação. O segmentador ou algoritmo multiresolução FNEA permite segmentar a imagem em níveis hierárquicos diferentes, utilizando os conceitos da lógica *fuzzy* e de semântica. Essa hierarquia possibilita diferenciar classes de objetos em diferentes escalas e a semântica possibilita inserir o conhecimento do usuário (BLASCHKE; LANG; HAY, 2008).

- ***Full-Lambda Schedule Algorithm* (FLSA)**

O algoritmo FLSA foi desenvolvido por Robinson, Redding e Crisp (2002) e proposto para ser usado com imagens SAR (*Syntetic Aperture Radar*), mas permitindo aplicabilidade em várias situações, o que favoreceu tornar-se de uso comum na atualidade. Este algoritmo permite a divisão de imagens em regiões

(objetos) e responsável pela fusão das regiões adjacentes, baseando-se na combinação da informação espectral e espacial ou de textura, resultando no agrupamento de *pixels* em forma de região, que devem estar espacialmente próximos, com informações espectrais semelhantes e informar seu atributo (espectral, espacial e textura) (ROBBINSON; REDDING; CRISP, 2002).

O trabalho de Prichoa (2012) relata que pelo fato desse algoritmo ser eficiente para definição do limite de segmentação com melhores divisões da imagem em regiões homogêneas, também se trata de algoritmo iterativo e multiescala garantindo bons resultados de segmentação.

O processo necessário para o preparo das imagens digitais para futura classificação temática em nível de regiões, ou orientada a objetos, é definido como análise orientada a objetos. Os elementos analisados são regiões, ou objetos, resultantes do agrupamento de *pixels* próximos e com características espectrais similares. Esse processo é considerado como pré-classificação de imagens. Em imagens de baixa resolução espacial o método mais utilizado é a análise baseada no *pixel*.

Entretanto, em imagens de alta resolução, o método sugerido e mais indicado na literatura é a Análise de Objetos Baseados em Imagem (OBIA), devido ao fato de considerar as características das classes, como as respostas espectrais, variantes geométricas, topológicas e de textura.

Os autores Cohenca e Carvalho (2015) realizaram a análise comparativa de três métodos de classificação de imagens OLI/*Landsat-8* de área com alta diversidade de uso do solo. Os métodos utilizados foram classificação supervisionada MaxVer, por MinDist e OBIA. As classes de referência em cada amostra foram obtidas com uso de imagens Rapideye, definida visualmente pelo interprete. De acordo com os resultados obtidos, os classificadores MaxVer e OBIA tiveram melhor desempenho em comparação com o MinDist. O desempenho mais fraco do método MinDist sugere que o uso do valor médio espectral somente para classificar *pixels* individuais não é apropriado para o processo de classificação.

Juntamente com as técnicas de segmentação como processo de pré-classificação, desenvolveu-se a análise orientada a objetos geográficos, que será utilizada nesta pesquisa e apresentada a seguir.

2.4.1 Análise de imagens baseada em objetos geográficos

A Análise de Imagens Baseada em Objetos Geográficos (GEOBIA) é o termo derivado da OBIA acrescentando o prefixo “geo”, que se refere ao componente georreferenciado, com foco principal em SR e SIG (BLASCHKE; LANG; HAY, 2008). Utilizada em diversas áreas, como Biomedicina, Astronomia, Visão Computacional e Geociências e considerada abordagem útil para o mapeamento da cobertura terrestre, devido ao fato de permitir a classificação mais significativa da paisagem quando comparada com métodos tradicionais baseados em *pixels* de processamento de imagem, especialmente usando dados espaciais de alta resolução (PRADO; CARVALHO, 2016).

A GEOBIA segmenta a imagem, analisando o *pixel* e sua vizinhança em função dos atributos espaciais (forma e tamanho), espectrais (cor e textura) e topologia (adjacência e contingência), possibilitando a classificação do uso do solo, pelo nível de abstração de regiões e não do *pixel*.

Solutions (2008) realizou estudos dos atributos espaciais, espectrais e de textura. Cada atributo possui valor numérico que é obtido por meio de fórmula. Os atributos espaciais, também conhecidos como atributos de forma, são as características geométricas que são mensuradas em forma de comprimentos, áreas ou relações matemáticas entre medidas da região e representam a forma do objeto matematicamente, de modo que o computador possa interpretar. O Quadro 1 descreve quais são os atributos espaciais juntamente com a fórmula para cálculo do valor.

Quadro 1 - Descrição dos atributos espaciais com suas respectivas fórmulas.

Atributo	Descrição	Fórmula
Área (area)	Valor da área do polígono.	
Comprimento (length)	Valor do comprimento dos limites de cada região.	
Compacidade (compact)	Indica quão compacto é a região. Regiões próximas do círculo são mais compactas.	$\frac{\sqrt{4 \cdot \frac{\text{Área}}{\pi}}}{\text{Perímetro}}$
Convexidade (convexity)	Corresponde se o polígono é convexo (valor 1) ou côncavo (valor menor que 1).	$\frac{\text{Comprimento Fecho Convexo}}{\text{Comprimento}}$
Solidez (solidit)	Compara a área com a área do fecho convexo no entorno do polígono.	$\frac{\text{Área Polígono}}{\text{Área Fecho Convexo}}$
Redondeza (roundness)	Compara a área com o quadrado do diâmetro máximo do polígono. Para o círculo o valor é 1.	$\frac{4 \cdot \text{Área}}{\pi \cdot (\text{Eixo Maior})^2}$
Fator de forma (formfactor)	Quanto mais aproximado do valor 1 mais circular é a região.	$\frac{4 \cdot \pi \cdot \text{Área}}{(\text{Perímetro})^2}$
Alongamento (elongation)	Região quadrada valor 1 e retangulares maiores que 1.	$\frac{\text{Eixo Maior}}{\text{Eixo Menor}}$
Retangularidade (rect_fit)	Refere-se ao quanto o polígono se aproxima da forma retangular. Para o retângulo o valor é 1.	$\frac{\text{Área}}{\text{Eixo Maior} \cdot \text{Eixo Menor}}$
Direção principal (maindir)	Indica a direção do eixo maior do polígono em relação ao eixo x.	
Eixo maior (majaxislen)	Comprimento do eixo maior de uma caixa envolvente em torno do polígono.	
Eixo menor (minaxislen)	Comprimento do eixo menor de uma caixa envolvente em torno do polígono.	
Número de furos (numholes)	Quantidade de furos na região	
Relação area/furos (holesolrat)	Se o polígono não possuir furos o valor é 1.	$\frac{\text{Área}}{\text{Área do contorno}}$

Fonte: Adaptado de Solutions (2008).

Os atributos espectrais estão relacionados às características espectrais dos *pixels* inseridos na região, como se observa no Quadro 2. Os atributos valor mínimo, valor máximo, média dos valores e desvio padrão são calculados separados para cada banda espectral da imagem.

Quadro 2 - Descrição dos atributos espectrais.

Atributo	Descrição
Matiz (hue)	Tonalidade do sistema de cores HSI (<i>Hue, Saturation, Intensity</i>) que descreve o objeto.
Saturação (saturation)	Saturação do sistema de cores HSI que descreve o objeto.
Intensidade (intensity)	Intensidade do sistema de cores HSI que descreve o objeto.
Razão entre bandas (bandratio)	Valor determinado pela razão dos valores digitais de duas bandas espectrais.
Valor mínimo de <i>pixel</i> banda (minband)	Menor valor digital observado nos <i>pixels</i> do objeto.
Valor máximo de <i>pixel</i> banda (maxband)	Maior valor digital observado nos <i>pixels</i> do objeto.
Média dos valores digitais banda (avgband)	Valor médio dos valores digitais dos <i>pixels</i> do objeto.
Desvio padrão dos valores digitais (stdband)	Desvio padrão da média dos valores digitais dos <i>pixels</i> do objeto.

Fonte: Adaptado de Solutions (2008).

No Quadro 3 são apresentados os atributos de textura, que servem para estimar a textura da região na imagem. Para isso, analisa a distribuição ou a frequência do Nível Digital (ND) dentro da imagem.

Quadro 3 - Descrição dos atributos de textura.

Atributo	Descrição
Intervalo de Textura (tx_range)	Intervalo de valores dos <i>pixels</i> compreendidos dentro da janela.
Média de textura (tx_mean)	Média dos valores dos <i>pixels</i> compreendidos dentro da janela.
Variância da textura (tx_varianc)	Variância dos valores dos <i>pixels</i> compreendidos dentro da janela.
Entropia da textura (tx_entropy)	Valor médio da entropia dos valores dos <i>pixels</i> compreendidos dentro da janela.

Fonte: Adaptado de Solutions (2008).

Prado e Carvalho (2016) avaliaram a precisão temática e geométrica do mapeamento da cobertura do solo por meio da GEOBIA, objetivando identificar e analisar as principais fontes de erros. O mapa gerado a partir de imagem Rapideye, integrando quatro medidas de similaridade (Forma, Temática, Borda e Posição) para gerar a matriz de erro. Com a metodologia proposta concluíram que é aplicável e eficaz para validar erros de mapeamentos temáticos gerados pela GEOBIA.

Segundo os autores Folharini e Souza (2016) a GEOBIA tem importância na atualidade, pois interpreta os padrões espectrais e de forma da imagem, tornando a classificação mais acurada. Ao aplicar essa técnica em imagens de alta resolução Rapideye os resultados são mais detalhados. Em seu trabalho, avaliaram o diagnóstico do uso e ocupação da bacia do córrego Monte Alegre aplicando a GEOBIA, testando a acurácia com pontos validados em campo. Obtiveram resultados satisfatórios, com índice kappa 0,85 e exatidão global 0,86.

2.5 ANÁLISE MULTIVARIADA

A análise multivariada de agrupamento dos dados constitui-se do conjunto de técnicas estatísticas que trata diversos atributos simultaneamente. Essa técnica possibilita classificar antecipadamente atributos similares, em grupos, ou seja, a criação de grupos de atributos semelhantes baseados em características em comum (SANTOS *et al.*, 2015).

A análise do dendrograma é realizada por meio da homogeneidade dos atributos próximos e valores similares para os atributos medidos. Quanto maior a aproximação entre as medidas relativas aos atributos, maior a semelhança entre elas. Dessa forma, geram grupos de acordo com os atributos associados, como a distância encontrada entre eles (PRICHOA, 2012).

Os autores Silva e Sano (2016) analisaram imagens Rapideye para a discriminação de classes de cobertura do solo. Utilizaram a análise de agrupamento hierárquico para separar em grupos de bandas espectrais da imagem que mostravam maior potencial para definir as classes de uso do solo. Os resultados obtidos pelo dendrograma permitiu separar em grupos de bandas para composição da imagem para a classificação, que obteve exatidão global acima de 80%.

Maciel e Vinhas (2015) utilizaram algoritmos de agrupamento para a separação de culturas agrícolas de tipos de uso e cobertura do solo, considerando diferentes conjuntos de atributos radiométricos. Comparando com outros métodos de seleção de dados, por meio do dendrograma obtiveram o agrupamento de atributos com maior homogeneidade que melhor separou as culturas agrícolas.

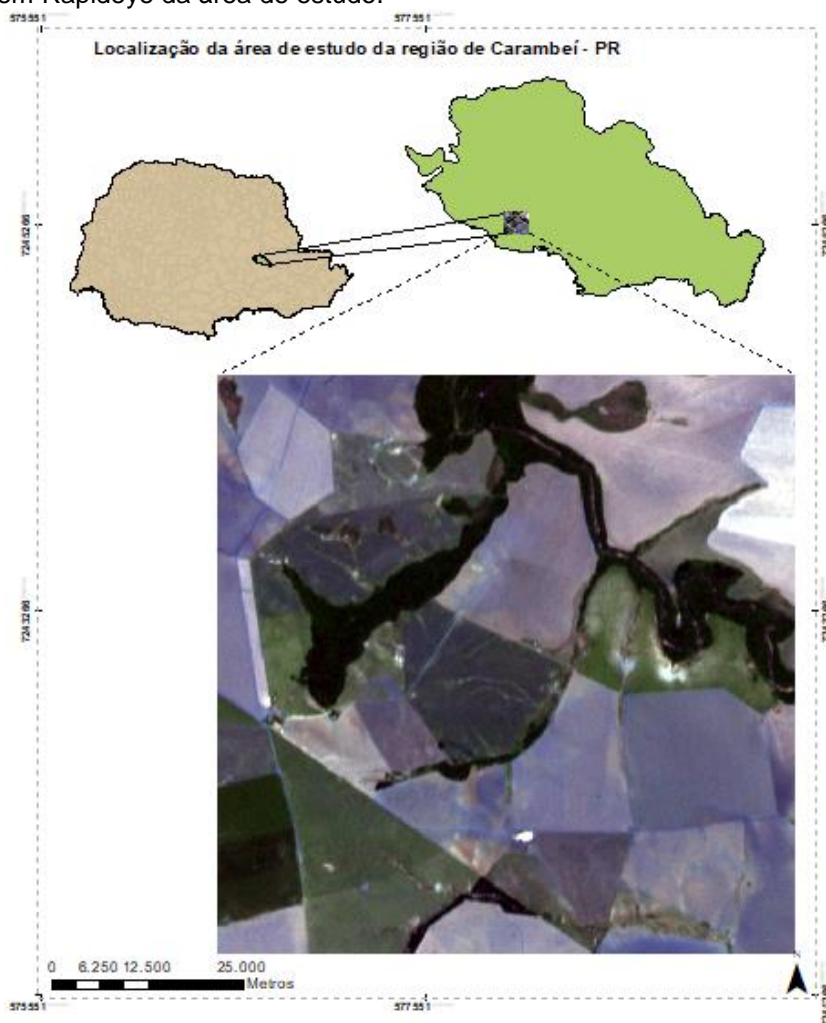
3 METODOLOGIA

3.1 ÁREA DE ESTUDO

A área de estudo situa-se na região de Carambeí, no estado do Paraná. Está localizada entre as latitudes $24^{\circ}54'36.20''$ e $24^{\circ}56'33.90''$ S e as longitudes $50^{\circ}12'35.73''$ e $50^{\circ}14'44.80''$ O. Atravessando a essa área tem-se o rio Pitanguí que é o principal da bacia hidrográfica da região, pois forma o reservatório de Alagados, utilizado para produção de energia hidrelétrica e abastecimento de água para o município de Ponta Grossa.

Na Figura 1 encontra-se o mapa do Paraná, destacando a região de Carambeí, com enfoque para a imagem Rapideye na verdadeira cor (composição azul, verde e vermelho) da área utilizada para estudo.

Figura 1 - Imagem Rapideye da área de estudo.



Fonte: A autora.

3.2 MATERIAL E MÉTODOS

Hardware utilizado:

- Notebook Dell Inspiron 15, com processador Intel(R) Core(TM) i5-4221U CPU 1.7GHz, 4GB de memória RAM, sistema operacional Windows 8.1 de 64 bits.

Softwares utilizados:

- Neo4j 3.0.1: banco de dados NoSQL orientado a grafos desenvolvido pelo *Neo Technology, Inc.* Utilizado para armazenamento e manipulação dos dados.
- Visual Studio 2015 Community: ambiente de desenvolvimento integrado da Microsoft. Utilizado para desenvolvimento dos softwares, na linguagem de programação cSharp (c#), para inserção e consulta dos dados no banco de dados Neo4j.
- Arcgis 10.3: versão lançada em 2004, utilizada para visualização dos arquivos *shapefile*, criação do mapa da área de estudo.
- Envi 4.7: usado para processamento das imagens digitais, como recortes e extração dos atributos espaciais, espectrais e de textura.
- Statistica 10: usado para normalização dos dados, cálculos estatísticos e agrupamento dos dados.

Imagem orbital utilizada:

Imagem capturada por satélite com sensor Rapideye, disponibilizada pelo curso de Pós-Graduação em Geografia (Mestrado e Doutorado) da Universidade Estadual de Ponta Grossa, obtida em agosto de 2011, com as seguintes características de resolução:

- Espacial de 5 metros;
- Radiométrica de 12 bits;
- Temporal de 1 dia.

A imagem orbital Rapideye da área de estudo é composta por 5 bandas espectrais que estão apresentadas na Tabela 2 com os respectivos valores da faixa espectral em que cada banda se encontra.

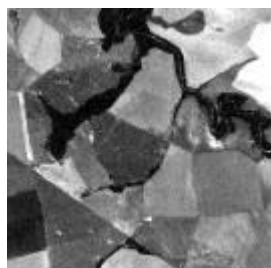
Tabela 2 - Bandas espectrais com respectivos valores de faixa espectral da imagem Rapideye.

Número	Faixa espectral (nm)	Banda
1	440 – 510	Azul
2	520 - 590	Verde
3	630 - 685	Vermelho
4	690 - 730	<i>Red-Edge</i>
5	760 - 850	Infravermelho próximo

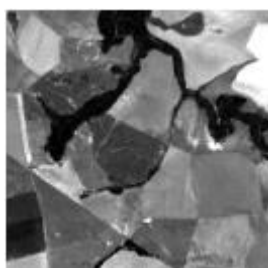
Fonte: (PLANET, 2017).

A Figura 2 mostra as bandas espectrais monocromáticas da imagem Rapideye, onde cada banda corresponde ao comprimento de onda do espectro eletromagnético, citado anteriormente. Com base nessas bandas espectrais apresentadas é possível fazer as composições coloridas delas, utilizando o sistema de cores RGB (*red, green, blue*), para realizar a fotointerpretação com diferentes composições que realçam diferentes feições em cada banda da imagem da área de estudo.

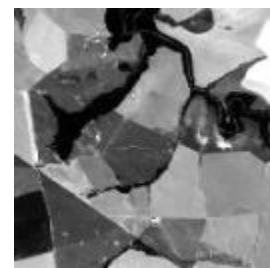
Figura 2 - Representação das bandas espectrais na escala cinza da imagem Rapideye da área de estudo.



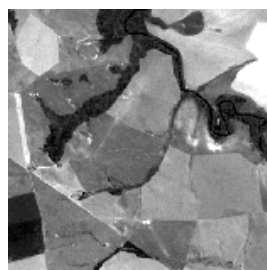
(a) Banda 1 – Azul



(b) Banda 2 – Verde



(c) Banda 3 – Vermelho



(d) Banda 4 – *Red-edge*

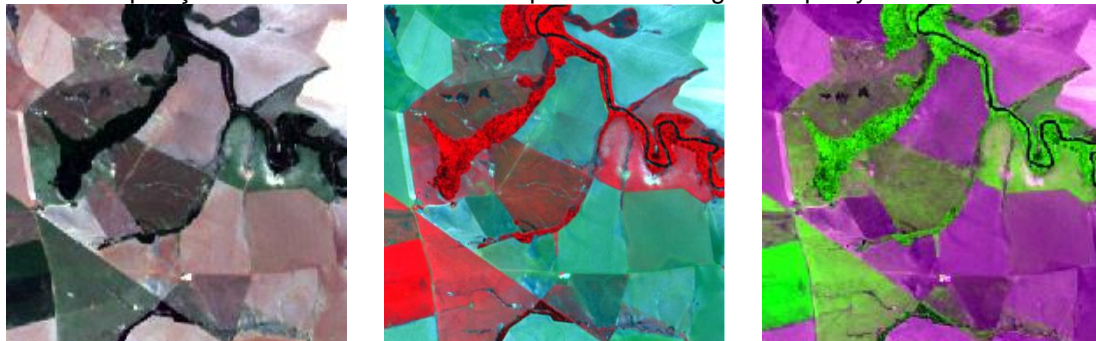


(e) Banda 5 – Infravermelho próximo

Fonte: A autora.

Diferentes composições coloridas formadas pelas bandas espectrais da imagem Rapideye são apresentadas na Figura 3.

Figura 3 - Composições coloridas das bandas espectrais da imagem Rapideye.



(a) Bandas 3R 2G 1B

(b) Bandas 5R 3G 1B

(c) Bandas 4R 5G 3B

Fonte: A autora.

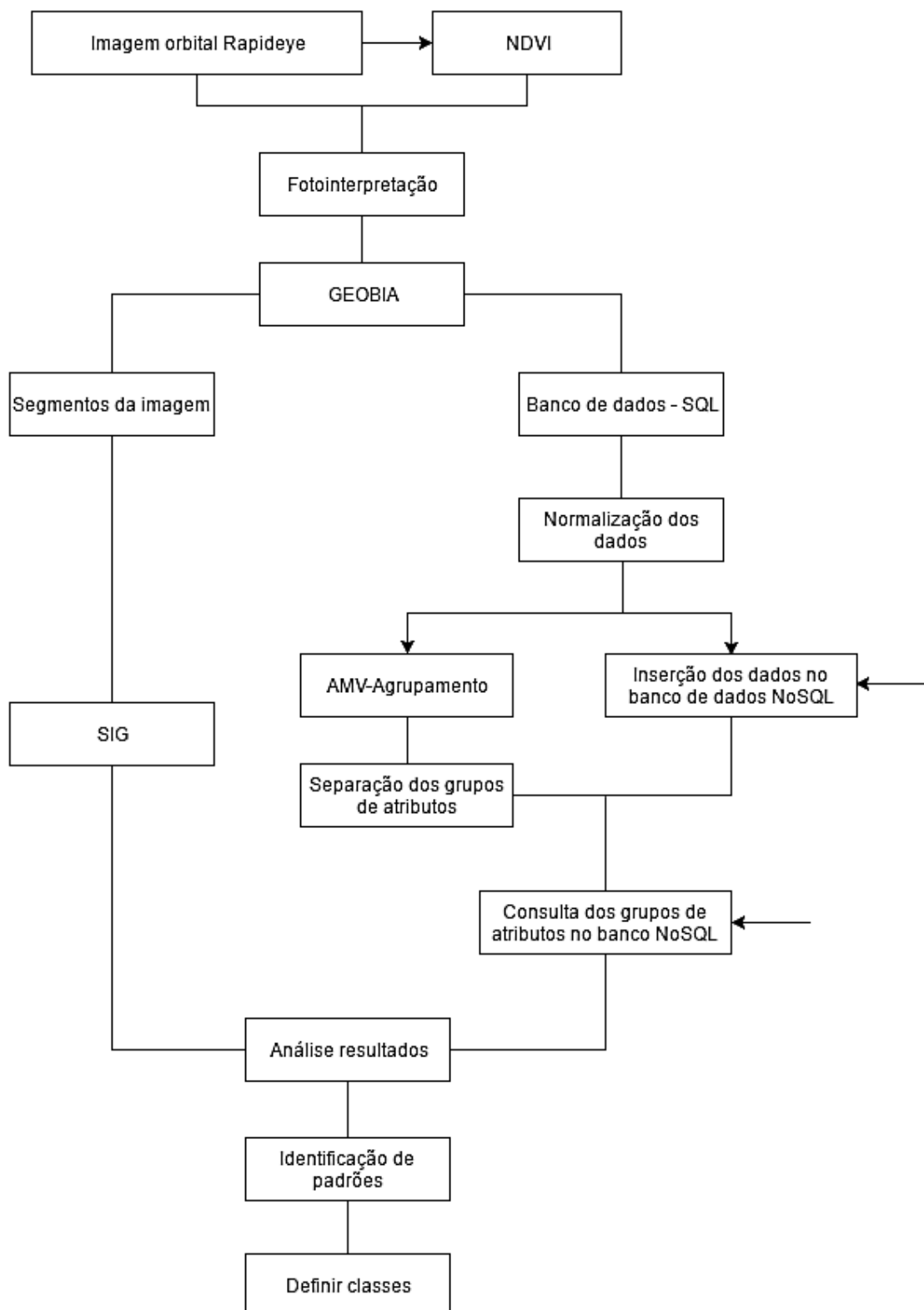
Nota-se que na Figura 3a é composta pela banda 3 na cor vermelha, banda 2 na cor verde e a banda 1 na cor azul, essa composição chama-se verdadeira cor, pois é a representação real das cores presentes na imagem, ou seja, como é visto pelo olho humano. Nesta, é possível identificar a maioria das classes para a classificação, como as classes de vegetação e cultivos.

A composição da Figura 3b é feita pelas bandas do NIR, vermelho e azul. Essa combinação de cores é chamada de falsa cor, devido ao fato de não representar as cores reais e sim as cores resultantes da composição de bandas. Nota-se que a mata ciliar (na cor vermelha) destaca-se das demais classes de cultivo.

Já na Figura 3c, também na composição falsa cor, formada pelas bandas do *red edge*, NIR e vermelho, é possível identificar claramente a classe pertencente ao rio e evidencia também a diferença entre as classes de cultivo. A partir de várias composições entre bandas percebe-se a importância dessa análise como auxílio para definição das classes durante o processo de fotointerpretação.

Para esta pesquisa a metodologia proposta e desenvolvida é explicada por meio do diagrama de blocos mostrada na Figura 4. O diagrama exemplifica as etapas realizadas na pesquisa.

Figura 4 - Diagrama de blocos para a metodologia proposta.



Fonte: A autora.

Iniciando pela imagem orbital Rapideye escolhida para a área de estudo, com as 5 bandas espectrais, aplica-se o NDVI, utilizado pelos autores Souza *et al.* (2015) com a intenção de realçar as feições e realizar a diferenciação de tipos diferentes de cobertura vegetal. Com a aplicação do índice de vegetação, a imagem fica composta com 6 bandas espectrais, sendo a banda 6 artificial correspondente ao resultado do cálculo do NDVI, para o realce das feições. Com a composição entre bandas é possível o reconhecimento das classes de interesse para executar o processo de classificação por fotointerpretação, baseada no Google Earth.

Como mostra o diagrama, para a segmentação da imagem Rapideye com as 6 bandas espectrais é aplicado por meio da técnica da GEOBIA (PRADO; CARVALHO, 2016). Nesse processo ocorre a extração das características espaciais, espectrais e de textura da imagem dividindo-a em segmentos ou regiões. A segmentação resulta no banco de dados com a topologia das regiões formadas juntamente com os atributos (espaciais, espectrais e textura) e no arquivo *shapefile* que contém o vetor de regiões geradas.

O arquivo *shapefile* com as regiões segmentadas da imagem é inserido no SIG para posteriormente analisar os resultados das classificações e reconhecer as classes. Simultaneamente, ocorre a normalização dos dados presentes no banco de dados relacional proveniente da segmentação. Com os dados normalizados é possível realizar o processo de agrupamento dos atributos por meio da análise multivariada e a inserção dos dados no banco de dados NoSQL.

A análise multivariada de agrupamentos tem a finalidade de agrupar dados e apresentar por meio do dendrograma os grupos formados. Ao obter o dendrograma dos atributos resultantes da GEOBIA, separa os grupos com o propósito de usá-los para consultas no banco de dados NoSQL.

Os autores Queiroz, Monteiro e Câmara (2013) utilizaram banco de dados NoSQL para atender a necessidade de armazenar, processar e analisar dados coletados por meio do SR e integrados com SIG. Portanto, Aqui neste estudo, definiu o banco de dados NoSQL orientado a grafos para armazenamento e manipulação dos dados já normalizados.

Com os grupos de atributos formados e o banco de dados NoSQL com os dados já inseridos, possibilita consultar as regiões semelhantes com os atributos pertencentes ao grupo selecionado, dentro do grafo gerado pelo banco de dados

NoSQL. Os resultados das consultas são regiões similares encontradas no banco de dados. Juntando os resultados das consultas com os segmentos da imagem previamente inseridos no SIG realiza a análise dos resultados, identificando padrões nas regiões semelhantes e definindo-as como classes.

Ao identificar e definir as classes resultantes das consultas no banco de dados NoSQL obtém-se a imagem classificada mediante aos grupos de atributos formados pelo grafo. Assim, sendo possível comparar essa nova imagem classificada com a imagem anteriormente classificada pela fotointerpretação e pelo método de classificação supervisionada escolhido.

Comparando as imagens classificadas observa-se a relação entre os atributos espaciais, espectrais e textura resultantes da segmentação inseridos em banco de dados NoSQL orientado a grafos, com o intuito de encontrar classes de interesse e reconhecer padrões para analisar o uso e ocupação do solo da área de estudo.

3.3 DEFINIÇÃO DAS CLASSES

Primeiramente, definiu a imagem orbital Rapideye da área de estudo, composta pelas 5 bandas espectrais correspondentes as faixas do azul, verde, vermelho, *red edge* e NIR. Utiliza-se o software Envi 4.7, desenvolvido pela empresa ITT VIS (*Visual Information Solutions*) para calcular o índice de vegetação.

O índice de vegetação escolhido é o NDVI, pois reduz efeitos de iluminação, destaca as feições não evidenciadas pelas bandas espectrais e diferencia classes de vegetação de interesse para o trabalho, assim como delimitador para a classe do rio. Esse índice é definido pela razão entre bandas espectrais, faixa do NIR e do vermelho, como mostra a equação (1):

$$NDVI = \frac{NIR - Vermelho}{NIR + Vermelho} \quad (1)$$

Após o cálculo é realizada a aplicação do NDVI na imagem de estudo, que fica composta por 6 bandas espectrais, sendo a banda 6 artificial representada pelo índice de vegetação.

O processo de fotointerpretação teve como base a imagem da área de estudo localizada pelo Google Earth, para melhor visualização da área e diferenciação das classes de uso e ocupação do solo. O Quadro 4 apresenta as classes de interesse definidas no processo de fotointerpretação, com as composições formadas pelo sistema de cores RGB, que correspondem ao vermelho, verde e azul com as bandas espectrais da imagem.

Quadro 4 - Composições para definição das classes da fotointerpretação.

Classe	Composição 3R2G1B	Composição 5R3G1B	Composição 4R5G3B
Rio			
Mata ciliar			
Solo exposto			
Cultivo 1			
Cultivo 2			
Cultivo 3			
Cultivo 4			
Cultivo 5			

Fonte: A autora.

A composição das bandas do vermelho, verde e azul no sistema RGB refere-se à verdadeira cor, ou seja, a cor real da imagem. Nesse processo definiram-se 8 classes de interesse, sendo elas: rio, mata ciliar, solo exposto, cultivo 1, cultivo 2, cultivo 3, cultivo 4 e cultivo 5, também verificadas no Google Earth.

Durante a fotointerpretação analisa quais as classes presentes na imagem e define estas como classes de interesse para realizar a classificação de imagens (COHENCA; CARVALHO, 2015). Classificar a imagem é importante nesta pesquisa para usá-la como parâmetro de comparação com as classes resultantes das consultas por meio do banco de dados NoSQL. Dessa forma, realiza-se a classificação supervisionada da imagem por meio do software Envi 4.7, utilizando os classificadores *pixel a pixel* pelo algoritmo de MinDist e MaxVer (NERY *et al.*, 2013).

Os autores Cattani *et al.* (2013) e Soares, Santos e Silva (2016) comparam esses métodos de classificações em imagem Rapideye, mostrando que são eficientes para mapear as áreas de interesse na cobertura do solo.

Pelo fato dos sensores Rapideye disponibilizam de imagens de alta resolução, com alto nível de detalhes e a presença da banda *red edge* (SCHUSTER; FÖRSTER; KLEINSCHMIT, 2012), (SOUZA *et al.*, 2015) e (COSTA *et al.*, 2016), é possível realizar a classificação orientada a objetos, com o algoritmo KNN. Prado e Carvalho (2016) avaliaram a eficiência desse método para classificação utilizando características não apenas do *pixel*, mas do *pixel* e sua vizinhança.

Para ambas as classificações, as amostras de treinamento foram coletadas. Definiu-se a quais classes cada amostra pertence, considerando a Verdade Terrestre vista pela imagem presente no Google Earth.

O resultado das classificações serão as novas imagens com as correspondentes classes pré-definidas na fotointerpretação, que terá a finalidade de análise de qual algoritmo melhor classifica as regiões e assim realizar a comparação com as classes definidas pelo banco de dados NoSQL.

3.4 SEGMENTAÇÃO POR MEIO DA GEOBIA

Para obter melhores resultados na classificação orientada a objetos, é necessário realizar o processo de segmentação da imagem para identificar regiões semelhantes. Portanto, definiu a segmentação realizada mediante a GEOBIA. Esse método extrai características das imagens digitais utilizando as propriedades básicas dos valores de níveis digitais juntamente das propriedades espaciais como forma, textura, compacidade e suavidade das regiões formadas, separando as que possuem propriedades semelhantes e espacialmente próximas.

Utilizando o software Envi 4.7, na funcionalidade *Feature Extration*, para a segmentação mediante GEOBIA está implementado o algoritmo FLSA. De acordo com Robinson, Redding e Crisp (2002) esse algoritmo consiste em duas etapas, divisão e fusão, onde é necessário regular cuidadosamente os índices *Scale Level* (SL) para escala de divisão e *Merge Level* (ML) para fusão. A primeira etapa é definir para o índice SL qual algoritmo de detecção de bordas *Intensity* ou *Edge* será usado. O valor apropriado para a escala de segmentação deve ser regulado de acordo com o interesse da pesquisa. Esse valor de escala afeta no tamanho e na quantidade das regiões formadas, ou seja, se manter valores baixos de SL o resultado da segmentação terá regiões pequenas e em grande quantidade, que pode causar a supersegmentação.

A segunda etapa é definir o valor para o índice ML que realiza a fusão das regiões similares, eliminando as regiões menores ou ruídos (GONZALEZ; WOODS, 2010). Se aumentar o valor do ML resulta-se na fusão de regiões pequenas formadas durante a divisão, formando regiões maiores, minimizando as possibilidades de supersegmentação.

A GEOBIA resulta no arquivo *shapfile* com os vetores segmentados da imagem e o banco de dados relacional com topologia e atributos (espacial, espectral e textura). O arquivo com os vetores de regiões são visualizados no software Arcgis 10.3, da empresa ESRI, para posteriormente ser usado como parâmetro de comparação entre as imagens classificadas durante o processo de fotointerpretação e a imagem classificada de acordo com as consultas dos atributos no banco de dados NoSQL orientado a grafos.

Os dados presentes no banco de dados resultante da GEOBIA necessitam de preparação, ou seja, deixar todos os valores na mesma escala, realizando a

normalização dos dados. Para isso, utilizou a normalização por meio do mínimo e máximo dos valores, como segue a equação (2):

$$(Z_i^k)_N = \frac{Z_i^k - Z_{min}^k}{Z_{max}^k - Z_{min}^k} \quad (2)$$

Onde k são os atributos, i representa as alternativas e tem-se o valor Z_i^k . O cálculo utiliza o valor mínimo e máximo dos atributos para deixar os valores na mesma escala, tornando possível a aplicação do método de análise multivariada de agrupamento dos dados para a seleção de atributos.

3.5 APLICAÇÃO DA ANÁLISE MULTIVARIADA

A análise multivariada de agrupamento dos dados objetiva a classificação de agrupamento de atributos, ou seja, a criação de grupos de atributos semelhantes baseados em características em comum (SANTOS *et al.*, 2015). Para gerar esses agrupamentos a estatística disponibiliza de vários algoritmos, onde se deve escolher o mais conveniente para a análise dos dados em questão.

Para a análise multivariada será aplicado o método de agrupamento hierárquico, implementado no software Statistica 10, desenvolvido pela empresa StatSoft. Esse método tem como resultado o dendrograma que mostra o agrupamento final dos atributos e permite a interpretação dos atributos que se agruparam, com as respectivas distâncias de cada grupo (PRICHOA, 2012).

Com os dados normalizados aplicados no software, inicia o processo da análise agrupando atributos semelhantes, gerando grupos únicos de atributos. O procedimento se repete com cada atributo que se parece similar, formando outros grupos. Com a redução da semelhança entre atributos, os grupos formados se unem gerando único grupo com todos os atributos, identificados pela distância de cada grupo (SILVA; SANO, 2016).

Ao término da análise é gerado o dendrograma delimita-se a linha de corte na distância desejada para selecionar os grupos que ficarão abaixo dessa linha. Os atributos de cada grupo serão usados para futuras consultas no banco de dados NoSQL explicadas no próximo item.

3.6 BANCO DE DADOS NOSQL

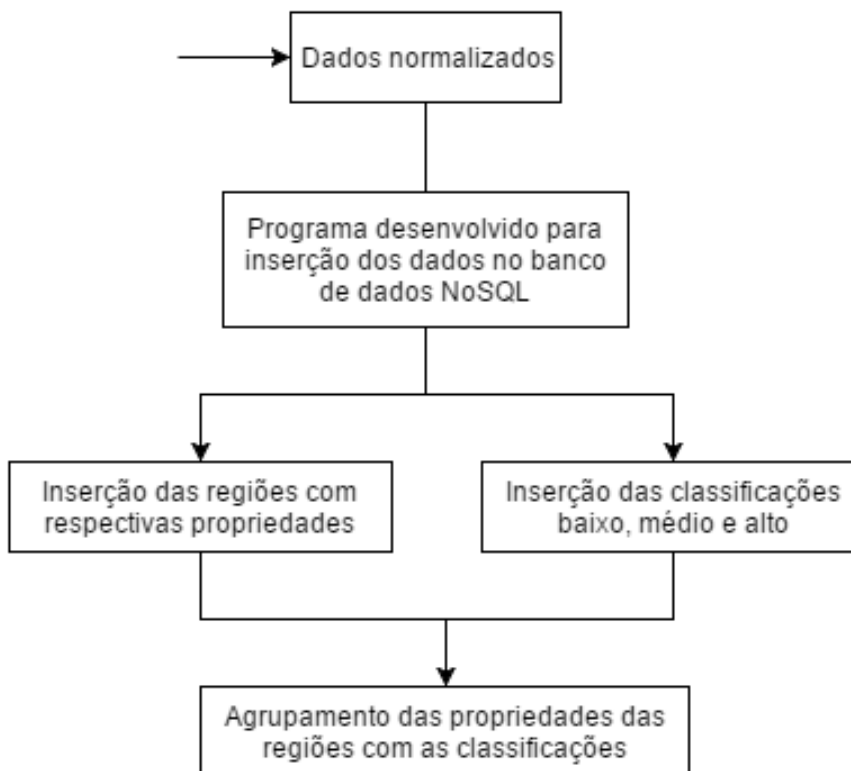
Como parte resultante da GEOBIA, os dados das regiões e seus atributos, estão armazenados em tabelas, que podem ser manipuladas pelos bancos de dados relacionais dentro do SIG. Porém, nesta pesquisa, para trabalhar com os dados optou-se por usar banco de dados NoSQL, pois este possui forma diferenciada de armazenamento e manipulação dos dados, adaptando o banco de dados ao tipo de aplicação desejada. A estrutura de banco de dados NoSQL escolhida é orientada a grafos, que permite visualizar relacionamentos entre os dados e também formar agrupamentos entre eles.

O banco de dados NoSQL utilizado é o Neo4j 3.0, desenvolvido pela *Neo Technology, Inc.* (NEO TECHNOLOGY, 2014) e possui alta capacidade de armazenamento e manipulação de dados, devido ao fato da estrutura ser baseada em grafos. Segundo a *Neo Technology* (2014), o banco de dados Neo4j suporta grandes bases de dados contendo muitos nós, relacionamentos e propriedades.

Semelhante ao banco de dados relacional é possível realizar consultas entre os nós com as *cypher queries*, que tem a mesma função das *queries* (consultas) em bancos relacionais. Para a inserção dos dados normalizados no banco de dados NoSQL orientado a grafos é necessário executar as *cypher queries*, de acordo com a sintaxe do banco de dados, para criar os nós e relacionamentos dentro do banco de dados Neo4j.

O diagrama de blocos da Figura 5 mostra a metodologia aplicada para inserir os dados no banco de dados NoSQL Neo4j. O processo inicia com os dados resultantes da tabela de dados da GEOBIA normalizados, ou seja, preparados para a inserção.

Figura 5 - Diagrama de blocos: inserção dos dados no banco de dados NoSQL.



Fonte: A autora.

Considerando a quantidade de dados a serem inseridos e o tempo para criar cada *cypher query*, implementou-se o programa para a execução desse processo, como forma de automatizar e facilitar a inserção dos dados. O programa foi desenvolvido na linguagem cSharp utilizando o ambiente de desenvolvimento *Visual Studio 2015 Community*, da empresa Microsoft.

A entrada dos dados no programa deu-se a partir da normalização dos dados resultantes da GEOBIA, realizada anteriormente. Ao executar o programa, todos os dados são inseridos no banco de dados Neo4j formando a estrutura de grafo. Essa estrutura permite que tenha nós de diferentes tipos no mesmo grafo, assim, na pesquisa serão utilizados nós de 2 tipos: nós do tipo região e nós do tipo classificação.

Inicialmente, o programa insere 3 nós do tipo classificação, denominados baixo, médio e alto, com propriedades 1, 2 e 3 respectivamente. Definiu-se esse tipo de nó a fim de, posteriormente, criar os relacionamentos entre nós.

Após, ocorre a inserção das 345 regiões resultantes da segmentação, para isso, definiu-se outro tipo de nó, os nós do tipo regiões. Cada região será um nó do tipo região e seus 46 atributos extraídos mediante GEOBIA, entre eles, espaciais, espectrais e de textura serão as propriedades desse nó.

Como todo grafo possui relacionamentos, também conhecidos como arestas, o programa desenvolvido gera também os relacionamentos entre nós. Assim, cada atributo das regiões é relacionado a determinado nó do tipo classificação, ou seja, para criar os relacionamentos, cada atributo (espaciais, espectrais e de textura) do nó do tipo região é ligada ao nó do tipo classificação.

Esse relacionamento se dá pela média simples dos valores dos atributos de cada região, assim, os atributos são classificados em baixo, médio e alto, e geram os relacionamentos, formando a estrutura do grafo. Essa classificação de atributos é realizada pela média dos valores de cada atributo, como mostra a equação 3.

$$y = \max(\text{ValorAtributo}) / 3$$

$$\text{Baixo: } \exists x \mid x < y$$

$$\text{Médio: } \exists x \mid y \leq x < y * 2 \quad (3)$$

$$\text{Alto: } \exists x \mid x \geq y * 2$$

Onde y é dito o valor resultante usado para classificar em baixo, médio ou alto cada atributo de determinada região.

Ao término da execução do programa, os dados estão inseridos no banco de dados NoSQL, formando o grafo, possibilitando realizar as consultas (buscas), visualizar agrupamentos de nós e analisar as regiões e classificações dos atributos em baixo, médio e alto. Para obter o valor de quantos relacionamentos (arestas) são criadas no grafo é calculada a seguinte equação 4:

$$\text{total relacionamentos} = n^{\circ} \text{vértices} * (n^{\circ} \text{atributos} - 1) \quad (4)$$

Com os dados armazenados no banco de dados é possível dar continuidade a pesquisa, realizando as consultas dos grupos de atributos. Para consultas no banco de dados Neo4j é necessário criar as *cypher queries* de consultas, de acordo

com a sintaxe do banco. Os dados utilizados para as consultas são os dados resultantes da análise multivariada que exhibe o agrupamento estatístico dos atributos extraídos da GEOBIA, formado por meio do dendrograma. A partir do dendrograma gerado observam-se os atributos com certo grau de semelhança que se unem em grupos a certa distância um do outro. Para obter a separação desses grupos é determinada a linha de corte no dendrograma, onde se devem separar e utilizar os grupos que são formados abaixo dessa linha de corte.

Nesta pesquisa, definiu-se linha de corte na distância de 1,75 para a separação dos grupos. Definido o valor para a linha de corte do dendrograma, ocorre a separação dos grupos de atributos, os quais são usados para as consultas no banco de dados NoSQL. Essas consultas visam buscar no grafo formado pelo banco de dados Neo4j as regiões semelhantes, ou seja, que possuem os atributos, pertencentes ao grupo, com as mesmas classificações entre baixo, médio ou alto, a fim de encontrar e selecionar no banco de dados as regiões com propriedades semelhantes.

Para essa finalidade, desenvolveu-se outro programa na linguagem c# para realizar essas consultas no banco de dados NoSQL de maneira automática, visto que são várias regiões e inúmeros relacionamentos entre elas, esse programa agiliza as consultas e torna menos propício a erros de sintaxe durante a construção das *cypher queries*.

A entrada dos dados no programa são os atributos pertencentes ao grupo que são utilizados para as consultas. Ao executar as consultas, o programa encontra as regiões que possuem os atributos relacionados a mesma classificação (baixo, médio ou alto) por meio de combinações dos atributos, por exemplo: todos atributos relacionados a classificação alto, todos atributos relacionados a classificação baixo.

Os resultados são salvos em arquivos (.txt) que serão utilizados dentro do SIG para classificação da imagem da área de estudo na análise dos resultados.

3.7 ANÁLISE DOS RESULTADOS

As consultas realizadas no banco de dados NoSQL resultaram em regiões semelhantes. Para a análise dos resultados, insere o arquivo com os segmentos da imagem resultantes da GEOBIA e a partir das regiões trazidas pelas consultas do banco de dados realizou a classificação da imagem.

A análise dos resultados se dá por meio do *shapefile* com os segmentos da imagem resultante da segmentação, aplicados no software Arcgis 10.3. O software possui a funcionalidade de executar consultas na tabela de dados onde é possível selecionar as regiões desejadas na imagem.

Dessa forma, selecionam-se as regiões resultantes das consultas pelo programa desenvolvido. A cada seleção de regiões feita no software é classificada como classes de uso do solo. Ao término terá a nova imagem classificada de acordo com o resultado oriundo das consultas do banco de dados NoSQL.

Com as imagens resultantes da classificação pelos algoritmos de MinDist, MaxVer, KNN e pelo agrupamento de atributos no banco de dados NoSQL orientado a grafos, analisa o delineamento das regiões classificadas em ambas imagens e compara com as classes de interesse pré-definidas no processo de fotointerpretação baseada no Google Earth.

3.8 VERDADE TERRESTRE

As classificações da área de estudo são validadas por meio da Verdade Terrestre (VT), que é o conhecimento da área de estudo por meio de trabalhos realizados em campo, da interpretação visual ou outros levantamentos feitos na mesma área.

Andrade (2013) utilizou a matriz de confusão para avaliar a classificação realizada em seu trabalho, optou-se usar esse método de validação. A matriz de confusão é a representação da qualidade obtida da classificação de imagem medindo a correlação dos dados de referência (VT) com os dados classificados.

A partir da matriz de confusão, para medir a acurácia da classificação calcula-se a exatidão global, índice kappa, erros de inclusão e erros de omissão.

O cálculo do índice kappa é utilizado para análise de resultados e amplamente aceito e adotado pelo SIG e SR. Esse índice é que retrata o grau de

concordância dos dados, apresentando aspecto de confiabilidade e precisão dos dados classificados (PRINA; TRENTIN, 2015). O valor obtido pelo índice kappa, varia no intervalo de 0 a 1, sendo que quanto mais próximo a 1, melhor a qualidade dos dados classificados (FONSECA, 2000). A Tabela 3 exibe os valores de referências do índice kappa.

Tabela 3 - Valores de referência do índice kappa.

Índice kappa	Desempenho
<0	Péssimo
$0 < k \leq 0,2$	Ruim
$0,2 < k \leq 0,4$	Razoável
$0,4 < k \leq 0,6$	Bom
$0,6 < k \leq 0,8$	Muito Bom
$0,8 < k \leq 1,0$	Excelente

Fonte: (FONSECA, 2000).

Como imagem da VT utilizou-se a imagem da área de estudo apresentada pelo Google Earth. As mesmas amostras (*pixels*) coletadas na imagem Rapideye foram identificadas como pertencentes a mesma classe proposta. Algumas áreas são difíceis de ser classificadas devido as diferentes respostas espectrais, porém é importante a definição das mesmas classes em ambas as imagens para comparar a imagem que possui a VT com a imagem obtida pela classificação.

Para gerar a matriz de confusão utilizou-se o software Envi 4.7, comparando a imagem da VT e a sua correspondência na imagem resultante da classificação.

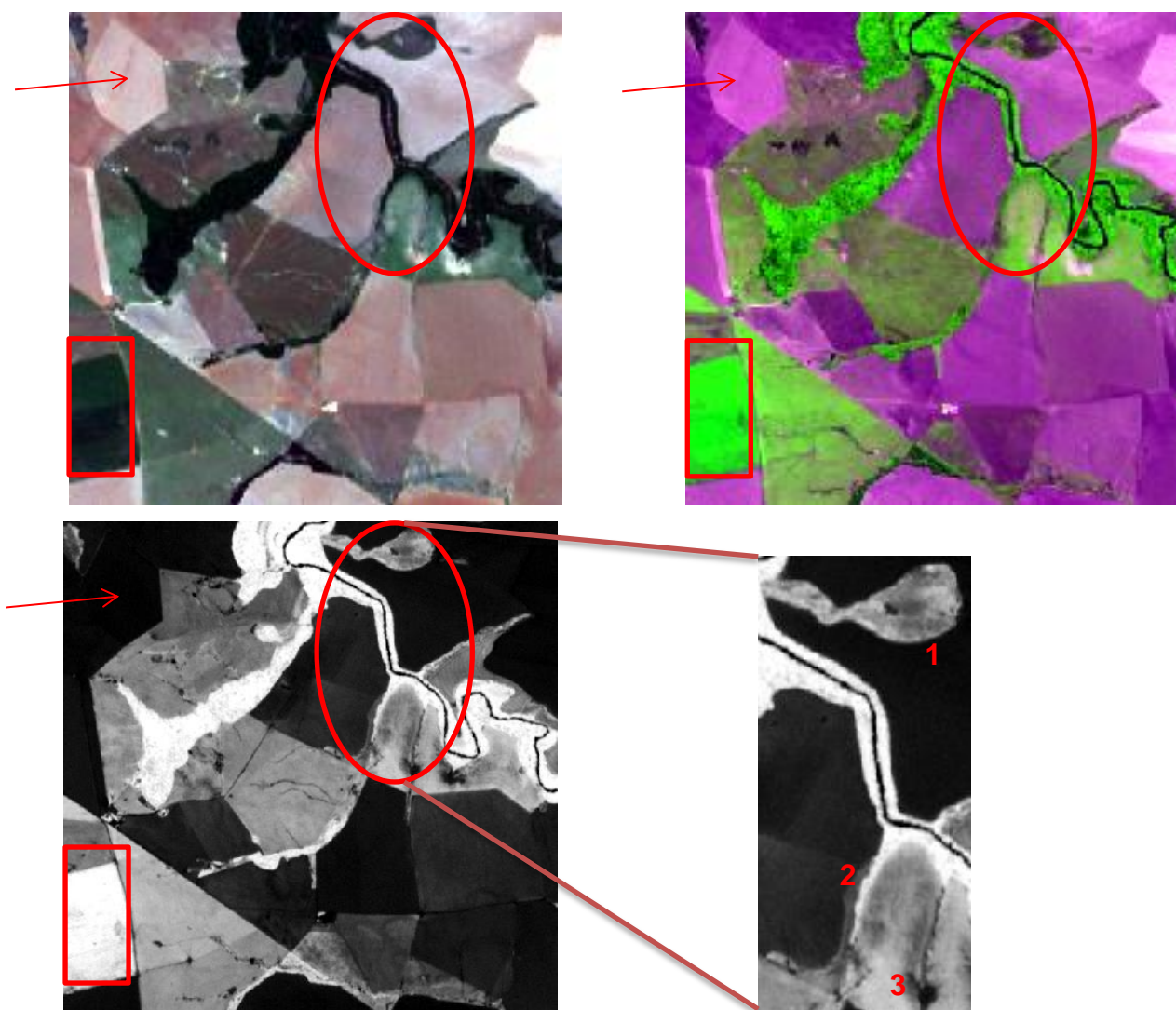
Esta seção abordou a metodologia definida para a pesquisa, explicando as etapas e os métodos utilizados. Na seção 4 estão apresentados os resultados obtidos em cada etapa proposta por meio dos métodos selecionados.

4 RESULTADOS E DISCUSSÃO

4.1 FOTOINTERPRETAÇÃO

Ao calcular e aplicar o NDVI na imagem Rapideye da área de estudo, esta passou a ser composta por 6 bandas espectrais. A Figura 6 exibe a imagem da área de estudo representada pela banda artificial correspondente ao NDVI.

Figura 6 - Imagens da área de estudo apresentada pela verdadeira cor, pela composição das bandas 4R5G3B e a banda artificial NDVI.



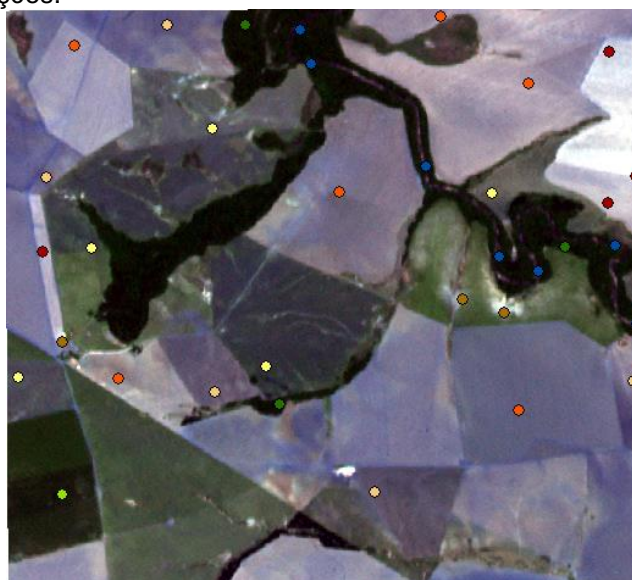
Fonte: A autora.

Esse método foi utilizado para destacar as feições da imagem. Pode-se analisar que houve diferenciação das classes de uso e ocupação do solo. Conforme o destaque em forma de elipse, ampliada para melhor visualização, na imagem representada pela banda NDVI, observa-se que a partir da aplicação do índice de vegetação da área, correspondente à mata ciliar (1 e 2) se destacou, o que não era identificada pela verdadeira cor e composição de bandas.

A área 3 exibe o destaque do solo exposto encontrado na imagem. As áreas marcadas pela seta mostram a ausência de vegetação. Pode-se afirmar que a área é de cultivo, porém não está em fase de plantio. A presença de vegetação na área demarcada pela forma retangular mostra que a área é de cultivo e está em época de plantio.

Com o NDVI aplicado à imagem, realizou-se o processo de fotointerpretação baseando-se na verdade terrestre da imagem do Google Earth, resultando na definição das classes de uso e ocupação do solo. O resultado da fotointerpretação definiu 8 classes de interesse, sendo elas: rio, mata ciliar, solo exposto, cultivo 1, cultivo 2, cultivo 3, cultivo 4 e cultivo 5. A Figura 7 apresenta o resultado da coleta das amostras para as classes de interesse utilizadas na classificação.

Figura 7 - Resultado das amostras coletadas na fotointerpretação para as 8 classes de interesse utilizadas nas classificações.



Cultivo 1
 Cultivo 3
 Cultivo 5
 Rio

Cultivo 2
 Cultivo 4
 Mata ciliar
 Solo exposto

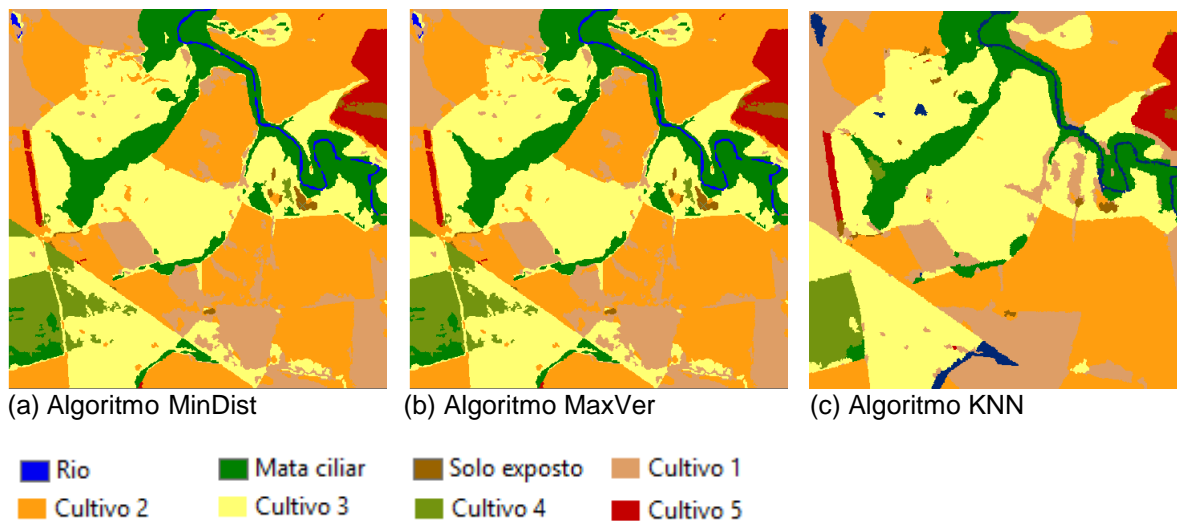
Fonte: A autora.

Após a definição das classes, classificou-se a imagem da área de estudo pelos métodos *pixel a pixel* com os algoritmos MinDist e MaxVer, utilizando as amostras coletadas.

O primeiro método de classificação supervisionada utilizou o algoritmo MinDist, conforme observa-se o resultado na Figura 8a. A segunda classificação usou o algoritmo MaxVer, tendo seu resultado apresentado na Figura 8b.

Por fim, o terceiro método de classificação supervisionada utilizado é orientado a objetos, pelo algoritmo KNN. Para essa classificação, segmentou-se a imagem em regiões com características espaciais, espectrais e texturas similares e está exibido na Figura 8c.

Figura 8 - Imagem resultante da classificação supervisionada pelo algoritmo de Distância Mínima, com 8 classes de uso e ocupação do solo.



Fonte: A autora.

De acordo com a Figura 8a, o método de classificação pelo algoritmo MinDist, classificou a imagem com todas as classes de interesse, porém, observa-se que a classe rio teve falhas durante a classificação. A mata ciliar se classificou bem, mas as demais classes de cultivos não se definiram, houve ruídos na separação entre elas.

A Figura 8b, mostra a classificação pelo algoritmo MaxVer, classificou a classe rio com falhas no seu percurso e as classes de cultivos 1, 2 e 3 não mostraram separação bem definida entre elas. A classe mata ciliar se classificou de forma adequada.

Com o método de classificação pelo algoritmo KNN, apresentado na Figura 8c, percebe-se claramente a delimitação das classes. Nesta, a classe rio se classificou sem falhas e as classes de cultivos ficaram devidamente separadas. Pode-se afirmar, que as 8 classes de uso e ocupação foram bem definidas e não apresentaram ruídos na classificação.

4.2 GEOBIA

O processo de segmentação da imagem Rapideye realizado pelo método da GEOBIA, separou as regiões com características espectrais, espaciais e texturas semelhantes. Conforme citado anteriormente, no processo de GEOBIA é importante a definição dos valores da escala e fusão, pois estes parâmetros interferem no resultado da segmentação e agrupamento de objetos de interesse para a classificação de uso e cobertura da terra. Devido a isso, foram realizados testes variando os valores de escala e fusão na etapa de segmentação, com a intenção de encontrar o teste que melhor representasse a delimitação dos objetos na imagem.

De acordo com os testes de segmentação aplicados pelos autores Ribeiro, Almeida e Ribeiro (2016) com diferentes valores de escala e fusão evidencia a importância da definição desses índices na obtenção de bom resultado na segmentação. Ao usar valores inadequados o resultado gera vetores prejudiciais para o reconhecimento dos objetos presentes na imagem.

A Tabela 4 descreve os testes utilizados para gerar os resultados da segmentação mediante a GEOBIA, utilizando o algoritmo de escala *Edge* e algoritmo de fusão FLSA.

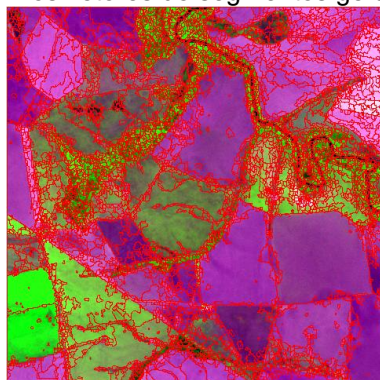
Tabela 4 - Testes de valores dos índices de escala e fusão para o algoritmo de segmentação FLSA.

Índice	Teste 1	Teste 2	Teste 3	Teste 4
Escala	40	50	70	56
Fusão	60	80	85	96

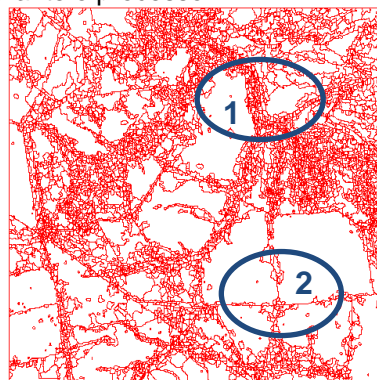
Fonte: A autora.

Realizados quatro testes com diferentes valores de escala e fusão para a segmentação da imagem, gerando resultados diferentes durante a delimitação dos objetos presentes na imagem. A Figura 9 mostra os resultados dos testes realizados, ou seja, a formação das regiões na imagem Rapideye da área de estudo.

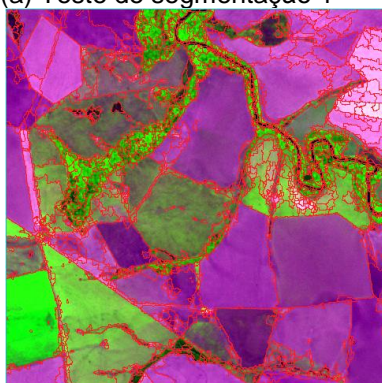
Figura 9 - Resultados dos testes de segmentação com o algoritmo de escala Edge e algoritmo de fusão FLSA, com os vetores de segmentos gerados durante o processo.



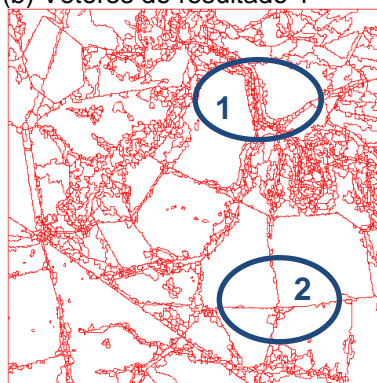
(a) Teste de segmentação 1



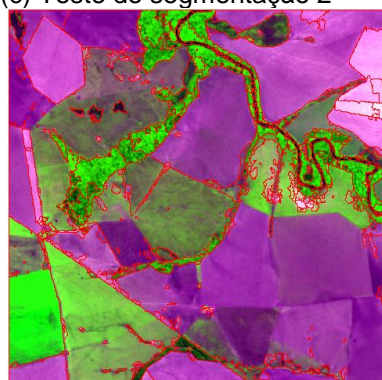
(b) Vetores do resultado 1



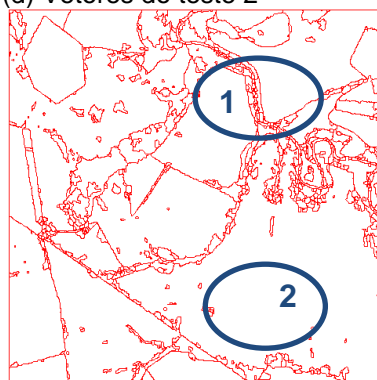
(c) Teste de segmentação 2



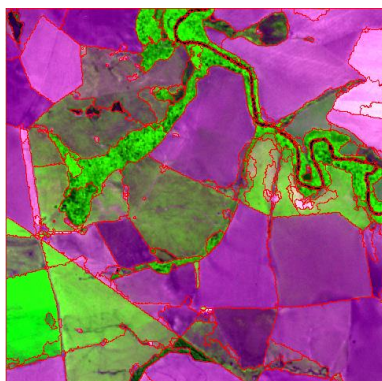
(d) Vetores do teste 2



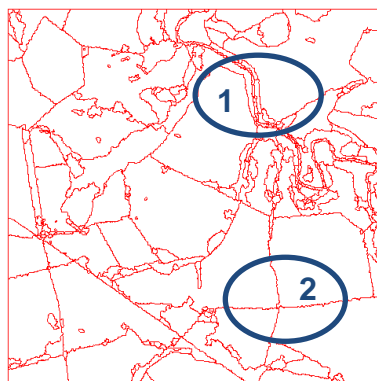
(e) Teste de segmentação 3



(f) Vetores do teste 3



(g) Teste de segmentação 4



(h) Vetores do teste 4

Fonte: A autora.

De acordo com os resultados dos testes, é possível observar que quando os valores de escala e fusão são pequenos, conforme o teste 1 gera supersegmentação, ou seja, são formados muitos segmentos na imagem o que dificulta a separação dos mesmos. Quando usados valores altos, assim como no teste 3 são perdidos os limites de objetos.

O teste 1 com a segmentação com escala de 40 e fusão de 60, foram gerados os segmentos mostrados na Figura 9a, onde percebe-se a quantidade de segmentos criados. Em 9b é exibido somente os vetores gerados e onde se identifica a quantidade de informações desnecessárias. É possível analisar a supersegmentação da imagem, ou seja, existem muitos ruídos que prejudicam na identificação das classes de cobertura e uso da terra.

O segundo teste de segmentação foi com escala de 50 e fusão de 80. Aumentando a escala em 10 e a fusão em 20. Nas Figura 9c e 9d estão os resultados desta segmentação. Nota-se a diminuição dos pequenos segmentos comparados com a segmentação do teste 1, entretanto, ainda há ruídos na classe mata ciliar e rio.

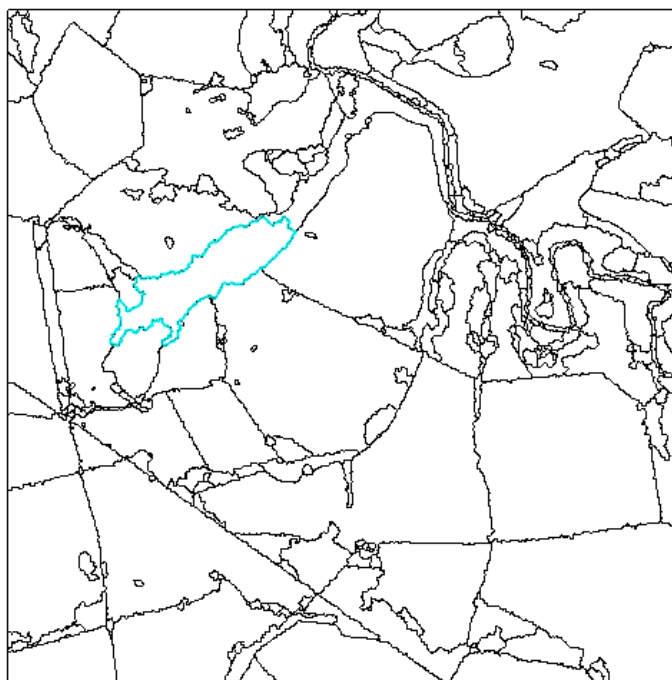
As Figura 9e e 9f ilustram a segmentação do teste 3 com os índices escala de 70 e fusão de 85. Aumentando a fusão em 15 com a intenção de agrupar ruídos, porém nota-se a perda de informação das regiões na área de estudo, ocasionando em segmentos maiores. Dessa maneira, não é permitido o contorno das fronteiras entre os cultivos.

Para o teste 4 aumentou-se em 6 o valor da escala do teste 2, que ficou em 56 e acrescentou-se em 11 o valor da fusão do teste 3, tornando em 96, com intuito de reduzir ruídos dentro de segmentos maiores. Conforme se observa nas Figura 9g e 9h com o valor de fusão maior dos demais testes, os pequenos segmentos se fundiram com segmentos maiores sem perder informações da área de estudo.

Observando o destaque representado pela forma de elipse 1, o rio para o Teste 1 apresenta muitos pequenos segmentos, ruídos, dificultando a delimitação da classe. Já para o Teste 4, os segmentos diminuíram pois houve o aumento do valor de fusão, permitindo a diferenciação da classe. Para o destaque representado pela forma de elipse 2, é possível ver pelo resultado do Teste 3 que o limite do objeto foi perdido devido aos valores de escala e fusão utilizados, já no Teste 4, ocorreu a separação dos objetos sem ruídos, diferentemente dos testes 1 e 2.

O resultado da segmentação do teste 4 foi escolhido, pois melhor representou as delimitações entre as classes de uso e ocupação do solo. A segmentação gerou os atributos espaciais, espectrais e de textura, resultando na formação do banco de dados relacional, conforme visto na Figura 10.

Figura 10 - Apresentação das regiões e do banco de dados espacial gerados a partir da segmentação mediante GEOBIA.



(a) Região selecionada

REGION_ID	AREA	LENGTH	COMPACT	CONVEXITY	SOLIDITY	ROUNDNESS	FORMFACTOR	ELONGATION	RECT_FIT	MAINDIR
318.00000	262.50000	68.839145	0.26557300	1.0000000	1.0000000	0.55920200	0.69609500	1.3731130	0.60306600	77.932570
70.000000	29525.000	1582.3133	0.12253400	1.4444460	0.47910800	0.16011900	0.14818900	2.8287080	0.35573000	144.20903
246.00000	2250.0000	316.68566	0.16901200	1.2253090	0.64981900	0.27197800	0.28192600	2.3514360	0.50229300	162.50094
232.00000	737.50000	119.16226	0.25715600	1.0202110	0.90769200	0.42311900	0.65267100	1.7638660	0.58616200	29.096260
324.00000	1337.5000	226.82218	0.18193500	1.1776120	0.74825200	0.25682000	0.32668800	2.9609240	0.59723700	165.28492
198.00000	1662.5000	192.30608	0.23924500	1.0774940	0.78698200	0.57421800	0.56491800	1.3626020	0.61452000	40.839450
67.000000	13100.000	841.96954	0.15338900	1.2012350	0.69312200	0.15544700	0.23221400	4.5033870	0.54980800	148.69752
263.00000	437.50000	98.693844	0.23914100	1.0777230	0.85365900	0.46528100	0.56442700	1.4727470	0.53818700	33.141889
66.000000	131487.50	2216.7291	0.18458000	1.0801710	0.86016800	0.20218200	0.33625500	4.0102760	0.63680400	96.602930
323.00000	550.00000	107.34707	0.24651700	1.0577190	0.91666700	0.47292400	0.59978000	1.9163820	0.71180900	169.79087
320.00000	200.00000	60.000000	0.26596200	1.0513210	0.94117600	0.56588400	0.69813200	1.2000000	0.53333300	135.00000
69.000000	2462.5000	325.73779	0.17190000	1.3856430	0.70609300	0.41438600	0.29164200	1.8539880	0.50575700	176.98816
68.000000	158637.50	3207.0795	0.14013600	1.4871800	0.69021600	0.21668900	0.19381900	2.8816680	0.49042300	31.777225
313.00000	475.00000	97.486826	0.25226400	1.0470930	0.90476200	0.49455000	0.62807500	1.7320900	0.67277600	177.63152
71.000000	775.00000	127.53020	0.24631600	1.0548940	0.88571400	0.43385800	0.59880500	1.7385210	0.59240400	51.606262
73.000000	59450.000	1250.8312	0.21995400	1.2154740	0.90349500	0.48576700	0.47749000	1.5374080	0.58655300	123.25142
301.00000	150.00000	70.000000	0.19742500	1.1604300	0.66666700	0.30916000	0.38468500	1.7290740	0.41984300	136.95025
72.000000	725.00000	163.43551	0.18589900	1.1819730	0.60416700	0.32544900	0.34107900	1.8940500	0.48413300	136.28615
74.000000	5387.5000	338.56491	0.24462800	1.0717290	0.91507400	0.43207200	0.59062700	1.9807850	0.67217600	91.730803
76.000000	77800.000	3319.2361	0.11154200	2.0183060	0.40722300	0.35159600	0.088739000	1.0035990	0.27713600	47.688531
230.00000	150.00000	70.000000	0.19742500	1.1359440	0.66666700	0.47746500	0.38468500	1.3333330	0.50000000	89.999999
77.000000	712.50000	109.53308	0.27498100	1.0296350	0.93442600	0.75655500	0.74628400	1.1395490	0.67631100	98.125808
215.00000	175.00000	65.388436	0.22828300	1.0103070	0.87500000	0.27895700	0.51433500	2.0000000	0.43818500	132.73421
297.00000	150.00000	50.000000	0.27639500	1.0000000	1.0000000	0.84882600	0.75398200	1.5000000	1.0000000	0.00000000
268.00000	462.50000	129.76622	0.18700300	1.2033190	0.68518500	0.29596400	0.34514200	1.6767310	0.38975600	149.31514
325.00000	150.00000	50.000000	0.27639500	1.0000000	1.0000000	0.84882600	0.75398200	1.5000000	1.0000000	0.00000000
304.00000	762.50000	189.68865	0.16426100	1.1357520	0.53043500	0.19421200	0.26629700	2.5771220	0.39309900	135.85724
79.000000	17337.500	1024.1333	0.14507500	1.3616320	0.64272500	0.20341000	0.20772200	2.7577630	0.44057500	135.99258
75.000000	1425.0000	182.13627	0.23386500	1.0681270	0.80851100	0.38766100	0.53979900	1.9924030	0.60662300	156.40890
78.000000	6112.5000	502.26287	0.17564400	1.2403790	0.64682500	0.31749700	0.30448500	1.9027300	0.47446800	164.59005
256.00000	837.50000	138.31216	0.23609500	1.0264900	0.93056000	0.32604200	0.55014100	2.7406560	0.70180800	160.28819

(b) Apresentação dos atributos da região selecionada no banco de dados

Fonte: A autora.

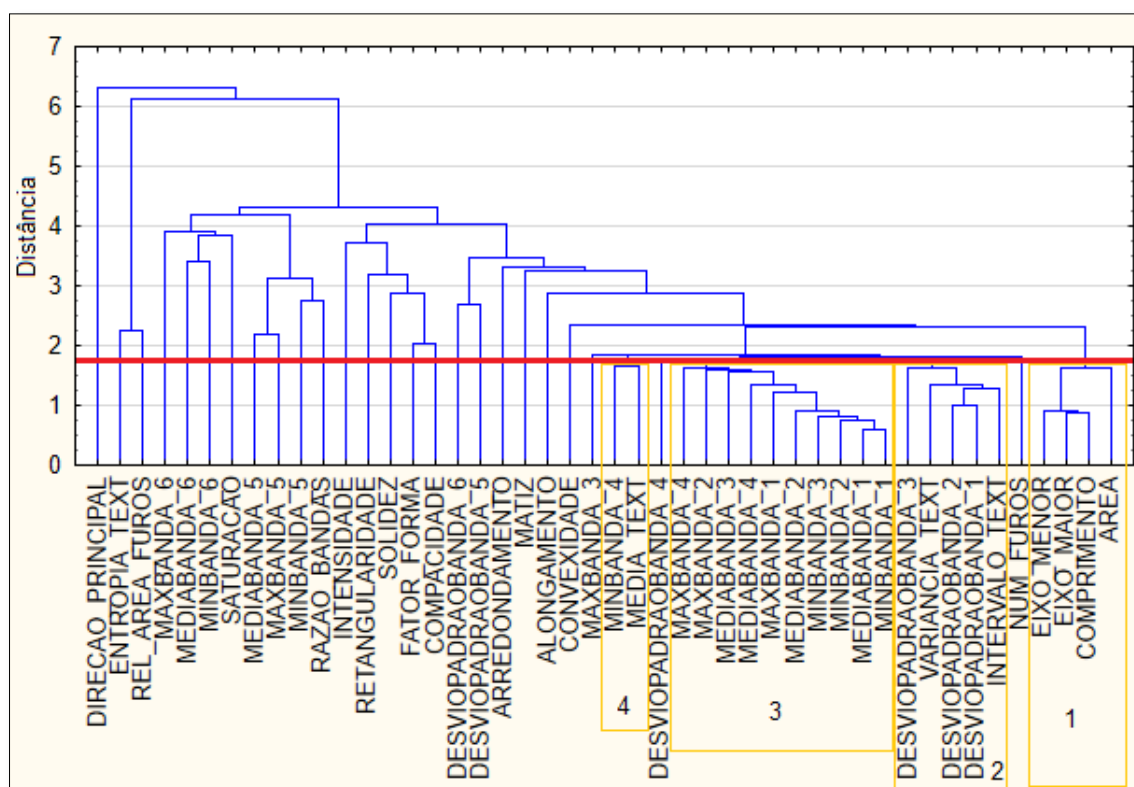
Observa-se na Figura 10a as regiões resultantes da segmentação do teste 4, onde apenas uma está selecionada para exibição de parte dos atributos gerados a partir de cada região da imagem no banco de dados relacional importado para o SIG, conforme a Figura 10b.

Com a imagem segmentada pelo método da GEOBIA e mediante os atributos do banco de dados, aplicou-se o pré-processamento a eles, ou seja, a normalização dos dados, para que estejam em mesma escala, visto que, possuem diferentes unidades de medida.

4.3 AGRUPAMENTO DE ATRIBUTOS

Com os dados resultantes da GEOBIA normalizados, aplicou-se a AVM para agrupar os atributos que se encontram na mesma distância por meio do dendrograma. A Figura 11 mostra o dendrograma gerado pela AVM com a linha de corte e os grupos formados abaixo dessa linha.

Figura 11 - Dendrograma dos 46 atributos provenientes da GEOBIA, as distâncias entre os grupos e com a linha de corte formando 4 grupos.



Fonte: A autora.

O dendrograma agrupou os 46 atributos que para a análise necessitam ser separados em grupos de atributos, utilizados para as consultas no banco de dados NoSQL orientado a grafos. Os grupos de atributos formados, abaixo da linha de corte definida inicialmente, foram utilizados para as consultas. De acordo com o dendrograma resultante, nota-se a formação de 4 grupos abaixo da linha de corte, sendo eles:

- Grupo 1: área, comprimento, eixo menor e eixo maior;
- Grupo 2: intervalo de textura, desvio padrão da banda 1, desvio padrão da banda 2, variância da textura e desvio padrão da banda 3;
- Grupo 3: maxbanda_2, mediabanda_3, mediabanda_4, maxbanda_1, mediabanda_2, minbanda_3, minbanda_2, mediabanda_1 e minbanda_1;
- Grupo 4: minbanda_4 e média de textura.

Com os atributos pertencentes a cada grupo criou-se as *cypher queries* e assim, realizadas as consultas no banco de dados Neo4j. Para isso, a inserção dos dados no banco de dados é fundamental e será explicada na próxima seção.

4.4 BANCO DE DADOS NOSQL NEO4J

O banco de dados gerado pela GEOBIA é relacional, porém para a pesquisa definiu-se o uso do banco de dados NoSQL orientado a grafos. Para isso, foi necessário transformar a tabela de dados para a sintaxe do banco de dados NoSQL.

Utilizando o banco de dados Neo4j, que possui estrutura baseada em grafos inseriu-se os dados normalizados resultante da GEOBIA por meio do programa desenvolvido para inserção. A Figura 12 mostra a tela do programa desenvolvido para inserção dos dados normalizados, resultantes da GEOBIA, no banco de dados Neo4j .

Figura 12 - Tela do programa para inserir os dados no banco de dados Neo4j.

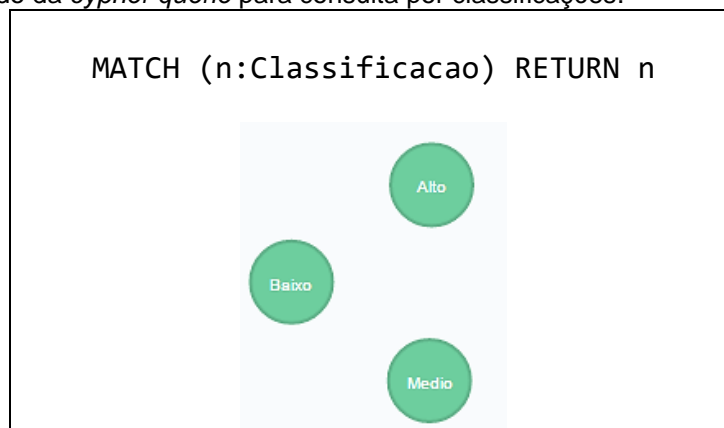
```

Mestrado em Computação Aplicada - Evelaine Berger Ribeiro
Programa para inserção dos dados da GEOBIA no banco de dados NoSQL Neo4j
Pressione uma tecla para iniciar...
-----
- Classificação 2 inserida!
- Classificação 3 inserida!
- Classificação 1 inserida!
- Região 1 inserida!
- Região 3 inserida!
- Região 7 inserida!
- Região 8 inserida!
- Região 4 inserida!
- Região 9 inserida!
- Região 5 inserida!
- Região 11 inserida!
- Região 2 inserida!
- Região 12 inserida!
- Região 6 inserida!
- Região 13 inserida!
- Região 16 inserida!
- Região 15 inserida!
- Região 18 inserida!
- Região 20 inserida!
- Região 21 inserida!
  
```

Fonte: A autora.

Inicialmente, o programa inseriu os nós do tipo classificação, resultando nos primeiros nós do grafo, conforme se observa na Figura 13.

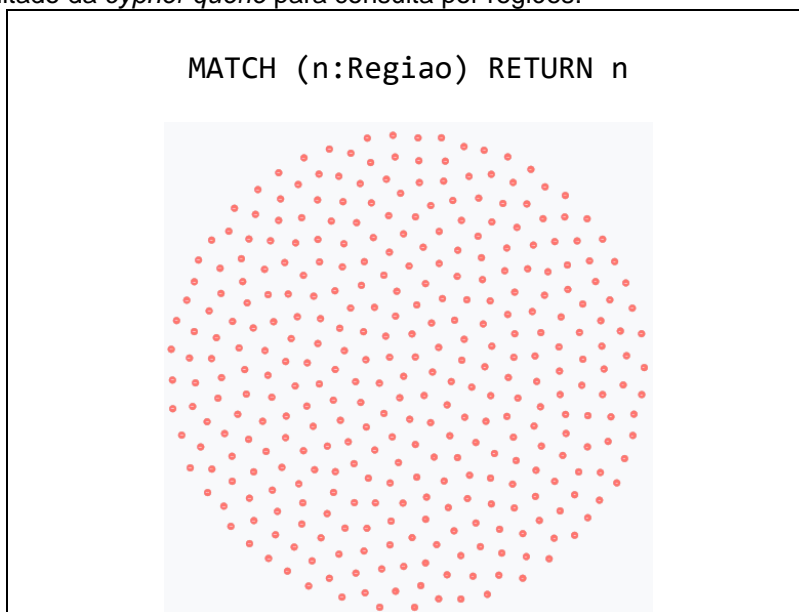
Figura 13 - Resultado da *cypher* *querie* para consulta por classificações.



Fonte: A autora.

Os nós do tipo classificação, baixo, médio e alto são utilizados para criar os relacionamentos entre nós. Após, inseriu-se os nós do tipo região, resultando nos nós apresentado na Figura 14.

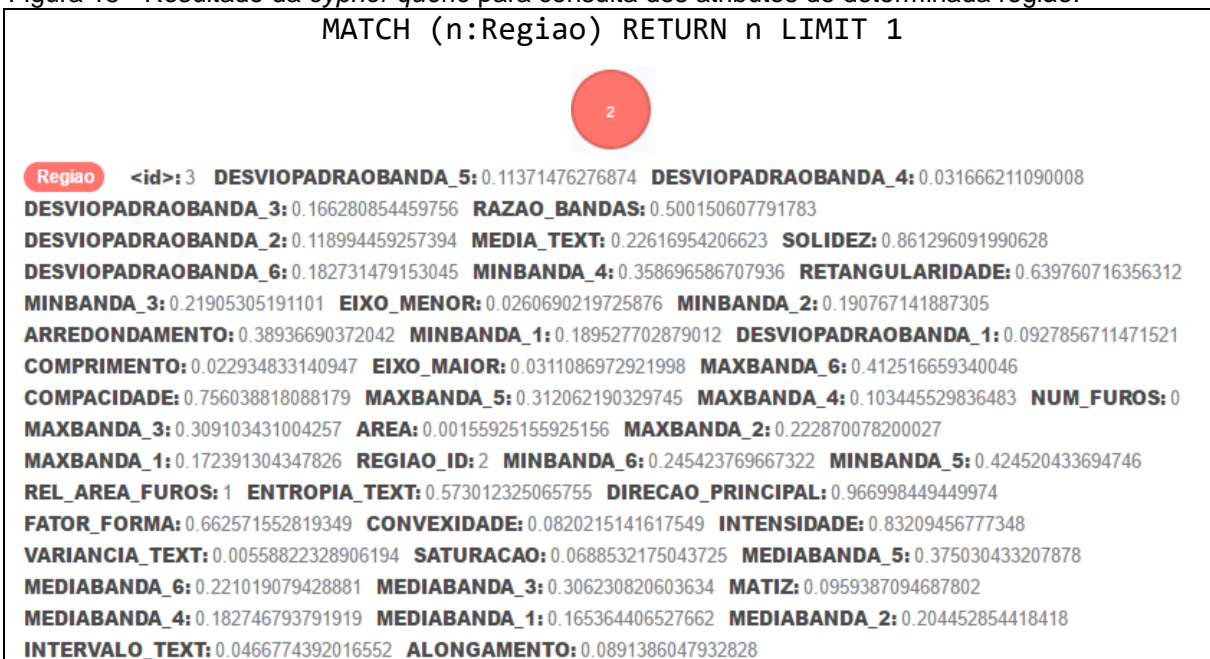
Figura 14 - Resultado da *cypher query* para consulta por regiões.



Fonte: A autora.

A inserção das regiões da segmentação resultou em 345 nós do tipo região inseridos no banco de dados orientado a grafos, cada nó do tipo região com 46 propriedades, que são atributos (espaciais, espectrais e textura) extraídos mediante GEOBIA, como se pode observar na Figura 15.

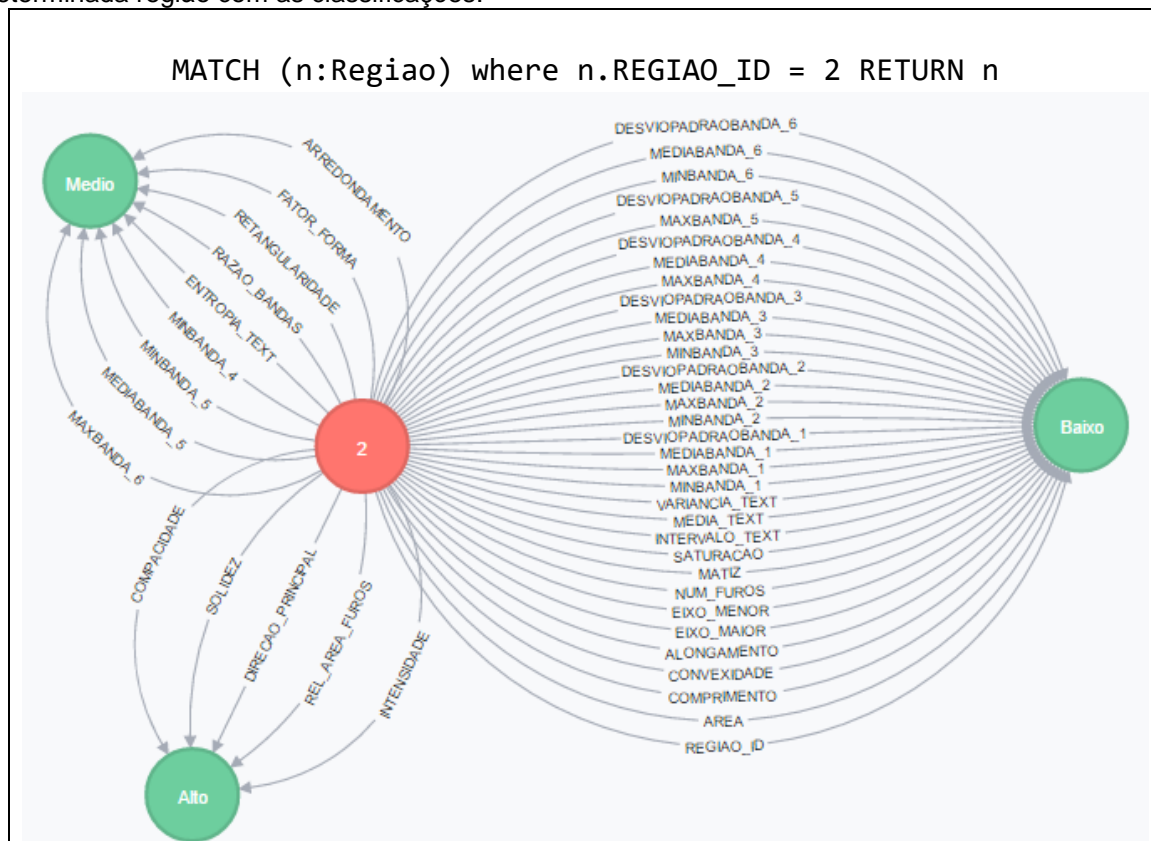
Figura 15 - Resultado da *cypher query* para consulta dos atributos de determinada região.



Fonte: A autora.

Por fim, o programa realizou o relacionamento entre cada atributo dos nós do tipo região com os nós do tipo classificação, como mostra a Figura 16.

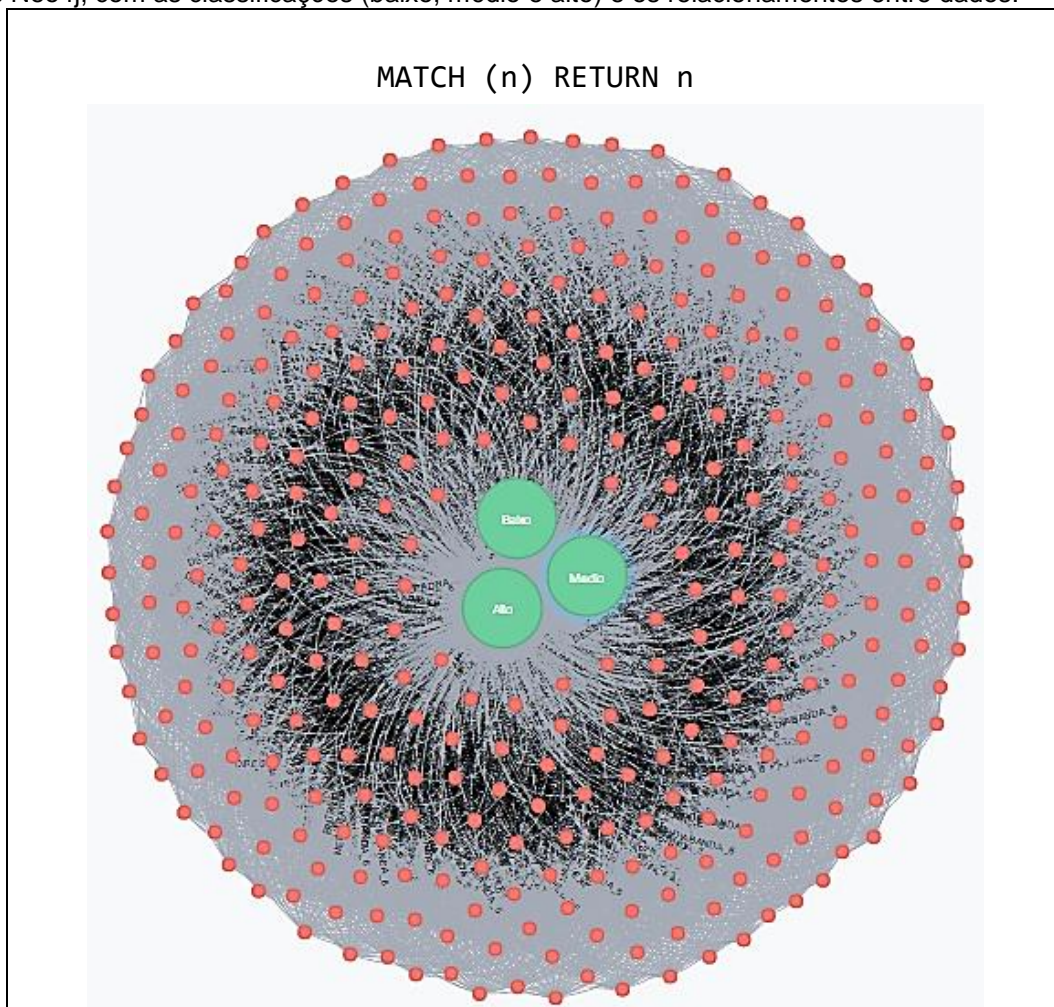
Figura 16 - Resultado da *cypher query* para consulta dos relacionamentos entre atributos de determinada região com as classificações.



Fonte: A autora.

Como pode ser visto, o nó do tipo região 2 possui atributos relacionados aos diferentes nós do tipo classificações. Dessa maneira, formaram-se os relacionamentos do grafo, totalizando em 16.215 relacionamentos. A Figura 17 mostra o resultado da inserção dos dados no banco de dados NoSQL.

Figura 17 - Apresentação de todos os dados resultantes da GEOBIA armazenados no banco de dados Neo4j, com as classificações (baixo, médio e alto) e os relacionamentos entre dados.

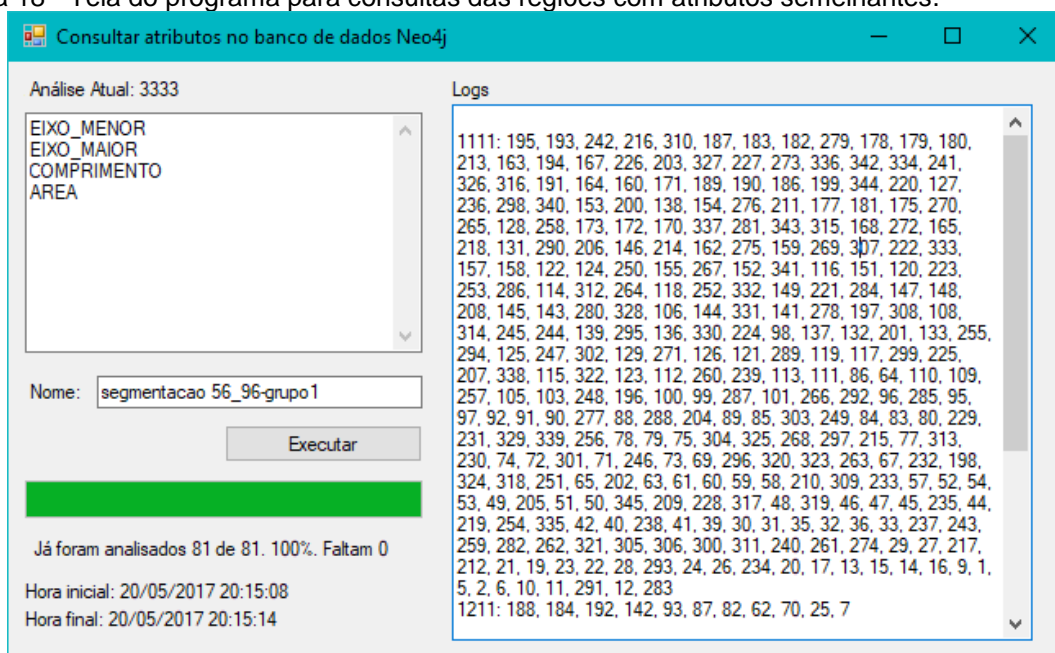


Fonte: A autora.

Com a inserção dos dados extraídos mediante GEOBIA no banco de dados Neo4j resultou na formação completa do grafo, com nós e relacionamentos, possibilitando as consultas dos atributos no banco de dados.

Para as consultas dos grupos de atributos formados pelo dendrograma, desenvolveu-se o programa para executar as *cypher queries* de consultas dos grupos de atributos como mostra a Figura 18.

Figura 18 - Tela do programa para consultas das regiões com atributos semelhantes.



Fonte: A autora.

Como resultado, o programa exibe as regiões com atributos espaciais, espectrais e de textura semelhantes para a realização da classificação da imagem da área de estudo pelo banco dados NoSQL.

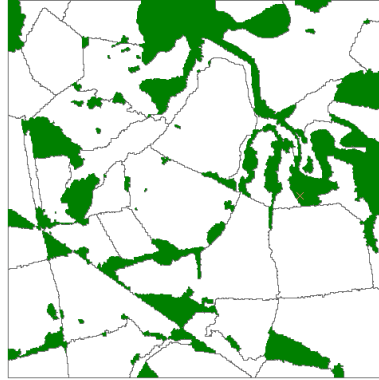
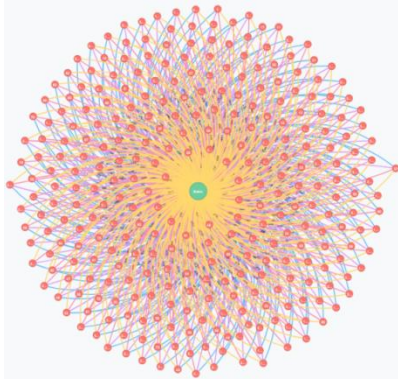
4.5 ANÁLISE DAS CLASSIFICAÇÕES

Os resultados das consultas são as regiões semelhantes encontradas por meio do agrupamento de atributos (espaciais, espectrais e de textura). O Quadro 5 alguns dos resultados das consultas com o banco de dados orientado a grafo, as regiões selecionadas e como resultaram na classificação da imagem.

Quadro 5 - Resultados das consultas realizadas no banco de dados NoSQL por meio do agrupamento dos atributos da GEOBIA.

Grupo 1: Eixo_Menor, Eixo_Maior, Comprimento, Area

Agrupamento: 1111- Todos relacionados na classificação baixo



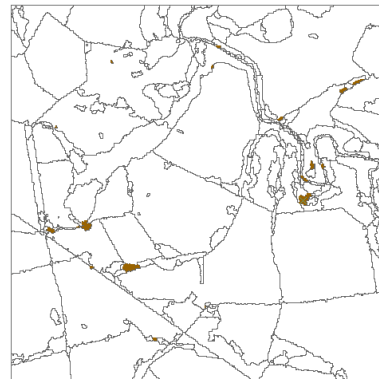
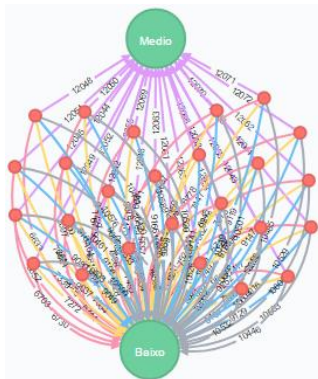
Classe: Mata ciliar

Regiões: 195, 193, 242, 216, 310, 187, 183, 182, 279, 178, 179, 180, 213, 163, 194, 167, 226, 203, 327, 227, 273, 336, 342, 334, 241, 326, 316, 191, 164, 160, 171, 189, 190, 186, 199, 344, 220, 127, 236, 298, 340, 153, 200, 138, 154, 276, 211, 177, 181, 175, 270, 265, 128, 258, 173, 172, 170, 337, 281, 343, 315, 168, 272, 165, 218, 131, 290, 206, 146, 214, 162, 275, 159, 269, 307, 222, 333, 157, 158, 122, 124, 250, 155, 267, 152, 341, 116, 151, 120, 223, 253, 286, 114, 312, 264, 118, 252, 332, 149, 221, 284, 147, 148, 208, 145, 143, 280, 328, 106, 144, 331, 141, 278, 197, 308, 108, 314, 245, 244, 139, 295, 136, 330, 224, 98, 137, 132, 201, 133, 255, 294, 125, 247, 302, 129, 271, 126, 121, 289, 119, 117, 299, 225, 207, 338, 115, 322, 123, 112, 260, 239, 113, 111, 86, 64, 110, 109, 257, 105, 103, 248, 196, 100, 99, 287, 101, 266, 292, 96, 285, 95, 97, 92, 91, 90, 277, 88, 288, 204, 89, 85, 303, 249, 84, 83, 80, 229, 231, 329, 339, 256, 78, 79, 75, 304, 325, 268, 297, 215, 77, 313, 230, 74, 72, 301, 71, 246, 73, 69, 296, 320, 323, 263, 67, 232, 198, 324, 318, 251, 65, 202, 63, 61, 60, 59, 58, 210, 309, 233, 57, 52, 54, 53, 49, 205, 51, 50, 345, 209, 228, 317, 48, 319, 46, 47, 45, 235, 44, 219, 254, 335, 42, 40, 238, 41, 39, 30, 31, 35, 32, 36, 33, 237, 243, 259, 282, 262, 321, 305, 306, 300, 311, 240, 261, 274, 29, 27, 217, 212, 21, 19, 23, 22, 28, 293, 24, 26, 234, 20, 17, 13, 15, 14, 16, 9, 1, 5, 2, 6, 10, 11, 291, 12, 283

(a) Regiões que representam a classe mata ciliar

Grupo 2: Desviopadraobanda_3, Variancia_Text, Desviopadraobanda_2, Desviopadraobanda_1, Intervalo_Text

Agrupamento: 21111



Classe: Solo exposto

Regiões: 314, 279, 326, 213, 163, 275, 332, 252, 148, 236, 128, 245, 302, 262, 335, 44, 57, 230, 91, 322, 106, 257, 330, 232, 317

(b) Regiões que representam a classe solo exposto

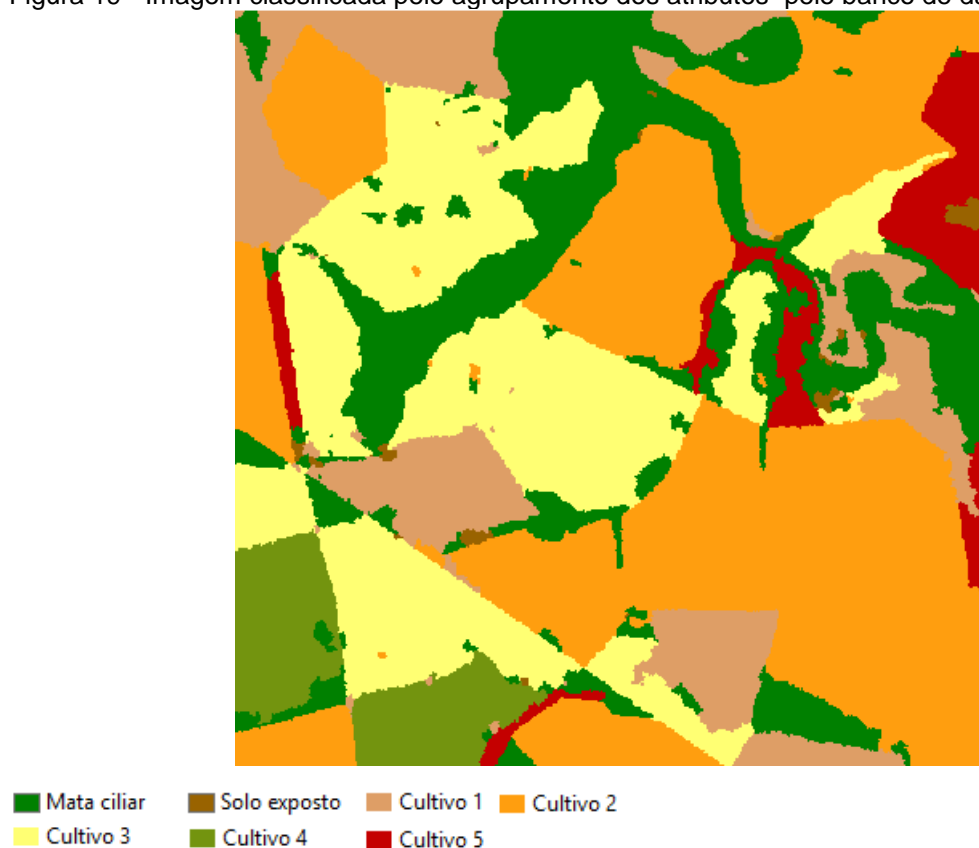
Fonte: A autora.

O Quadro 5a exibe a consulta no banco de dados NoSQL, onde o intuito era selecionar os nós do tipo região que possuem os atributos eixo menor, eixo maior, comprimento e área possuíam relacionados no nó do tipo classificação baixo. Como resultado, o grafo exibiu as regiões que satisfazem essa condição. Para a classificação da imagem, que também pode ser vista no quadro apresentado, é resultado das regiões selecionadas pelo banco de dados e dentro do SIG, utilizadas para a definir a quais classes pertencem. Nesta, nota-se que as regiões trazidas pelo grafo pertencem a classe mata ciliar, a consulta traz praticamente todas as regiões que compõem essa classe.

Para o Quadro 5b segue a mesma ideia, porém utilizando atributos do grupo 2, formado pelo dendrograma. Com os atributos apresentados no grupo e relacionados as classificações médio e baixo, o grafo selecionou pequenas regiões que representam a classe solo exposto. O resultado completo das consultas estão apresentados no APÊNDICE A, juntamente com a ilustração do agrupamento das regiões no grafo, mostrando os relacionamentos dos atributos e a definição das classes por regiões.

A partir das classes definidas pelas regiões trazidas pelo banco de dados, classificou-se a imagem da área de estudo, formando nova imagem classificada. A Figura 19 ilustra o resultado dessa classificação por meio dos agrupamentos dos atributos dos bancos de dados NoSQL.

Figura 19 - Imagem classificada pelo agrupamento dos atributos pelo banco de dados Neo4j.

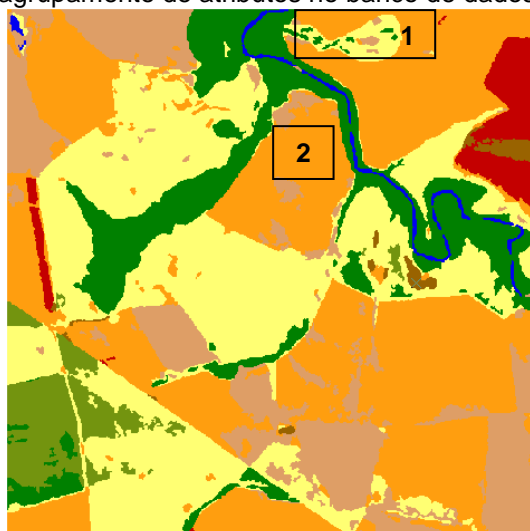


Fonte: A autora.

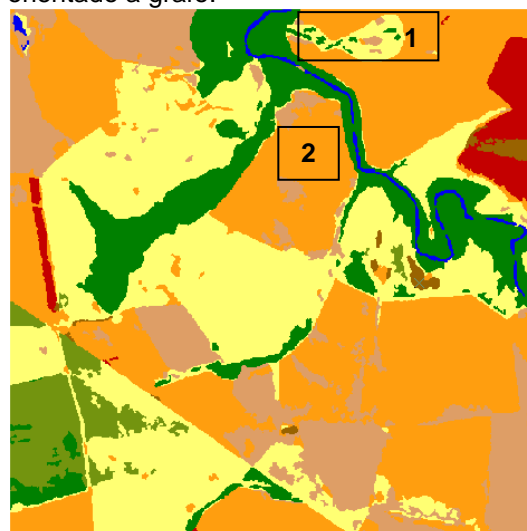
A classificação da imagem mediante as regiões com atributos semelhantes, agrupados no banco de dados orientado a grafos, permitiu reconhecer seis classes das oito classes definidas inicialmente no processo de fotointerpretação. Entre elas, a classe mata ciliar, de cultivos e o solo exposto ficaram bem definidas. Porém, a classe rio não se classificou neste método. A explicativa se dá pelo fato de atributos que representem as características dessa classe estejam agrupados em outros grupos formado pelo dendrograma que não foram utilizados na pesquisa.

Para realizar a análise compara-se as imagens classificadas anteriormente, pelos algoritmos de classificação supervisionada comumente usados, com a classificação pelo banco de dados orientado a grafos. A Figura 20 exhibe os resultados das classificações para comparação das classes.

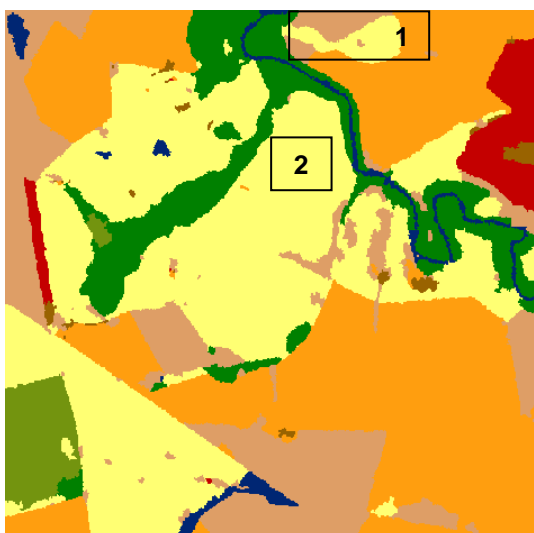
Figura 20 - Comparação das classificações supervisionadas com a classificação por meio do agrupamento de atributos no banco de dados NoSQL orientado a grafo.



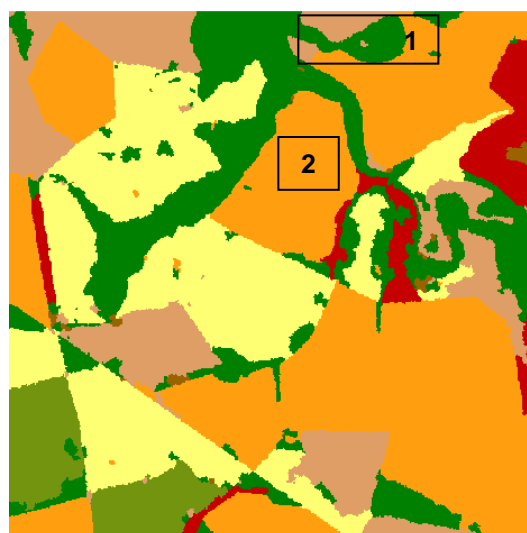
(a) Classificação pelo algoritmo MinDist



(b) Classificação pelo algoritmo MaxVer



(c) Classificação pelo algoritmo KNN



(d) Classificação pelo agrupamento do grafo



Fonte: A autora.

Observa-se pelo destaque da forma retangular 1, pelos algoritmos MinDist e MaxVer (Figura 20a e 20b), a região classificou-se como sendo maior parte cultivo 3 com pequenas regiões de mata ciliar, para o algoritmo KNN (Figura 20c), classificou em sua totalidade como cultivo 3. Já para o método de agrupamento pelo grafo (Figura 20d), classificou na sua totalidade como mata ciliar. Verificando com as classes pré-definidas na fotointerpretação, essa região é mata ciliar, permitindo

assim, afirmar que a classificação pelo banco de dados NoSQL melhor representou essa classe.

Para a classe solo exposto, as classificações por meio do grafo e pelos algoritmos MinDist, MaxVer encontraram as mesmas regiões, já o algoritmo por regiões KNN não definiu essa classe na sua totalidade. Estes são bons resultados para a classificação pelo banco de dados orientado a grafos, pois essa foi capaz de detectar as regiões pertencentes a essas classes por meio do agrupamento dos seus atributos.

Outro destaque representado na forma retangular 2 é a classe cultivo 2, que de acordo com a fotointerpretação, classificou-se corretamente com os algoritmos MinDist, MaxVer e pelo banco de dados NoSQL. Já para o algoritmo KNN a região foi classificada como cultivo 3.

Os algoritmos MinDist, MaxVer e KNN classificaram de forma adequada a classe rio, entretanto o KNN apresentou melhor delimitação do rio. A classificação pelo grafo não teve resultado para essa classe, isso se deve ao fato dos atributos que apresentam características espaciais, espectrais e de textura do rio estarem presentes em outros grupos formados pelo dendrograma, que não foram utilizados na pesquisa.

É importante validar os resultados das classificações para obter confiança nos dados apresentados. Para isso aplicaram-se métodos para avaliar a acurácia das classificações, resultando na validação dos dados apresentados.

4.6 VALIDAÇÃO DOS DADOS

Para avaliar o método de classificação que melhor representa o uso e cobertura do solo na área de estudo apenas a interpretação visual não é suficiente. Dessa forma, tornou-se necessário obter dados de referência para serem usados para gerar a matriz de confusão, conforme visto na Tabela 5 (a, b e c). O resultado da matriz de confusão permitiu avaliar e comparar os resultados obtidos a partir das classificações.

Tabela 5 - Matrizes de Confusão geradas a partir das classificações supervisionadas, em metros.

Classe	Cultivo 1	Cultivo 2	Cultivo 3	Cultivo 4	Cultivo 5	Mata ciliar	Rio	Solo exposto	Sem classe	Total
Cultivo 1	30.180	26.429	25.339	11	514	35	207	35	43	82.793
Cultivo 2	43.144	44.820	2.014	0	7.527	32	0	619	190	98.346
Cultivo 3	3.669	4.295	98.040	383	409	3.502	104	119	16	110.537
Cultivo 4	1	5	3.355	7.132	0	508	3	0	0	11.004
Cultivo 5	795	288	54	0	10.870	0	0	894	1	12.902
Mata ciliar	20	29	6.211	3.870	1	28.676	763	0	8	39.578
Rio	215	0	28	0	0	353	2.316	0	39	2.951
Solo exposto	0	24	13	0	432	0	0	1.410	10	1.889
Total	78.024	75.890	135.054	11.396	19.753	33.106	3.393	3.077	307	360.000

(a) Classificação pelo algoritmo MinDist

Classe	Cultivo 1	Cultivo 2	Cultivo 3	Cultivo 4	Cultivo 5	Mata ciliar	Rio	Solo exposto	Sem classe	Total
Cultivo 1	29.714	15.893	4.828	6	427	64	168	0	38	51.138
Cultivo 2	43.586	55.263	21.449	1	8.817	92	8	673	195	130.084
Cultivo 3	4.155	4.573	92.763	349	383	2.324	52	109	10	104.718
Cultivo 4	0	5	6.156	8.546	2	0	0	0	0	14.709
Cultivo 5	67	0	5	0	9.794	0	0	364	0	10.230
Mata ciliar	115	88	9.658	2.494	9	30.376	1.058	2	19	43.819
Rio	219	0	21	0	0	250	2.107	0	30	2.627
Solo exposto	168	68	174	0	321	0	0	1.929	15	2.675
Total	78.024	75.890	135.054	11.396	19.753	33.106	3.393	3.077	307	360.000

(b) Classificação pelo algoritmo MaxVer

Classe	Cultivo 1	Cultivo 2	Cultivo 3	Cultivo 4	Cultivo 5	Mata ciliar	Rio	Solo exposto	Sem classe	Total
Cultivo 1	46.070	3.141	2.161	0	0	0	468	369	121	52.330
Cultivo 2	8.955	79.337	243	0	2.566	0	0	0	0	91.101
Cultivo 3	24.782	0	117.828	0	0	1.779	4.426	18	0	148.833
Cultivo 4	0	0	0	11.396	0	0	0	0	0	11.396
Cultivo 5	0	0	0	0	8.636	0	0	1.132	0	9.768
Mata ciliar	0	0	5.012	0	0	30.029	6.085	0	0	41.126
Rio	0	0	0	0	0	0	3.799	0	88	3.887
Solo exposto	0	24	0	0	0	0	0	1.510	25	1.559
Total	79.807	82.502	125.244	11.396	11.202	31.808	14.778	3.029	234	360.000

(c) Classificação pelo algoritmo KNN

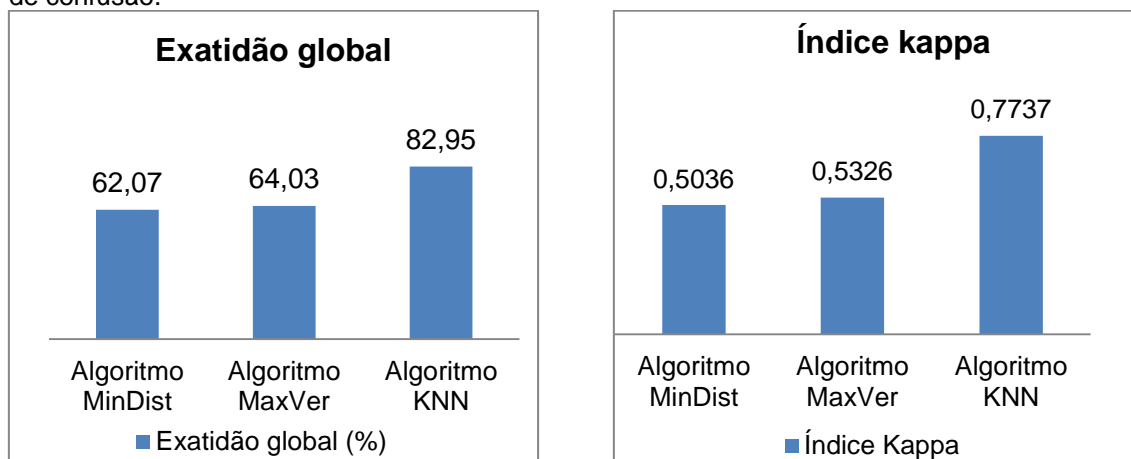
Fonte: A autora.

Como se observa na matriz de confusão de ambas as classificações presente na Tabela 5 (a, b e c), as classes de cultivos 1, 2 e 3 apresentaram grande confusão de classes, isso se dá pelo fato de que uma região classificada como cultivo, por exemplo, ter uma resposta espectral similar a outra região de cultivo em estágio inicial, dificultando a separação das classes.

Os valores apresentados na matriz de confusão da classificação pelo algoritmo KNN, visto na Tabela 5c, obteve pouca confusão entre as classes, principalmente a classe cultivo 4 que não teve divergência entre as classes. As classes cultivo 5, mata ciliar e solo exposto também apresentaram pouca confusão entre si. Já para as classes de cultivo 1, 2 e 3 tiveram resultados confusos na separação das classes, porém obteve melhores resultados comparados com os algoritmos MinDist e MaxVer.

A matriz de confusão permite ainda, o cálculo do Índice kappa que mede a acurácia das classificações em relação à verdade terrestre e calcular os erros de inclusão e omissão de amostras coletadas, como exibido na Figura 21.

Figura 21 - Resultados do cálculo da exatidão global e do índice kappa calculados por meio da matriz de confusão.



Fonte: A autora.

Observando a comparação dos resultados das classificações, destaca-se a classificação por meio do algoritmo KNN com exatidão global de 82,95% e índice kappa de 0,7737, classificando-se como “muito bom” (FONSECA, 2000). Nota-se que alcançou a mesma classificação resultante no trabalho de Folharini e Souza (2016), com índice 0,85. Os índices kappa para os algoritmos MinDist (0,5036) e

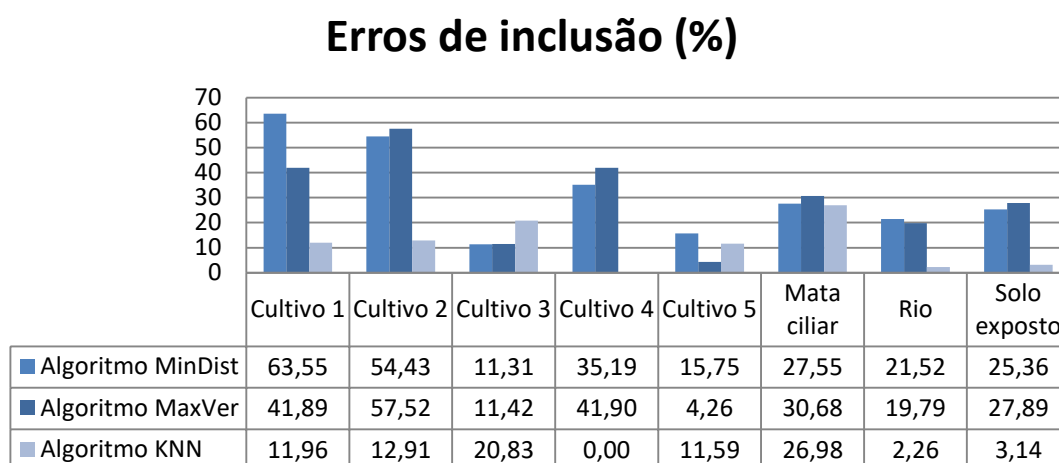
MaxVer (0,5326) se classificam como “bom”, também apresentando resultados positivos para os classificadores.

A acurácia para os algoritmos MinDist (62,07%) e MaxVer (64,03%) mantiveram variação de 2% de diferença, sendo que os dados de amostras foram os mesmos para as duas classificações.

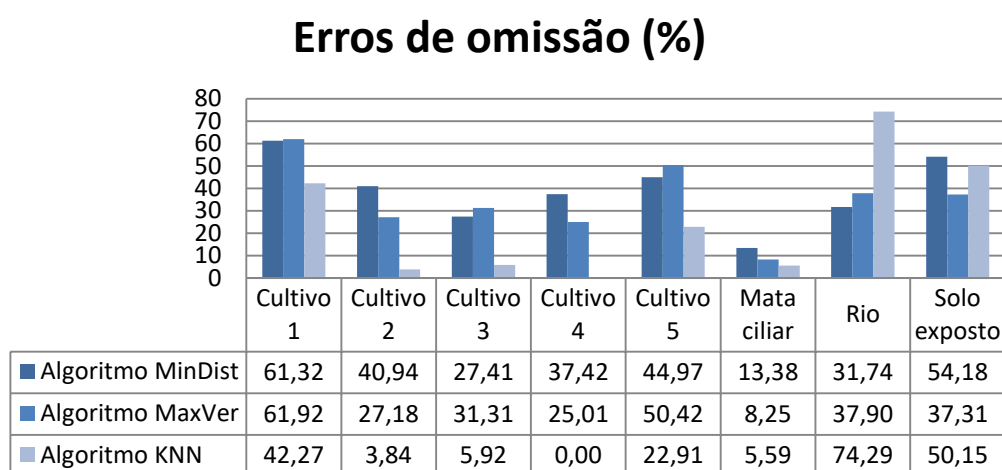
Pelos valores do índice kappa e exatidão global pode-se dizer que o algoritmo KNN melhor classificou a imagem Rapideye da área de estudo. Outro resultado gerado pela matriz de confusão foram os cálculos dos erros de inclusão e omissão, apresentados na

Figura 22.

Figura 22 - Resultados dos cálculos de erros e omissões gerados pela matriz de confusão.



(a) Erros de inclusão



(b) Erros de omissão

Fonte: A autora.

O maior erro de inclusão (63,55%) ocorreu na classe cultivo 1 pelo algoritmo MinDist, que, de acordo com a matriz de confusão apresentada, as classes cultivo 2 e 3 foram as que mais contribuíram para este erro. A classe cultivo 4 não apresentou erros de inclusão para o algoritmo KNN, porém para o MinDist (35,19%) e MaxVer (41,90%) teve erros significativos.

Os erros de omissão mais expressivos foram para as classes rio (74,29%), cultivo 1 (61,32%) e cultivo 2 (61,92%). A classe rio com o algoritmo KNN teve alta porcentagem de erros omitidos e mesmo assim, observando a imagem classificada por esse algoritmo foi a que melhor classificou a classe rio, sem falhas no seu percurso. Em relação aos erros das classes de cultivos 1 e 2, é destacado novamente a confusão gerada entre essas classes.

Com a validação feita pela matriz de confusão observou-se que a melhor classificação foi realizada pela classificação por regiões com o algoritmo KNN. Portanto, escolheu-se essa classificação para validar o método pelo banco de dados NoSQL, devido ao fato da imagem classificada pelo banco de dados não possuir o mesmo padrão de imagem de entrada no software realizado a matriz de confusão.

A partir das classificações pelo algoritmo KNN e banco de dados NoSQL orientado a grafos realizou-se a tabulação cruzada, cujo resultado pode ser visto na Tabela 6.

Tabela 6 - Resultado da tabulação cruzada entre a imagem classificada pelo algoritmo KNN e pelo banco de dados NoSQL orientado a grafos, em m².

CLASSE	Cultivo 1	Cultivo 2	Cultivo 3	Cultivo 4	Cultivo 5	Mata Ciliar	Solo Exposto	Rio
Cultivo 1	623.225	363.925	96.500			77.975	3.275	
Cultivo 2	354.275	2.432.475	475.925				4.100	
Cultivo 3	106.750		1.921.525				1.500	
Cultivo 4			215.875	249.900				
Cultivo 5	72.775		55.175		274.500	29.350		25.975
Mata Ciliar	223.150	82.200	316.500	14.950	5.725	778.800	23.125	124.925
Solo Exposto	7.675	275	7.250		825		26.675	2.925

Fonte: A autora.

Observa-se que a classe cultivo 1 sobrepôs as demais classes, exceto o cultivo 4, evidenciando a dificuldade em diferenciar as classes de interesse. Na classe cultivo 2 houve sobreposição com a classe cultivo 1, mata ciliar e solo exposto.

Para a classe cultivo 3 teve sobreposição em todas as demais classes, com maior divergência com a classe cultivo 2. Já com a classe mata ciliar sobrepôs 5.725m e com a classe solo exposto 825. A classe cultivo 5 não foi identificada em outros cultivos (1 ao 4) e com ela mesmo, identificou a área de 274.500m.

No cruzamento dos dados, a classe mata ciliar, sobrepôs 778.800m da classe mata ciliar, porém, os valores sobrepostos pelas outras classes resultam em 50% de acerto da classificação.

Com isso, pode-se dizer que os valores de cruzamento entre as classes demonstram que os métodos de classificação tiveram resultados semelhantes ao classificar as áreas nas mesmas classes definidas durante a fotointerpretação. Dessa forma, se valida o resultado da classificação da imagem da área de estudo por meio do banco de dados NoSQL orientado a grafos.

5 CONCLUSÃO

Esta pesquisa teve como objetivo a classificação do uso e cobertura do solo da imagem Rapideye da área de estudo por meio do banco de dados NoSQL orientado a grafos. Para isso, utilizou-se o método de segmentação mediante GEOBIA, agrupamento de dados pela AVM, classificação supervisionada pelos algoritmos de MinDist, MaxVer e KNN.

O uso da imagem Rapideye composta por 5 bandas espectrais, entre elas a *red edge*, foi determinante para que obtivesse resultados significativos. Pois, tratando-se de imagem de alta resolução realizou a classificação supervisionada por regiões. A aplicação do NDVI na imagem, deixando-a com 6 bandas espectrais, auxiliou na separação das classes, principalmente nas classes de cultivos e mata ciliar. Assim, foi possível delimitar as classes de interesse na imagem por meio da fotointerpretação com as composições coloridas mediante as bandas espectrais da Rapideye e NDVI.

A segmentação da imagem mediante GEOBIA separou a imagem da área de estudo em regiões, com isso, determinou os valores de escala e fusão que melhor representassem a separação dos mesmos. Dessa forma, se obteve os atributos espaciais, espectrais e de textura armazenados no banco de dados relacional. Ressalta a importância da escolha dos índices de escala e fusão para o processo de segmentação, de maneira que não ocorra a perda de informação das regiões de interesse e a diminuição de ruídos, auxiliando no processo de classificação.

Realizar o agrupamento dos dados pela AVM, gerando o dendrograma, possibilitou a separação dos grupos de atributos resultantes da GEOBIA, com a intenção de criar as *cypher queries* para consultas de regiões, com atributos semelhantes, no banco de dados NoSQL.

O armazenamento e manipulação dos dados no banco de dados NoSQL orientado a grafos permitiu analisar o agrupamento de regiões similares por meio do relacionamento dos atributos espaciais, espectrais e de textura. Com essas regiões foi possível definir a quais classes de interesse pertenciam e classificar a imagem da área de estudo. Destacam-se os desenvolvimentos dos programas utilizados para inserir e consultar dados no banco de dados auxiliaram na agilidade desses processos.

Com as classificações por diferentes métodos realizou-se a matriz de confusão, para comparação entre os algoritmos classificadores, a fim de validar o resultado da classificação por meio dos agrupamentos do banco de dados NoSQL orientado a grafos. Entre os algoritmos de classificação comumente encontrados na literatura, MinDist, MaxVer e KNN, o que melhor separou as classes foi o algoritmo KNN com maior acurácia e índice kappa 0,7737. Portanto, foi o escolhido para a validação da imagem classificada pelo agrupamento de regiões do banco de dados NoSQL por meio da tabulação cruzada.

O método da tabulação cruzada entre as classes das imagens pelo método da classificação pelo banco de dados NoSQL orientada a grafos e pelo algoritmo KNN, evidenciou que as classificações tiveram resultados similares. Os valores do cruzamento entre as imagens classificadas permitiu validar e afirmar a possibilidade de realizar a classificação da imagem por meio do agrupamento de regiões no grafo formado pelo banco de dados NoSQL.

Pode-se concluir que os métodos utilizados no estudo alcançaram os objetivos propostos nessa pesquisa, mostrando resultados satisfatórios, de maneira que apresentaram eficiência e agilidade na classificação do uso e cobertura do solo da área de estudo.

5.1 TRABALHOS FUTUROS

Como trabalhos futuros referentes a esta pesquisa, pretende-se:

- Testar outros valores da linha de corte no dendrograma para obter outros grupos de atributos.
- Testar no mesmo banco de dados NoSQL orientado a grafos, outros métodos para a seleção de atributos.
- Acrescentar no banco de dados NoSQL orientado a grafos outros nós do tipo classificação, além de baixo, médio e alto.
- Aplicar a metodologia proposta testando outros métodos de segmentação e comparar com a GEOBIA.

REFERÊNCIAS

ABREU, E. S. et al. Armazenamento e Processamento de Grandes Grafos em Bancos de Dados Geográficos. In: **XIII Brazilian Symposium on Geoinformatics**, São Paulo, 2012.

ACHARYA, T. D.; YANG, I. T.; LEE, D. H. Land Cover Classification Using a KOMPSAT-3A Multi-Spectral Satellite Image. **Applied Sciences**, v. 6, n. 11, p. 371, 2016.

AGUIAR, T. V. et al. Anati quanti: software de análises quantitativas para estudos em anatomia vegetal. **Planta Daninha**, Viçosa-MG, v. 25, n. 4, p. 659-659, 2007.

ALMEIDA, R. C. D.; BRITO, P. F. D. Utilização da Classe de Banco de Dados NOSQL como Solução para Manipulação de Diversas. In: **Encontro de Computação e Informática do Tocantins**, p. 152-161, 2012.

ANDERSON, J. C.; LEHNARDT, J.; SLATER, N. CouchDB: The Definitive. **O'REILLY**, San Francisco, p. 272, 2010.

ANDRADE, A. F. **Integração de Variáveis Espectrais e Forma na Classificação de Imagens de Alta Resolução Utilizando Redes Neurais Artificiais**. Universidade Federal do Paraná. Curitiba, p. 112. 2003.

BACANI, V. M.; SAKAMOTO, A. Y.; LUCHIARI, A. E. Q. H. Sensoriamento Remoto e SIG Aplicados à Avaliação da Fragilidade Ambiental de Bacia Hidrográfica. **Mercator (Fortaleza)**, Fortaleza, v. 14, n. 2, p. 119-135, 2015.

BLASCHKE, T.; LANG, S.; HAY, G. **Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline**. New York: [s.n.], 2008.

BRITO, C. R. et al. O Uso de SIG no Inventário de Árvores no Campus do Instituto Politécnico de Bragança, Portugal. **Geografia Ensino e Pesquisa**, v. 16, n. 3, 2012.

CÂMARA NETO, G. et al. **Anatomia de Sistemas de Informação Geográfica**. Campinas, SP: [s.n.], 1996.

CASANOVA, M. A. et al. **Bancos de Dados Geográficos**. Curitiba: Mundogeo, 2005.

CATTANI, C. E. V. et al. Desempenho de Algoritmos de Classificação Supervisionada para Imagens dos Satélites RapidEye. In: **Simpósio Brasileiro de Sensoriamento Remoto**, Foz do Iguaçu, PR, 2013.

CHODOROW, K.; DIROLF, M. MongoDB: The Definitive Guide. **O'REILLY**, p. 216, 2010.

COHENCA, D.; CARVALHO, R. Comparação de Métodos de Classificação OBIA, Máxima Verossimilhança e Distância Mínima em Imagem OLI/Landsat-8 em Área de Alta Diversidade de Uso do Solo. In: **Simpósio Brasileiro de Sensoriamento Remoto**, João Pessoa-PB, Brasil, p. 25 a 29, Abril 2015.

COSTA, E. M. et al. Processamento de Imagens RapidEye no Mapeamento de Uso do Solo em Ambiente de Mar de Morros. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 9, p. 1417-1427, Setembro 2016.

DATE, C. J. **An Introduction to Database Systems**. 8ª. ed. Boston, MA: Addison Wesley, 2004.

FACON, J. **Processamento e Análise de Imagens**. Pontifícia Universidade Católica do Paraná. Curitiba. 2002.

FOLHARINI, S. O.; SOUZA, S. O. Classificação GEOBIA Aplicada a Gestão de Bacias Hidrográficas: o Caso da Microbacia do Córrego Monte Alegre. **Revista Eletrônica do PRODEMA**, Fortaleza, Brasil, v. 10, n. 2, p. 37-48, Julho 2016. ISSN 1982-5528.

FONSECA, L. M. G. Processamento digital de imagens. **Instituto Nacional de Pesquisas**, 2000.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens**. São Paulo: Pearson Prentice Hall, 2010.

HUNGER, M. Neo4j: Java-based NoSQL Graph Database. **InfoQ**, 2010. Disponível em: <<http://www.infoq.com/news/2010/02/neo4j-10>>. Acesso em: 24 Nov. 2015.

IBM. IBM D2M User's Guide and Reference. **IBM**, 2002. Disponível em: <<http://www.ibm.com/>>. Acesso em: 15 Nov 2015.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma Perspectiva em Recursos Terrestres**. 2. ed. São José dos Campos: [s.n.], 2009.

LAKSHMAN, A.; MALIK, P. Cassandra: a Decentralized Structured Storage. **ACM SIGOPS Operating System Review**, v. 44, p. 35-40, Abril 2010.

LIZARDO, L. E. O.; MORO, M. M.; DAVIS JR, C. A. GeoNoSQL: Banco de Dados Geoespacial em NoSQL. In: **Computer on the Beach 2014**, Florianópolis, p. 303-309, Março 2014. ISSN 2358-0852.

LÓSCIO, B. F.; OLIVEIRA, H. R. D.; PONTES, J. C. D. NoSQL no Desenvolvimento de Aplicações Web Colaborativas. In: **Simpósio Brasileiro de Sistemas Colaborativos**, Paraty, v. 8, n. Anais.Paraty: SBC, 2007., 2011.

MACIEL, A. M.; VINHAS, L. C. G. Algoritmos de Clustering para Separação de Culturas Agrícolas e Tipos de Uso e Cobertura da Terra Utilizando Dados de

Sensoriamento Remoto. In: **Simpósio Brasileiro de Sensoriamento Remoto**, João Pessoa-PB, Brasil, Abril 2015.

MATSUSHITA, R. S. I.; NGUESSAN, D. Framework para integração de Serviços Móveis Baseado em Rede Social. **Fasci-Tech – Periódico Eletrônico da FATEC**, 2011. 141-153.

MICROSOFT. White Paper: Delivering Location Intelligence With Spatial Data. **Microsoft**, 2008. Disponível em: <<http://www.microsoft.com>>. Acesso em: 15 nov 2015.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e Metodologias de Aplicação**. 3ª. ed. [S.I.]: UFV, 2007.

NEO TECHNOLOGY, I. What is a Graph Database? **Neo4j**, 2014. Disponível em: <<http://neo4j.com/developer/graph-database/>>. Acesso em: 13 nov 2015.

NERY, C. V. M. et al. Avaliação das Técnicas de Classificação MAXVER, MAXVER – ICM e Distância Mínima Euclidiana de Acordo com Índice Kappa. **Revista Brasileira de Geografia Física**, v. 6, n. 2, p. 320-328, 2013. ISSN 1984-2295.

ORACLE. Spatial Developer's Guide. **ORACLE**, 2016. Disponível em: <http://docs.oracle.com/cd/B28359_01/appdev.111/b28400/sdo_intro.htm>. Acesso em: 15 out 2016.

PLANET. Planet Imagery Product Specification: Planetscope & Rapideye, 2017. Disponível em: <<https://www.planet.com>>. Acesso em: 30 mar 2017.

POKORNY, J. NoSQL Databases: A Step to Database Scalability in Web Environment. **International Journal of Web Information Systems**, v. 9, n. 1, p. 69 - 82, 2013.

POSTGIS. PostGIS 2.0 Manual. **PostGis**, 2005. Disponível em: <<http://postgis.net/>>. Acesso em: 15 Nov 2015.

PRADO, D. F. C.; CARVALHO, L. M. T. Application of Object-Based Accuracy Assessment for Land Cover Classification Using Rapideye Images in Southeastern Brazil. In: **GEOBIA 2016 : Solutions and Synergies**, set 2016.

PRICHOA, C. E. **Estudo em Duas Unidades de Paisagem da Bacia Hidrográfica do Rio Pitangui/ PR Mediante Estatística Multivariada e Análise Orientada a Objetos**. Universidade Estadual de Ponta Grossa. Ponta Grossa, p. 140. 2012.

PRINA, B. Z.; TRENTIN, R. GMC: Geração de Matriz de Confusão a Partir de uma Classificação Digital de Imagem do ArcGIS. **Simpósio Brasileiro de Sensoriamento Remoto**, João Pessoa-PB, Brasil, 25-29 abr 2015.

QUEIROZ, G. R.; MONTEIRO, A. M. V.; CÂMARA, G. Bancos de Dados Geográficos e Sistemas Nosql: Onde Estamos e para Onde Vamos. **Revista Brasileira de Cartografia**, São José dos Campos, p. 479-492, 2013.

RIBEIRO, E. B.; ALMEIDA, F. L. B.; RIBEIRO, S. R. A. Importância dos Valores de Escala e Fusão na Segmentação Mediante GEOBIA. **In: IV Congresso Agropecuário, Industrial e Tecnológico do Paraná**, Anais. Ponta Grossa: UEPG, 2016, 2016.

RIBEIRO, S. R. A.; BAHR, H. P.; CENTENO, J. S. Integração de Imagens de Satélite e Dados Complementares para a Delimitação de Unidades de Paisagem Usando uma Abordagem Baseada em Regiões. **Boletim Ciências Geodésicas**, Curitiba, v. 8, n. 1, p. 47-57, 2002.

ROBBINSON, D.; REDDING, N.; CRISP, D. Implementation of a Fast Algorithm for Segmenting SAR Imagery. **DSTO - Electronics and Surveillance Research Laboratory**, 2002.

SANTOS, A. A. et al. Análise Comparativa Entre Classificação Orientada ao Objeto e a Fotointerpretação do Uso do Solo no Parque Estadual da Serra do Rola Moça, MG. **in: Simpósio Brasileiro de Sensoriamento Remoto**, Foz do Iguaçu, PR, Brasil, abr 2013.

SANTOS, P. G. et al. Agrupamento de Pedons de Cambissolos Húmicos com Base em Atributos Físicos e Químicos Utilizando a Estatística Multivariada. **Revista Brasileira de Ciência do Solo**, 2015. 350-360.

SCHUSTER, C.; FÖRSTER, M.; KLEINSCHMIT, B. Testing the Red Edge Channel for Improving Land-use Classifications Based on High-resolution Multi-spectral Satellite Data. **International Journal of Remote Sensing**, Berlin, Alemanha, v. 33, p. 5583-5599, set 2012.

SILVA, A. J. F. M.; EDUARDO, B. F. S.; FAZAN, A. J. Avaliação da Qualidade Geométrica das Imagens RapidEye Ortoretificadas. **In: Simpósio Brasileiro de Sensoriamento Remoto**, Foz do Iguaçu, PR, Brasil, p. 8, abr 2013.

SILVA, F. M. M. **Integração de Dados de Imagens Orbitais de Alta Resolução e ALS para Detecção Semi-Automática de Edificações em Áreas Urbanas**. Universidade Federal do Paraná. Curitiba, p. 181. 2012.

SILVA, L. R.; SANO, E. E. Análise das Imagens do Satélite Rapideye para Discriminação da Cobertura Vegetal do Bioma Cerrado. **Revista Brasileira de Cartografia**, Rio de Janeiro, Brasil, v. 68, n. 7, p. 1269-1283, jul/ago 2016. ISSN 1808-0936.

SOARES, D. O.; SANTOS, A. C.; SILVA, E. B. Análise das Classificações Supervisionada e Não Supervisionada com o Uso de Imagens Landsat 5 TM e RapidEye, e suas Contribuições para o Mapeamento do Uso e Cobertura do Solo no

Parque Nacional das Emas. **Revista Brasileira de Geomática**, Pato Branco, PR, Brasil, v. 4, n. 2, p. 110-115, mai/ago 2016.

SOLUTIONS, I. V. I. ENVI Feature Extraction Module User's Guide. **Feature Extraction Module Version 4.6**, p. 26, Dezembro 2008.

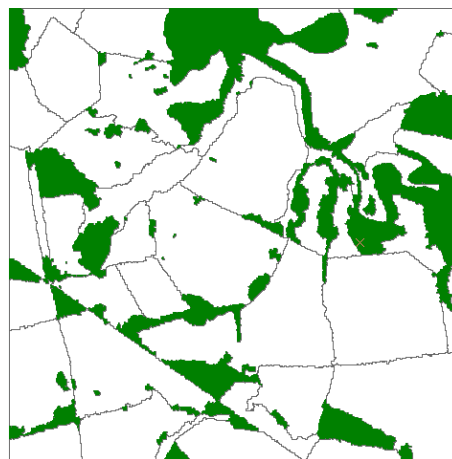
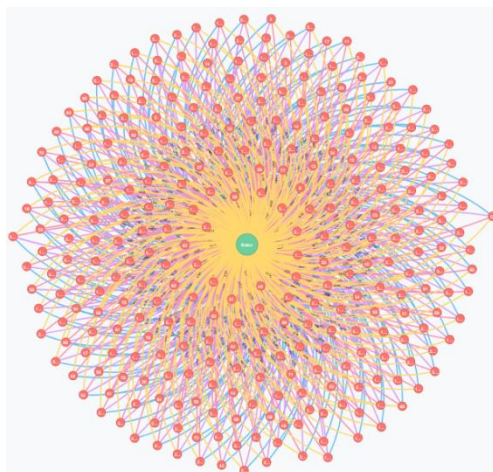
SOUZA, C. H. W. et al. Estudo da Banda Red Edge do Satélite Rapideye na Discriminação da Cobertura Vegetal. **in: Simpósio Brasileiro de Sensoriamento Remoto**, João Pessoa-PB, Brasil, 25 a 29 abr 2015. 1-7.

WIGGERS, K. L. **Análise em uma Imagem Orbital de Alta Resolução para Classificação do Uso e Cobertura da Terra de uma Área da Bacia do Pitangui - PR**. Universidade Estadual de Ponta Grossa. Ponta Grossa, p. 123. 2014.

APÊNDICE A – CONSULTA DOS ATRIBUTOS NO BANCO DE DADOS NEO4J

Grupo 1: Eixo_Menor, Eixo_Maior, Comprimento, Area

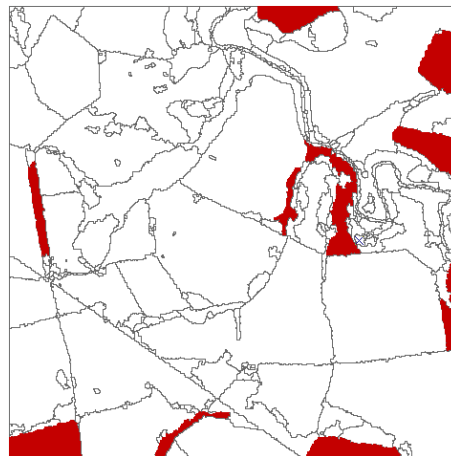
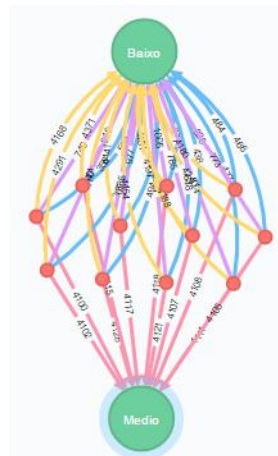
Agrupamento: 1111- baixo



Classe: Mata ciliar

Regiões: 195, 193, 242, 216, 310, 187, 183, 182, 279, 178, 179, 180, 213, 163, 194,
 167, 226, 203, 327, 227, 273, 336, 342, 334, 241, 326, 316, 191, 164, 160, 171, 189,
 190, 186, 199, 344, 220, 127, 236, 298, 340, 153, 200, 138, 154, 276, 211, 177, 181,
 175, 270, 265, 128, 258, 173, 172, 170, 337, 281, 343, 315, 168, 272, 165, 218, 131,
 290, 206, 146, 214, 162, 275, 159, 269, 307, 222, 333, 157, 158, 122, 124, 250, 155,
 267, 152, 341, 116, 151, 120, 223, 253, 286, 114, 312, 264, 118, 252, 332, 149, 221,
 284, 147, 148, 208, 145, 143, 280, 328, 106, 144, 331, 141, 278, 197, 308, 108, 314,
 245, 244, 139, 295, 136, 330, 224, 98, 137, 132, 201, 133, 255, 294, 125, 247, 302,
 129, 271, 126, 121, 289, 119, 117, 299, 225, 207, 338, 115, 322, 123, 112, 260, 239,
 113, 111, 86, 64, 110, 109, 257, 105, 103, 248, 196, 100, 99, 287, 101, 266, 292, 96,
 285, 95, 97, 92, 91, 90, 277, 88, 288, 204, 89, 85, 303, 249, 84, 83, 80, 229, 231, 329,
 339, 256, 78, 79, 75, 304, 325, 268, 297, 215, 77, 313, 230, 74, 72, 301, 71, 246, 73,
 69, 296, 320, 323, 263, 67, 232, 198, 324, 318, 251, 65, 202, 63, 61, 60, 59, 58, 210,
 309, 233, 57, 52, 54, 53, 49, 205, 51, 50, 345, 209, 228, 317, 48, 319, 46, 47, 45, 235,
 44, 219, 254, 335, 42, 40, 238, 41, 39, 30, 31, 35, 32, 36, 33, 237, 243, 259, 282, 262,
 321, 305, 306, 300, 311, 240, 261, 274, 29, 27, 217, 212, 21, 19, 23, 22, 28, 293, 24,
 26, 234, 20, 17, 13, 15, 14, 16, 9, 1, 5, 2, 6, 10, 11, 291, 12, 283

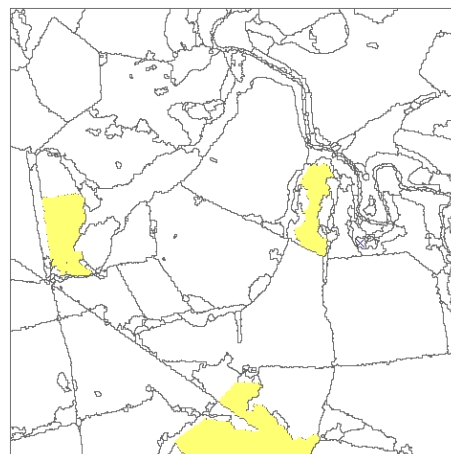
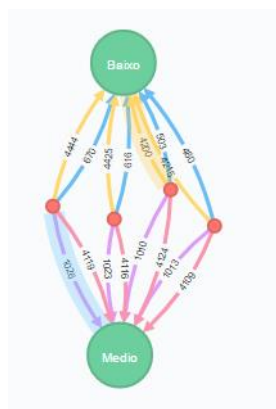
Agrupamento: 1211 -



Classe: Cultivo 5

Regiões: 188, 184, 192, 142, 93, 87, 82, 62, 70, 25, 7

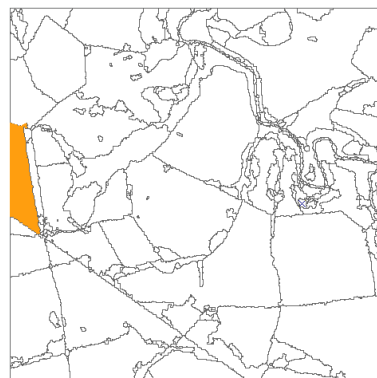
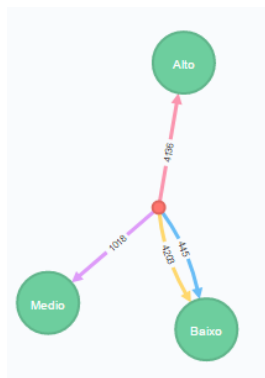
Agrupamento: 1221



Classe: Cultivo 3

Regiões: 185, 174, 102, 81

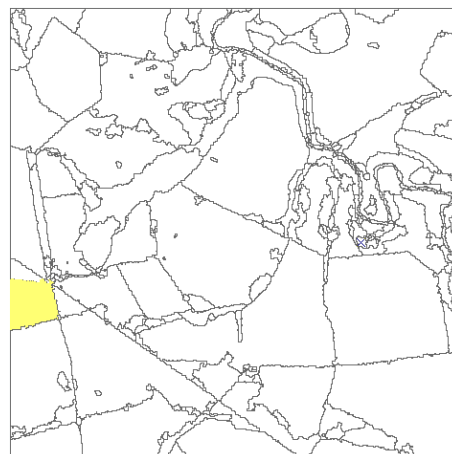
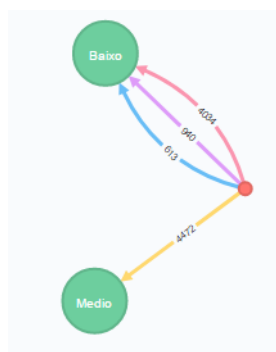
Agrupamento: 1321



Classe: Cultivo 2

Regiões: 66

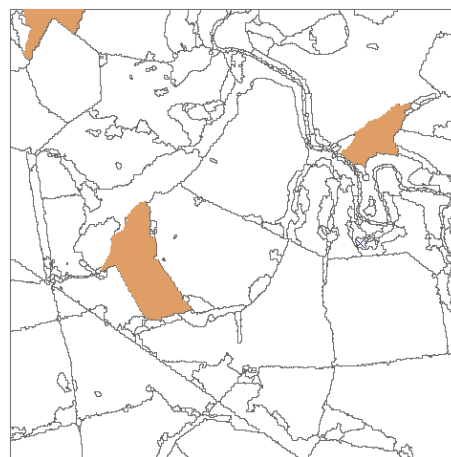
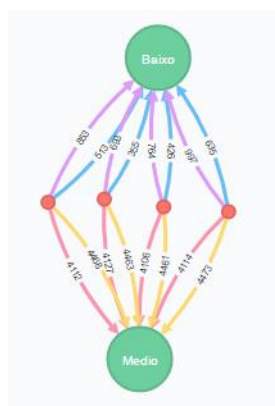
Agrupamento: 2111



Classe: Cultivo 3

Regiões: 150

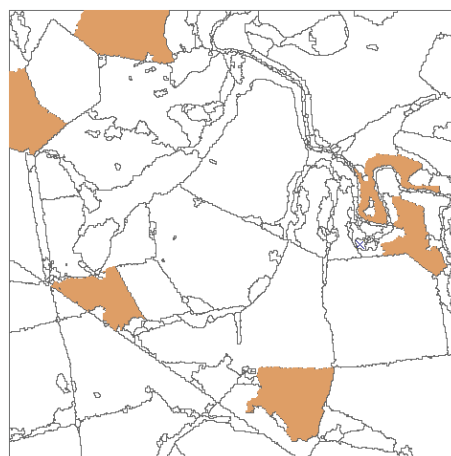
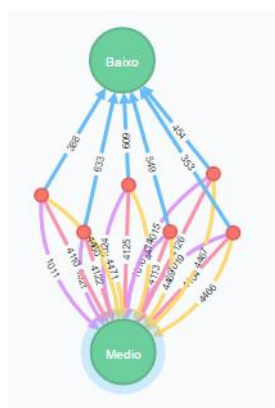
Agrupamento: 2211



Classe: Cultivo 1

Regiões: 135, 107, 55, 3

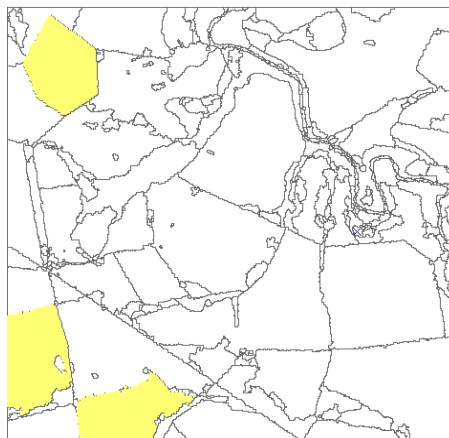
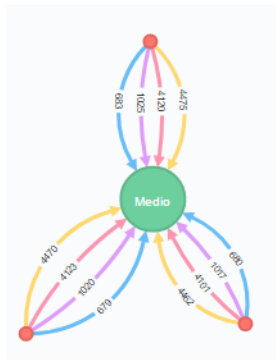
Agrupamento: 2221



Classe: Cultivo 1

Regiões: 140, 169, 104, 76, 38, 4

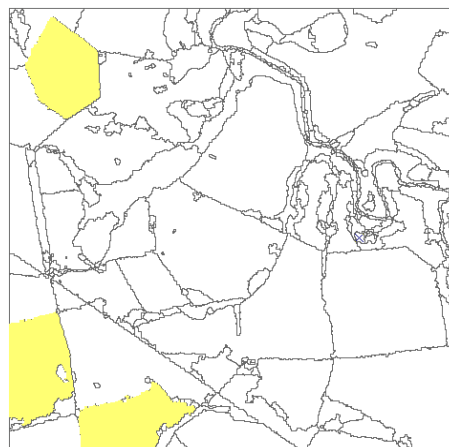
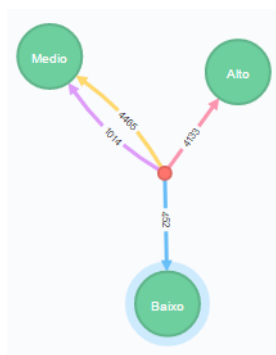
Agrupamento: 2222



Classe: Cultivo 3

Regiões: 176, 18, 161

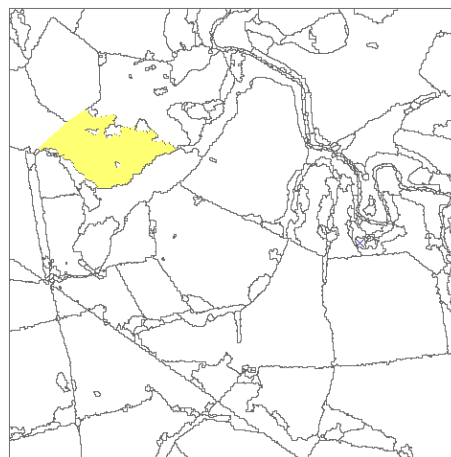
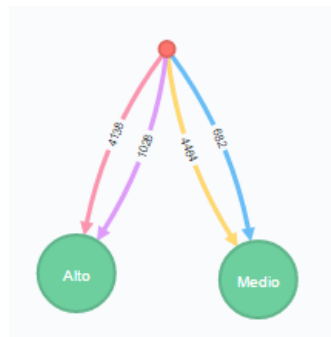
Agrupamento: 2321



Classe: Cultivo 3

Regiões: 68

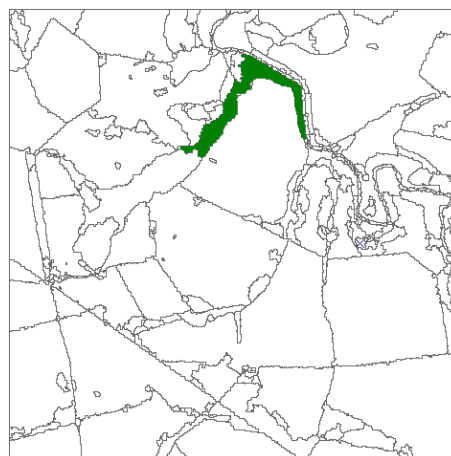
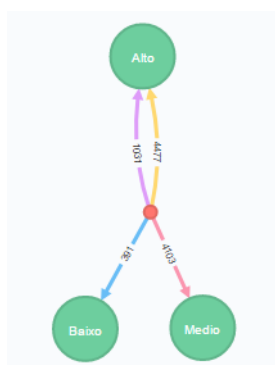
Agrupamento: 2332



Classe: Cultivo 3

Regiões: 56

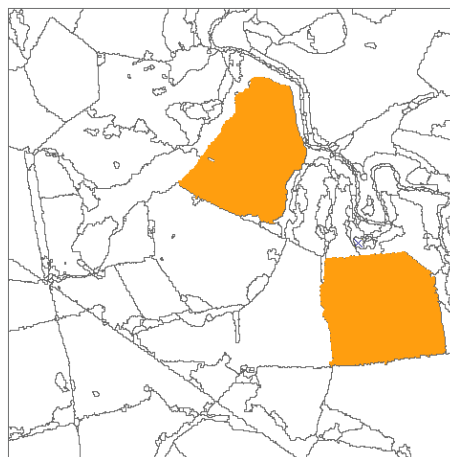
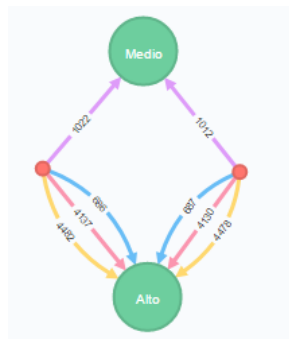
Agrupamento: 3231



Classe: Mata ciliar

Regiões: 37

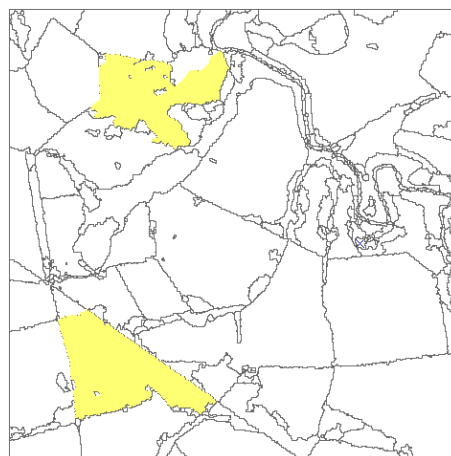
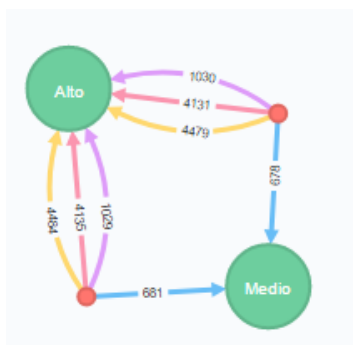
Agrupamento: 3323



Classe: Cultivo 2

Regiões: 43, 134

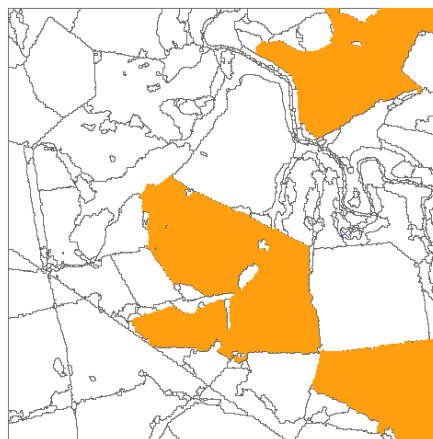
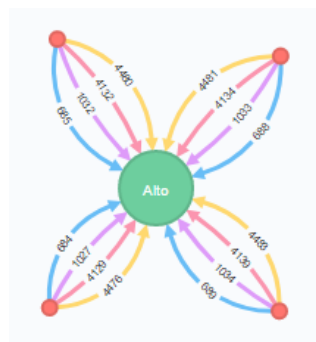
Agrupamento: 3332



Classe: Cultivo 3

Regiões: 156, 34

Agrupamento: 3333

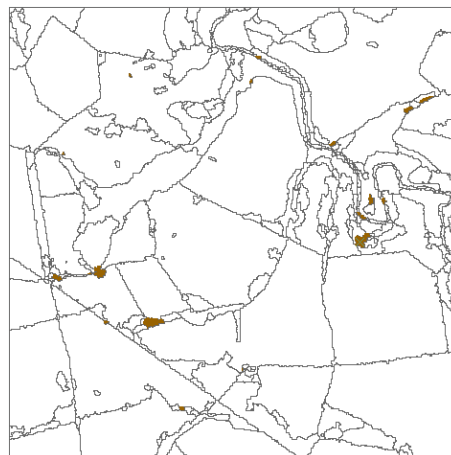
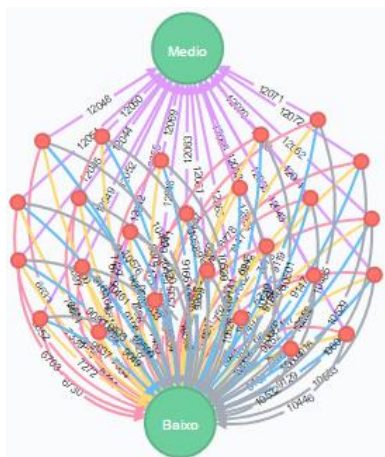


Classe: Cultivo 2

Regiões: 166, 130, 94, 8

**Grupo 2: Desviopadraobanda_3, Variancia_Text, Desviopadraobanda_2,
Desviopadraobanda_1, Intervalo_Text**

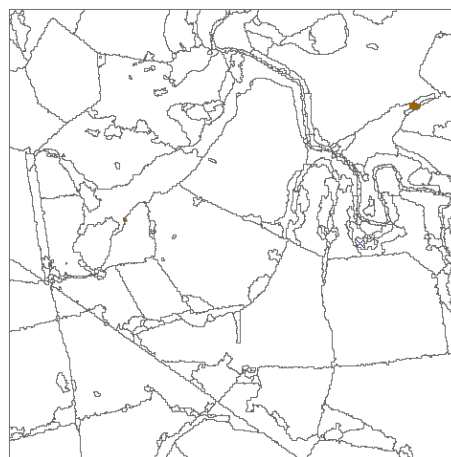
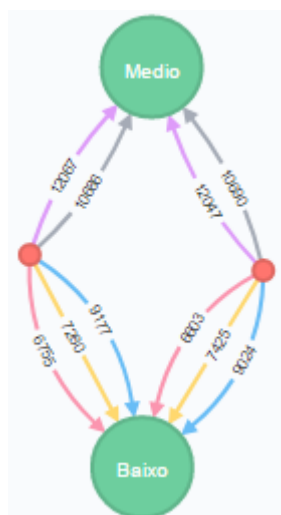
Agrupamento: 21111



Classe: Solo exposto

Regiões: 314, 279, 326, 213, 163, 275, 332, 252, 148, 236, 128, 245, 302, 262, 335,
44, 57, 230, 91, 322, 106, 257, 330, 232, 317

Agrupamento: 21211

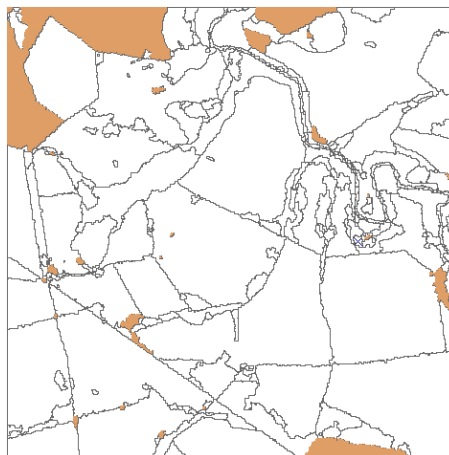
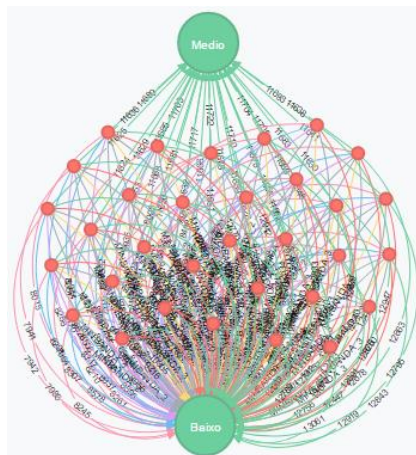


Classe: Solo exposto

Regiões: 292, 209

**Grupo 3: MAXBANDA_4, MAXBANDA_2, MEDIABANDA_3, MEDIABANDA_4,
MAXBANDA_1, MEDIABANDA_2, MINBANDA_3, MINBANDA_2, MEDIABANDA_1,
MINBANDA_1**

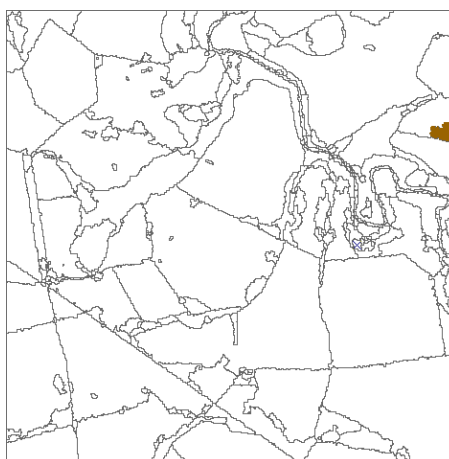
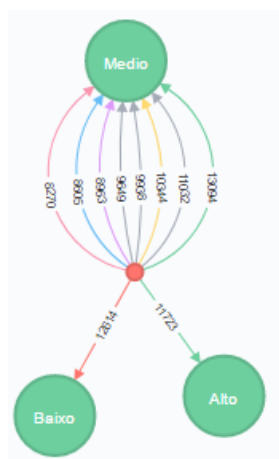
Agrupamento: 1121111111



Classe: Cultivo 1

Regiões: 195, 192, 216, 211, 298, 199, 165, 222, 160, 223, 147, 143, 144, 20, 277, 91, 330, 290, 139, 295, 218, 268, 24, 212, 27, 38, 45, 61, 1, 3, 4, 7

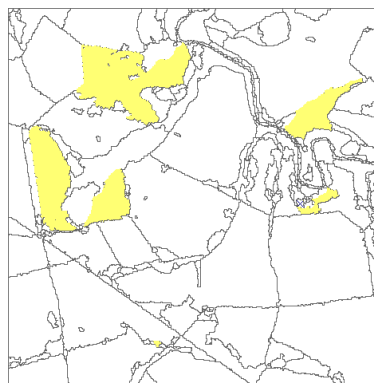
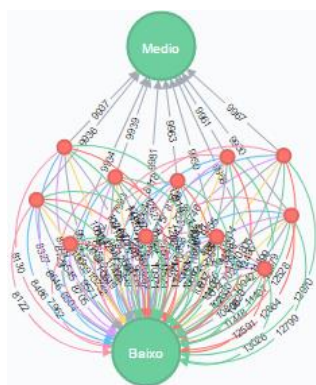
Agrupamento: 1232222222



Classe: Solo exposto

Regiões: 59

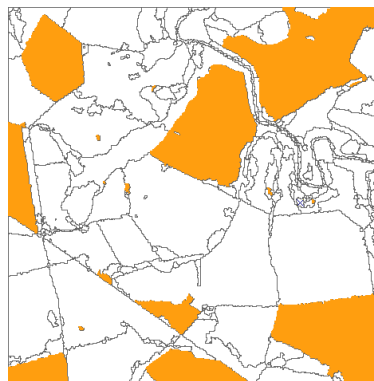
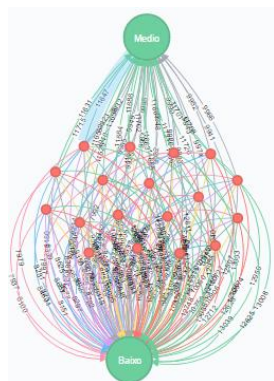
Agrupamento: 1121111111



Classe: Cultivo 3

Regiões: 200, 122, 107, 102, 73, 209, 34, 36, 57, 48, 55

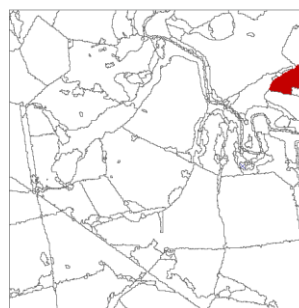
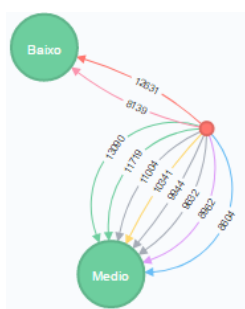
Agrupamento: 1221111111



Classe: Cultivo 2

Regiões: 188, 185, 270, 167, 168, 166, 162, 229, 116, 258, 292, 206, 18, 43, 66, 205, 317, 8

Agrupamento: 1222222221



Classe: Cultivo 5

Regiões: 46