



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE CIÊNCIAS GEOGRÁFICAS
CURSO DE BACHARELADO EM GEOGRAFIA**

JOSÉ VINICIUS ALVES

**ANÁLISE COMPARATIVA ENTRE MÉTODOS DE CLASSIFICAÇÃO
SUPERVISIONADA E NÃO SUPERVISIONADA PARA O MAPEAMENTO DO USO
E COBERTURA DA TERRA**

RECIFE

2023

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
CURSO DE BACHARELADO EM GEOGRAFIA

JOSÉ VINICIUS ALVES

**ANÁLISE COMPARATIVA ENTRE MÉTODOS DE CLASSIFICAÇÃO
SUPERVISIONADA E NÃO SUPERVISIONADA PARA O MAPEAMENTO DO USO
E COBERTURA DA TERRA**

TCC apresentado ao Curso de Geografia da
Universidade Federal de Pernambuco, como
requisito parcial para a obtenção do título de
Bacharel em Geografia.

Orientador(a): Prof^a Dr^a Josicleda
Domiciano Galvíncio

RECIFE

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Alves, José Vinicius.

ANÁLISE COMPARATIVA ENTRE MÉTODOS DE CLASSIFICAÇÃO
SUPERVISIONADA E NÃO SUPERVISIONADA PARA O MAPEAMENTO
DO USO E COBERTURA DA TERRA / José Vinicius Alves. - Recife, 2023.

65 p. : il., tab.

Orientador(a): Josicleda Domiciano Galvínio

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Filosofia e Ciências Humanas, Geografia -
Bacharelado, 2023.

1. Sensoriamento Remoto. 2. Bacia Hidrográfica. 3. Google Earth Engine
(GEE). 4. Random Forest. 5. K-means. I. Galvínio, Josicleda Domiciano.
(Orientação). II. Título.

910 CDD (22.ed.)

JOSÉ VINICIUS ALVES

**ANÁLISE COMPARATIVA ENTRE MÉTODOS DE CLASSIFICAÇÃO
SUPERVISIONADA E NÃO SUPERVISIONADA PARA O MAPEAMENTO DO USO
E COBERTURA DA TERRA**

TCC apresentado ao Curso de Geografia da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Geografia.

Aprovado em: ___/___/_____.

BANCA EXAMINADORA

Documento assinado digitalmente
 JOSICLEDA DOMICIANO GALVINCIO
Data: 22/11/2023 06:09:30-0300
Verifique em <https://validar.iti.gov.br>

Prof^a. Dr^a. Josicleda Domiciano Galvêncio (Orientador)
Universidade Federal de Pernambuco

Dr^a. Luciana Mayla de Aquino França (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Henrique dos Santos Ferreira (Examinador Externo)
Universidade Estadual do Piauí

*Dedico a todos que me suportaram
durante essa (longa) jornada.*

AGRADECIMENTOS

Agradeço, primeiramente, ao meu Senhor Jesus Cristo, que me deu forças para continuar essa caminhada em momentos difíceis.

Agradeço à minha família, minha mãe Maria José Alves e minha tia Maria do Carmo Alves do Nascimento, que mesmo com todas as dificuldades fizeram de tudo para que eu pudesse chegar aqui nesse momento. Todo meu amor a vocês.

Agradeço muito aos meus supervisores de estágio e amigos Henrique dos Santos e Marcelo Olímpio, assim como a todos do ICPS. Vocês foram parte essencial da minha jornada, espero que a vida continue nos agraciando com essa parceria.

Aos meus companheiros nesses quase 6 anos de curso, João Lucas, Florêncio, Paulo Henrique e Felipe (por um pouco menos de tempo). Não importa o que houvesse, sempre estávamos juntos, nos apoiando e dando forças uns aos outros nessa árdua jornada. Cito também Emile e Cecília que apesar de por menos tempo, também foram importantes.

Às minhas amigas Kelyane Vitória e Laís Anjos, que ouviram meus dramas, sofrimentos e alegrias e me suportaram por todo esse tempo: que ótimo ter vocês. Obrigado por tudo.

A todos os meus professores, muito obrigado por todo ensinamento transmitido.

À professora Josicleda, muito obrigado pela orientação e apoio e à Luciana Mayla, que aceitou rapidamente o convite de participar da banca.

Agradeço também à Vanessa, George e Luiz pelo tempo de estágio na Prefeitura de Paulista e Paulo, Edjane, Orlando e Ângela pelo tempo de estágio na CONDEPE/FIDEM. Foram meses de muito aprendizado.

Por fim, sou grato a todos que direta ou indiretamente fizeram parte dessa jornada. Tudo tem um propósito. Todos foram importantes. Muito obrigado.

*“já que o todo está para o maior
como o maior está para o menor
entre razões, fórmulas e mais
o que me difere de você?”*

phi – scalene (magnetite, 12, 2017)

RESUMO

O mapeamento de uso e cobertura do solo vem ganhando cada vez mais relevância no que diz respeito à gestão de recursos hídricos podendo monitorar a distribuição e a dinâmica temporal das formas de ocupação da terra. Desta forma, este estudo aborda o gerenciamento de bacias hidrográficas com foco no mapeamento de uso e cobertura da terra na Bacia Hidrográfica GL2, utilizando a plataforma Google Earth Engine (GEE). O objetivo desse estudo foi realizar o mapeamento de uso e cobertura da terra na bacia e comparar métodos de classificação supervisionados e não-supervisionados. Para isso, foram coletadas amostras espectrais para treinar os classificadores e avaliar sua precisão por meio de validação com matriz de erro, acurácia global e índice kappa. O método supervisionado, com o algoritmo Random Forest, apresentou excelente desempenho, superando a classificação não-supervisionada pelo algoritmo K-means. No entanto, houve a necessidade de reduzir a área de estudo devido a um problema não resolvido na plataforma durante a comparação dos métodos. Foram obtidos resultados que indicam uma leve superioridade ao método supervisionada através do Random Forest. Este estudo demonstra o potencial do GEE para o mapeamento de uso da terra em bacias hidrográficas e fornece insights valiosos para pesquisas futuras.

Palavras-chave: Google Earth Engine (GEE); Bacia Hidrográfica; Sensoriamento Remoto; Random Forest; K-means.

ABSTRACT

Mapping land use and cover is gaining increasing relevance with regard to water resources management, being able to monitor the distribution and temporal dynamics of forms of land occupation. Thus, this study addresses watershed management with a focus on mapping land use and cover in the GL2 Watershed, using the Google Earth Engine (GEE) platform. The objective of this study was to map land use and coverage in the basin and compare supervised and unsupervised classification methods. For this, spectral samples were collected to train the classifiers and evaluate their accuracy through validation with error matrix, global accuracy and kappa index. The supervised method, using the Random Forest algorithm, demonstrated excellent performance, surpassing the unsupervised classification by the K-means algorithm. However, there was a need to reduce the study area due to an unresolved issue in the platform during method comparison. The results obtained indicate a slight superiority of the supervised method using Random Forest. This study demonstrates the potential of GEE for land use mapping in watersheds and provides valuable insights for future research.

Keywords: Google Earth Engine (GEE); Watershed; Remote Sensing. Random Forest; K-means.

LISTA DE FIGURAS

Figura 01 – Representação do algoritmo Random Forest.....	23
Figura 02 – Representação do algoritmo K-means.....	25
Figura 03 – Resultado e Desempenho do Índice Kappa.....	26
Figura 04 – Mapa de Localização da Bacia GL2.....	28
Figura 05 – Mapa Pedológico da Bacia GL2.....	29
Figura 06 – Mapa Geomorfológico da Bacia GL2.....	30
Figura 07 – Mapa de Cobertura Vegetal da Bacia GL2.....	31
Figura 08 – Fluxograma metodológico.....	34
Figura 09 - Mosaico "True Color".....	36
Figura 10 - Mosaico "False Color".....	36
Figura 11 - Mosaico "Mangrove".....	37
Figura 12 - Mosaico "Agricultura".....	37
Figura 13 - Mosaico de Clusters.....	42
Figura 14 - Mapa gerado pelo GEE de Uso e Cobertura da Terra por Classificação Supervisionada.....	48
Figura 15 - Mapa Final de Uso e Cobertura da Terra por Classificação Supervisionada.....	52

Figura 16 - Mapa de Localização da área de interesse para comparativo de classificações.....	53
Figura 17 - Mapa comparativo de Uso e Cobertura da Terra entre Classificação Supervisionada e não-supervisionada.....	54

LISTA DE QUADROS

Quadro 01 - Bandas Sentinel-2A.....	32
Quadro 02 - Definição das classes de Uso.....	39
Quadro 03 - Parâmetros Random Forest no GEE.....	40
Quadro 04 - Parâmetros "seedGrid" no GEE.....	40
Quadro 05 - Parâmetros "SNIC" no GEE.....	41
Quadro 06 - Parâmetros "reduceToVectors" no GEE.....	43
Quadro 07 - Parâmetros K-means no GEE.....	44
Quadro 08 - Exemplos de correções realizadas e não realizadas na fase de pós-classificação.....	45
Quadro 09 - Área representada por cada classe na comparação entre a Classificação Supervisionada e não-supervisionada	55
Quadro 10 - Relação de Acurácia entre a Classificação Supervisionada e não-supervisionada.....	56

LISTA DE TABELAS

Tabela 01 - Coleta de Amostras.....	38
Tabela 02 - Matriz de erro da Classificação Supervisionada gerada pelo GEE.....	49
Tabela 03 - Relação de pixels corrigidos na fase de pós-classificação.....	51
Tabela 04 - Área representada por cada classe no Mapa Final de Classificação Supervisionada.....	53
Tabela 05 - Matriz de erro do mapeamento de Classificação Supervisionada na área recortada.....	56
Tabela 06 - Matriz de erro do mapeamento de Classificação não-supervisionada na área recortada.....	56

LISTA DE ABREVIações

ANA	Agência Nacional de Águas e Saneamento Básico
APAC	Agência Pernambucana de Águas e Chuvas
BDiAWeb	Banco de Dados de Informações Ambientais
GEE	Google Earth Engine
GIS	Geographic Information System
GL2	Grupo de Pequenos Rios Litorâneos 2
IBGE	Instituto Brasileiro de Geografia e Estatística
MSI	Multi Spectral Instrument
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
OSGEO	Open Source Geospatial
QGIS	Quantum GIS
RF	Random Forest
SIG	Sistema de Informação Geográfica
SNIC	Simple Non-Iterative Clustering
SR	Sensoriamento Remoto
UP	Unidade de Planejamento

SUMÁRIO

1 INTRODUÇÃO	17
2 REFERENCIAL TEÓRICO.....	19
2.1 Uso e Cobertura da Terra	19
2.1 Bacia Hidrográfica.....	20
2.3 Sistemas de Informação Geográfica e Sensoriamento Remoto	21
2.3.1 Google Earth Engine	22
2.4 Random Forest	23
2.5 Clusterização e K-means	24
2.6 Validação dos resultados	25
3 OBJETIVOS.....	27
3.1 Objetivo Geral.....	27
3.2 Objetivos Específicos	27
4 METODOLOGIA	28
4.1 Caracterização da Área de Estudo	28
4.2 Materiais.....	31
4.2.1 Google Earth Engine	31
4.2.2 Sentinel-2A.....	32
4.2.3 Python	32
4.2.4 QGIS	33
4.3 Métodos.....	33
4.3.1 Pré-processamento	34
4.3.2 Classificação Supervisionada.....	35
4.3.3 Classificação não-supervisionada.....	40
4.3.4 Validação.....	44
4.3.5 Pós-classificação	44

4.3.6 Finalização	46
5 RESULTADOS.....	48
5.1 Classificação Supervisionada.....	48
5.2 Comparativo das classificações	53
6 CONCLUSÃO	58
REFERÊNCIAS.....	59

1 INTRODUÇÃO

O conceito de Bacia Hidrográfica tem se expandido progressivamente e está sendo cada vez mais empregado como uma unidade de gestão da paisagem no contexto do planejamento ambiental (SCHIAVETTI e CAMARGO, 2002). O gerenciamento das bacias hidrográficas é crucial para estruturar e orientar o uso da terra e dos recursos naturais. Seu objetivo central é coordenar a interação entre o uso do solo, a vegetação e os recursos hídricos, visando à produção eficiente de bens e serviços.

Nesse cenário, a utilização de ferramentas de geotecnologia e sensoriamento remoto na análise e descrição de bacias hidrográficas pode ampliar significativamente a abordagem espacial do ambiente em estudo. Além disso, essas tecnologias possibilitam acelerar o processo de geração de informações, aumentando a eficiência e permitindo uma manipulação mais flexível dos dados. Elas também viabilizam atualizações em tempo real e são econômicas tanto na aquisição quanto na operação.

Um aspecto relevante é o mapeamento de uso e cobertura do solo, que vem ganhando cada vez mais relevância no que diz respeito à gestão de recursos hídricos podendo monitorar a distribuição e a dinâmica temporal das formas de ocupação da terra. Para tal, o uso de imagens de satélite de alta resolução são de extrema importância, pois capturam informações detalhadas da superfície terrestre, permitindo uma visão abrangente e precisa dos mais diversos modos de ocupação, como áreas urbanas, vegetações, corpos d'água, agriculturas e outros.

Apesar de grandes vantagens, algumas dificuldades ainda são enfrentadas, sendo muitas vezes difíceis de serem superadas. Entre essas, destacam-se o mapeamento de áreas altamente fracionadas, a diferenciação entre classes de cobertura do solo e a presença muito acentuada de nuvens em algumas regiões, obstruindo a visão da superfície e afetando a precisão das análises. Não obstante, a elaboração de mapas de cobertura do solo em regiões extensas, demanda a gestão de volumes significativos de dados geoespaciais. É imprescindível dispor de uma infraestrutura de armazenamento de grande capacidade, processamento robusto e uma abordagem flexível para uma implementação bem-sucedida. E algumas dessas complexas questões encontraram suas soluções por meio da plataforma Google Earth Engine (GEE). (XIE et al., 2019)

Gorelik (2017) define o GEE como uma plataforma poderosa que combina acesso a dados geoespaciais, recursos de processamento de alto desempenho e ferramentas para facilitar a análise e o uso de informações geográficas na escala necessária, mesmo que global. A plataforma disponibiliza um vasto catálogo de dados geoespaciais para análise e oferece uma variedade de ferramentas e algoritmos para processar esses dados, tornando o processo de análise mais eficiente.

Através de sua vasta coleção de imagens de satélite e ferramentas de processamento de dados, o GEE facilita a criação de mapas detalhados e atualizados de uso e cobertura do solo, permitindo uma compreensão precisa da paisagem. Além disso, a plataforma oferece recursos avançados de classificação por aprendizado de máquina e análise orientada a objetos, melhorando a precisão das classificações e a capacidade de detectar padrões complexos de uso da terra.

O objetivo principal desta pesquisa, portanto, é usar a plataforma do Google Earth Engine para produzir um mapa detalhado do uso e cobertura da terra da bacia hidrográfica do grupo de pequenos rios litorâneos (GL2) de Pernambuco, além de comparar métodos de classificação supervisionada e não-supervisionada.

2 REFERENCIAL TEÓRICO

2.1 Uso e Cobertura da terra

O uso da terra costumava ser considerado predominantemente uma questão ambiental local, porém, está gradativamente adquirindo uma relevância global (Foley et al., 2005). Xian e Crane (2006 apud LELIS et al., 2016) destacam que o atual padrão de uso e cobertura do solo com a expansão das terras agrícolas, pastagens, plantações e áreas urbanas em todo o mundo, acompanhada por um notável consumo de energia, água e fertilizantes, acarreta diversos desafios socioambientais, como a degradação da qualidade da água, poluição atmosférica, perda da biodiversidade, fenômenos de ilha de calor urbano, desigualdades socioeconômicas e fragmentação das áreas florestais. Essas transformações globais estão sendo impulsionadas pela necessidade de atender às demandas de quase oito bilhões de pessoas alimentos, água e habitação. (CHU, 2019)

Com essa perspectiva, a questão sobre o uso e cobertura da terra está estritamente relacionada aos conceitos de espaço geográfico, que Santos explica como o desdobramento das mudanças que o ser humano introduz na natureza ao longo do tempo. Nesse contexto, o estudo de uso e cobertura da terra torna-se essencial para compreender as mudanças no espaço, oferecendo diagnósticos e projeções de grande relevância e contribuindo para o processo de tomada de decisão na busca por soluções adequadas em uma gestão integrada. (LEITE, 2012)

Estudos para compreender a dinâmica do uso e cobertura da terra são fundamentais para mensurar os efeitos e compreender as causas do uso e ocupação desordenada dos recursos naturais, assim como permitem a tomada de decisão mais pontual e eficiente pelos gestores, combatendo o uso exploratório, a degradação ambiental, a contaminação e as perdas de recursos hídricos e de biodiversidade, assegurando o desenvolvimento mais consciente por parte das populações. (LELIS et al., 2016)

O uso e a cobertura do solo representam a informação primordial para compreender a dinâmica ambiental nas bacias hidrográficas, servindo como o ponto de partida para uma variedade de estudos. Essa informação é fundamental, sendo utilizada principalmente como base para desenvolver diretrizes, planos de ação e

propostas destinadas à ordenação, ao planejamento e à gestão ambiental das áreas em questão. (BRAZ, 2017)

2.2 Bacia Hidrográfica

Uma bacia hidrográfica é uma região natural onde a água da chuva é naturalmente coletada e direcionada para um único ponto de saída, conhecido como exutório. (TUCCI, 1997) Essa bacia é então composta por uma combinação de superfícies inclinadas formadas pelo solo e por uma rede de drenagem composta por cursos d'Corpos Hídricos que se encontram gradualmente até alcançarem um único leito nesse ponto de saída. (FINKLER, 2013)

A Lei nº 9.433, de 1997 (BRASIL, 1997), que estabelece a Política Nacional de Recursos Hídricos, designa a bacia hidrográfica como a unidade de gerenciamento territorial do Brasil e segundo Schiavetti e Camargo (2002), o conceito de bacia hidrográfica no planejamento voltado para a conservação dos recursos naturais está em processo de expansão, indo além das definições que anteriormente se limitavam aos aspectos hidrológicos. Agora, busca incorporar o entendimento da estrutura biofísica da bacia hidrográfica, bem como as transformações no uso do solo e suas implicações ambientais. E são instrumentos importantes no planejamento e gestão em função da integração entre o ser humano e o meio que o cerca.

Para Tundisi (2006), o planejamento e o gerenciamento dos recursos hídricos devem atuar de forma integrada, preditiva e em nível de ecossistemas, para incorporar as dimensões sociais e econômica de forma mais abrangentes.

De acordo com Garcez e Alvarez (2002), o comportamento hidrológico de uma bacia é significativamente influenciado por diversas características, tais como topografia, geologia, geomorfologia, solos, temperatura, e pela natureza da cobertura vegetal presente na mesma. Portanto, é de suma importância realizar medições quantitativas para avaliar essas influências de forma precisa.

Para isso, a obtenção de uma visão abrangente e em constante evolução de uma bacia, possibilitada pelo sensoriamento remoto com o auxílio de ferramentas de sistema de informação geográfica, torna-se fundamental. Isso porque essa abordagem pode destacar tanto os efeitos gerados por fenômenos naturais quanto

aqueles decorrentes da intervenção humana no uso e ocupação do solo nessa área. (FLORENZANO, 2007).

2.3 Sistemas de Informação Geográfica e Sensoriamento Remoto

Os SIG (Sistema de Informação Geográfica) são ferramentas computacionais elaboradas para o geoprocessamento, incorporando informações de diversas fontes em bancos de dados georreferenciados. (CÂMARA, 1998) Para D'alge (2003), geoprocessamento é uma área na qual se usa de técnicas computacionais e matemáticas para o tratamento de processos que ocorrem no espaço geográfico.

De acordo com Ferreira (1997), os Sistemas de Informações Geográficas (SIGs) podem ser entendidos como uma ferramenta essencial para o mapeamento e fornecimento de soluções para as diversas questões relacionadas ao planejamento urbano e regional, à agricultura, bem como para a avaliação dos recursos renováveis. Esses sistemas desempenham um papel fundamental na caracterização dos processos de mudança que ocorrem no meio ambiente e no apoio ao planejamento e gestão sustentável dos recursos naturais em áreas específicas, pois possibilita a realização de análises espaciais envolvendo dados georreferenciados geograficamente, permitindo o desenvolvimento de aplicações cartográficas em bancos de dados geográficos. (SANTOS, 2018)

O Sensoriamento Remoto (SR) é descrito por Florenzano (2007) como uma técnica para obter e aplicar informações sobre um objeto sem a necessidade de contato físico direto. Para Novo (2010), o Sensoriamento Remoto é definido como a utilização coordenada de sensores, dispositivos de processamento de dados e sistemas de transmissão de informações com o propósito de investigar eventos, fenômenos e processos que ocorrem na superfície terrestre.

As plataformas de sensoriamento remoto definem o nível de aquisição de dados, sendo eles orbital, aéreo e terrestre. Já os sistemas dividem-se em duas categorias, sendo o primeiro o de coleta ou de aquisição de dados, que depende da ocorrência de alguns fenômenos, como a fonte de radiação, a propagação da radiação pela atmosfera, a ocorrência de interações entre a radiação e os componentes da atmosfera, a incidência da radiação e os objetos da superfície e a produção de radiação que retorna o sensor, após propagar-se pela atmosfera; o segundo é o de análise e utilização dos dados, que envolve a interpretação dos produtos de sensoriamento remoto e o seu uso, incluindo os dados de campo,

participando desta fase, o pré-processamento, a conversão da forma analógica para forma digital e vice-versa. (SANTOS, 2018)

Por fim, Sensoriamento Remoto é reconhecido como uma tecnologia que viabiliza a aquisição de informações da superfície terrestre por meio da captura e registro da energia refletida ou emitida por essa superfície. Isso é alcançado por meio da detecção e quantificação das respostas resultantes das interações entre a radiação eletromagnética e os materiais presentes na Terra. (FLORENZANO, 2002; LILLESAND e KIEFER, 1997)

De acordo com Santos (2018), a combinação entre os SIG's e o Sensoriamento Remoto e sua popularização nos estudos ambientais, tem proporcionado resultados altamente satisfatórios em relação aos fenômenos do meio ambiente no que diz respeito à rapidez, quantidade e qualidade dessas informações. Através da utilização de imagens de satélite, tornou-se possível explorar uma ampla gama de escalas e aplicações, incluindo o mapeamento e a obtenção de dados sobre áreas minerais, bacias hidrográficas, agricultura, florestas, poluição de rios e reservatórios, erosão, deslizamentos de terra, secas, desmatamento, entre outros (PONZONI e SHIMABUKURO, 2007). Assim, o sensoriamento Remoto representa atualmente uma das tecnologias mais eficazes para a coleta automatizada de dados destinados ao mapeamento e monitoramento dos recursos terrestres em escala global. (MENESES e ALMEIDA, 2002)

2.3.1 Google Earth Engine

O Google Earth Engine surge como um grande expoente dessa interação entre SIG e Sensoriamento Remoto, sendo uma plataforma *big data* que fornece imagens temporais e dados vetoriais temporais globais, computação baseada em nuvem e acesso a software e algoritmo para processamento de dados. (GORELICK, 2017).

A plataforma conta com uma poderosa capacidade de processamento e armazenamento, que corresponde ao ambiente computacional da nuvem da Google. Além disso, oferece um ambiente de desenvolvimento que suporta a implementação de análises e processamento de dados geoespaciais nas linguagens JavaScript e Python, conforme destacado por Gorelick (2013).

No catálogo do Google, podemos identificar diversos dados geoespaciais de destaque, abrangendo produtos de cobertura regional e global provenientes de sensores como ASTER, MODIS, LANDSAT, NOAA, SENTINEL e outros. Além das opções já disponíveis na infraestrutura do Google, os usuários têm a possibilidade de enviar seus próprios dados para integração e processamento remoto. (GANEM, 2017)

2.4 Random Forest

A utilização de algoritmos de aprendizado de máquina tem sido fundamental para a classificação do uso e cobertura do solo, pois permitem a criação de modelos que viabilizam a automatização de um mapeamento e o Random Forest (RF) é o principal expoente dessa popularização atualmente. (PAL, 2005)

Random Forest é uma técnica de classificação criada por BREIMAN (2001) que se baseia na combinação de múltiplas árvores de decisão para resolver problemas de classificação. Como demonstra a Figura 01, cada árvore de decisão é construída a partir de uma amostra inicial aleatória dos dados, e em cada divisão desses dados, um subconjunto aleatório de m atributos é usado para a seleção dos atributos mais informativos. (VIEIRA, 2015)

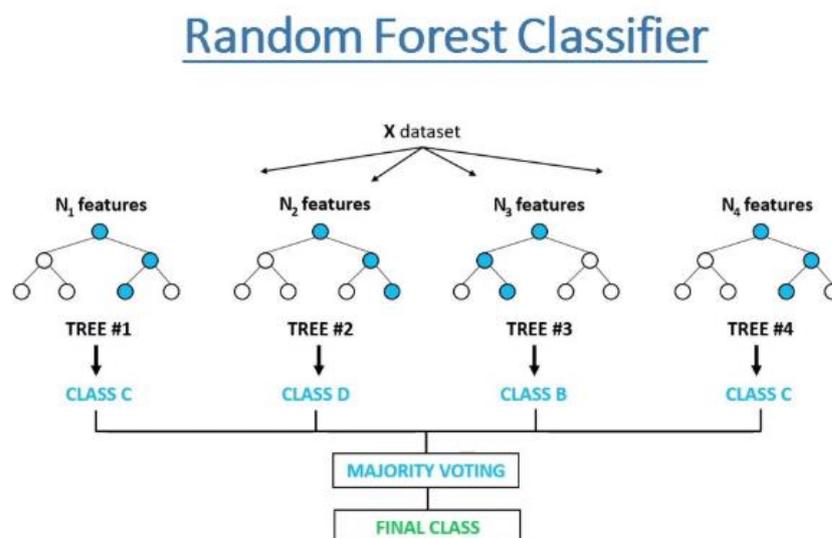


Figura 01 – Representação do algoritmo Random Forest.
Fonte: Furuya (2020)

Em termos gerais, o Random Forest busca empregar um conjunto de características que sejam representativas para a classificação de tipos específicos de uso e cobertura do solo. Essas características são derivadas das informações espectrais e temporais do produto e, em seguida, são utilizadas como dados de entrada para o algoritmo. O resultado é um classificador capaz de identificar uma ou várias classes com um determinado grau de precisão, de acordo com o modelo. Essa classificação possibilita a criação de mapas de uso e cobertura do solo, que podem ser usados como novas referências para futuras classificações. (GUEDES e GUIMARÃES, 2013)

2.5 Clusterização e K-means

A clusterização é uma técnica empregada para agrupar informações com base em seu grau de similaridade, resultando em um número específico de grupos. Ao explorar os métodos de clusterização, é possível identificar diversas técnicas para o agrupamento de informações e a principal delas é o uso do algoritmo k-means. (OLIVEIRA, 2022)

O algoritmo K-means, proposto por MacQueen em 1967, continua sendo amplamente utilizado devido à sua simplicidade, facilidade de implementação e velocidade de execução. Seu principal objetivo é particionar uma imagem em K grupos, com a finalidade de minimizar a distância dentro de cada grupo (intra-cluster) e maximizar a distância entre os grupos (inter-cluster). (ROCHA, 2021)

O método de agrupamento K-Means requer uma atribuição inicial dos vetores de medição possíveis para um número limite de grupos, especificado pelo usuário. O algoritmo, em seguida, arbitrariamente “semeia”, ou localiza, os centros dos grupos no espaço espectral, conforme é representado na Figura 02. Cada pixel da imagem é então atribuído ao grupo cujo vetor médio arbitrário está mais próximo. O procedimento continua até que não haja mudanças significativas na localização da classe dos vetores médios entre as iterações. (CAMPOS et al., 2016)

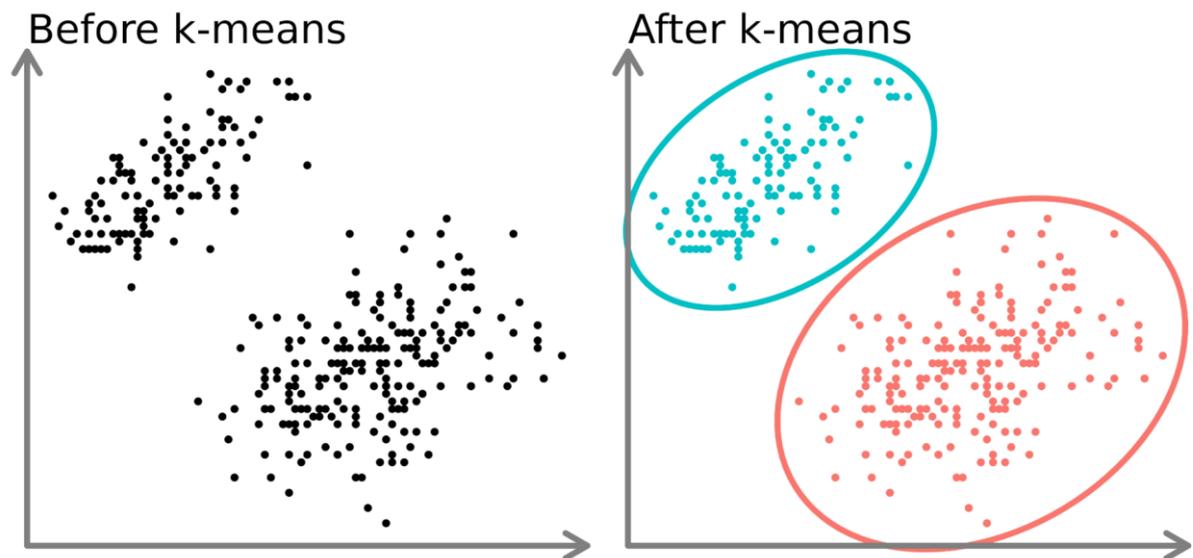


Figura 02 – Representação do algoritmo K-means.
Fonte: Anello (2023)

2.6 Validação dos resultados

A confiabilidade de qualquer mapa, como, por exemplo, um mapa de uso da terra, está diretamente relacionada à sua precisão, que indica o quão próxima uma observação ou medida está do seu valor real. Portanto, após a realização da classificação de imagens digitais, é necessário empregar algum procedimento estatístico para avaliar a acurácia ou precisão dessa classificação, independentemente de ela ter sido realizada de forma automatizada ou visual. (Bernardes, 2006)

Dentre as diversas formas de avaliar a acurácia de um classificador, grande maioria delas são relacionadas a criação de uma matriz de erro (também conhecida como matriz de confusão ou matriz de contingência (FOODY, 2020). A matriz de erro é uma ferramenta utilizada para avaliar o desempenho de uma classificação ao comparar os resultados da classificação com os dados de uma base de referência, organizando essas informações em forma de tabela, proporcionando um índice global de acerto. (SILVA, 2009)

Essa matriz é uma representação quadrada que mostra o número de unidades de amostra (pixels, grupos de pixels ou polígonos) atribuídas a categorias específicas em comparação com a categoria geral. As colunas representam os

dados de referência, enquanto as linhas indicam a classificação resultante dos dados de Sensoriamento Remoto. A diagonal principal da matriz lista as unidades de amostra que foram classificadas corretamente. (QUEIROZ, 2021)

O Índice Kappa é um produto derivado da matriz de erro que auxilia na avaliação do classificador e se caracteriza por ser uma técnica multivariada discreta usada na avaliação da precisão de mapeamentos temáticos, utilizando todos os elementos da matriz de erros em seu cálculo. (SILVA, 2009)

Os valores de Kappa variam de 0 a 1, sendo que valores próximos a 1 indicam uma melhor qualidade na classificação dos dados. Esta avaliação pode ser visualizada na Figura 03, seguindo a categorização qualitativa sugerida por Landis e Koch (1977).

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

Figura 03 – Resultado e Desempenho do Índice Kappa.
Fonte: Landis e Koch (1977)

3 OBJETIVOS

3.1 Objetivo Geral

Realizar análises de acurácia da qualidade do mapeamento dos classificadores Random Forest (supervisionado) e K-means (não supervisionado) na bacia hidrográfica do Grupo de Pequenos Rios Litorâneos 2 (GL2).

3.2 Objetivos Específicos

- Produzir o mapeamento de uso e cobertura do solo da área de estudo, empregando imagens do satélite Sentinel-2A;
- Realizar análise quantitativa da cobertura da terra na área de estudo.
- Comparar os resultados obtidos de classificação supervisionada e não-supervisionada.

4 METODOLOGIA

4.1 Caracterização da área de estudo

A Unidade de Planejamento 15 (UP 15) corresponde à bacia hidrográfica do segundo grupo de pequenos rios litorâneos (GL2) e está localizada na costa do Estado de Pernambuco, Brasil (Figura 04). Ela se estende entre as latitudes de 8°02'42" e 8°25'59" sul, bem como as longitudes de 34°52'27" e 35°52'27" oeste. De acordo com informações da Agência Pernambucana de Águas e Climas (APAC), a bacia hidrográfica GL2 abrange uma área de 1.264,94 km², o que corresponde a cerca de 1,29% do território estadual. Esta região é habitada por uma população de 1.347.053 pessoas. A área total de drenagem da bacia é de 1.022,01 km², incluindo os municípios de Cabo de Santo Agostinho, Jaboatão dos Guararapes, a sede de Moreno, além de partes dos municípios de Escada, Ipojuca, Pombos, Recife, São Lourenço da Mata e Vitória de Santo Antão. (TEIXEIRA e GALVÍNCIO, 2010)

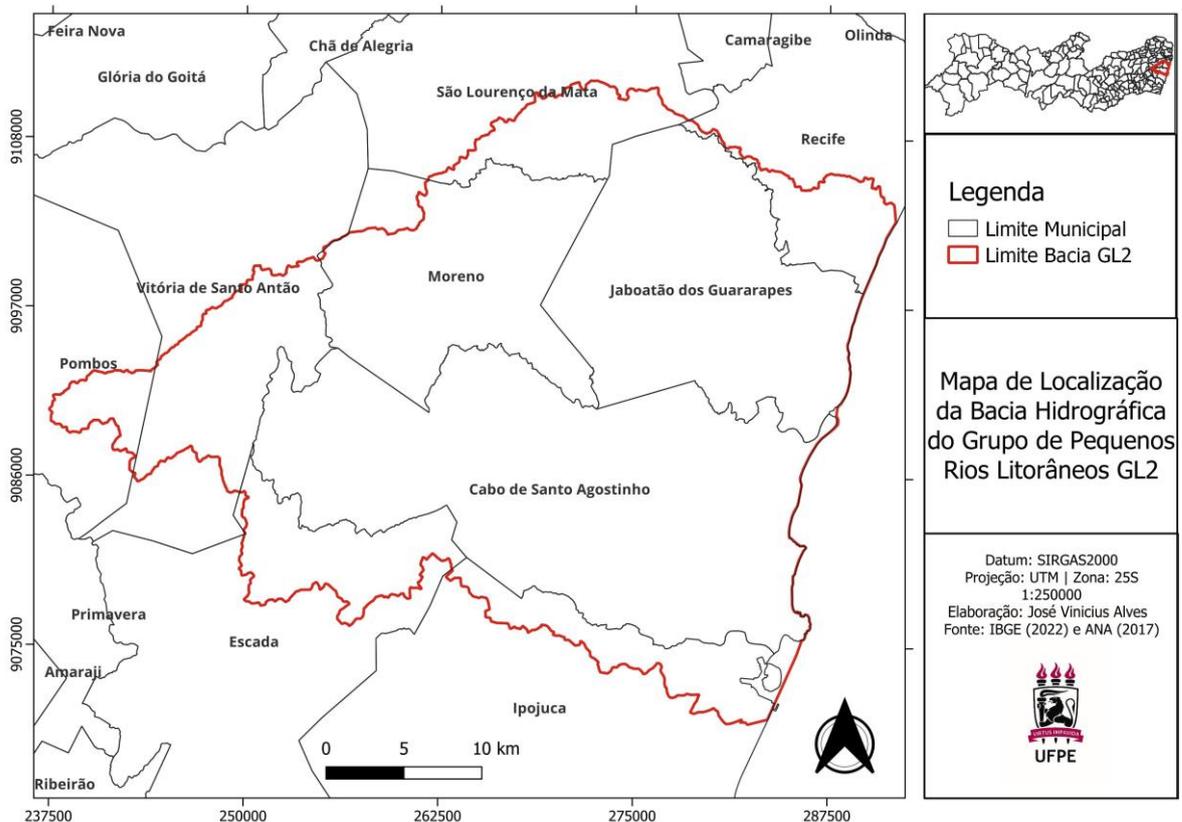


Figura 04 – Mapa de Localização da Bacia GL2.
 Fonte: Elaborado pelo autor.

Conforme os dados fornecidos por Pernambuco (1998), a região apresenta um clima quente e úmido, marcado por um período de chuvas prolongado que abrange, pelo menos, seis meses, indo de março a agosto, seguido por uma estação seca que se estende de setembro a fevereiro. No que diz respeito à distribuição das precipitações, a maior parte, aproximadamente 75 a 80%, ocorre durante os meses chuvosos, enquanto os restantes 20 a 25% estão concentrados durante o período seco.

A distribuição dos tipos de solo na região se concentra em uma faixa relativamente uniforme que predomina ao longo de todo o litoral sul de Pernambuco. Nessa faixa, encontram-se diversas classes de solos, que variam desde os predominantemente arenosos, localizados em terraços costeiros, até os solos de textura argilosa que cobrem os morros e colinas na Planície Costeira conforme a Figura 05. (COSTA e RODRIGUES, 2022)

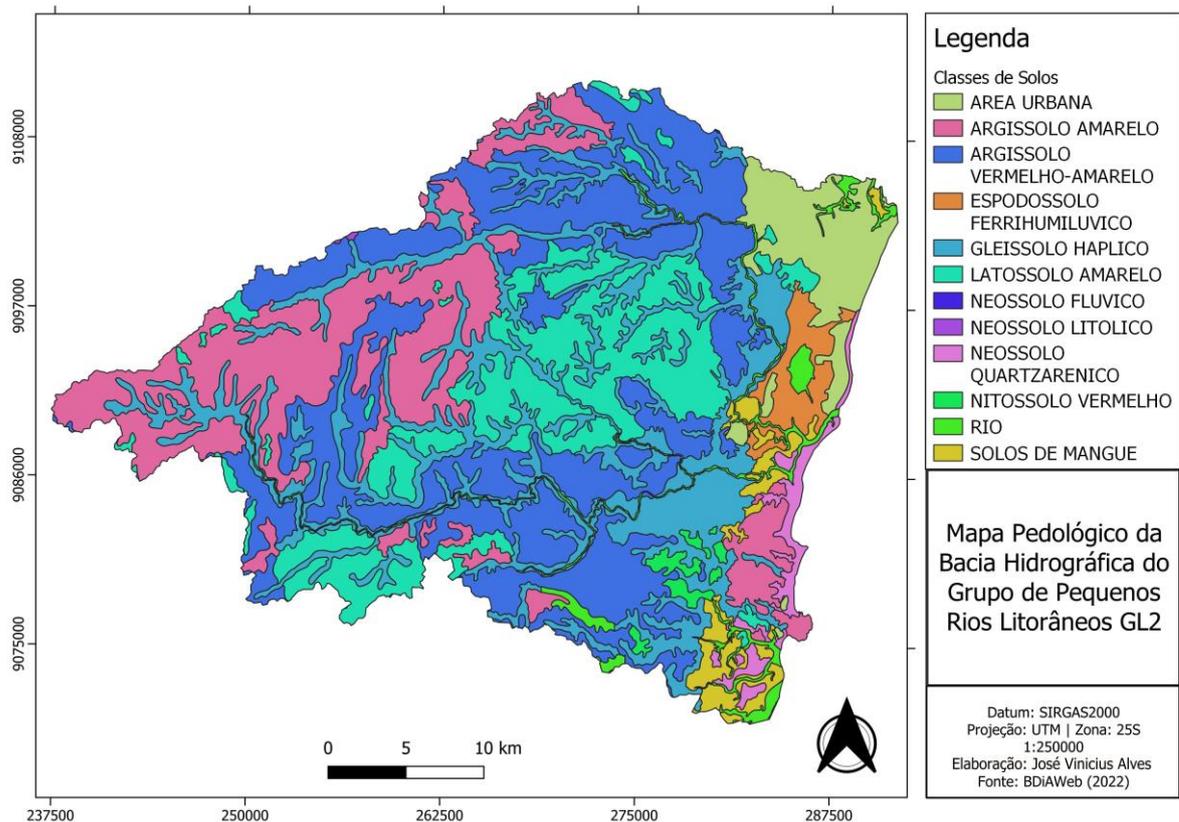


Figura 05 – Mapa Pedológico da Bacia GL2.
 Fonte: Elaborado pelo autor.

A geomorfologia da bacia hidrográfica GL2 pode ser visualizada na Figura 06, apresentando características como Piemonte e Encostas do Planalto da Borborema, além de Planícies e Terraços Fluviais e Tabuleiros. Litoral e Planícies Litorâneas completam a geomorfologia.

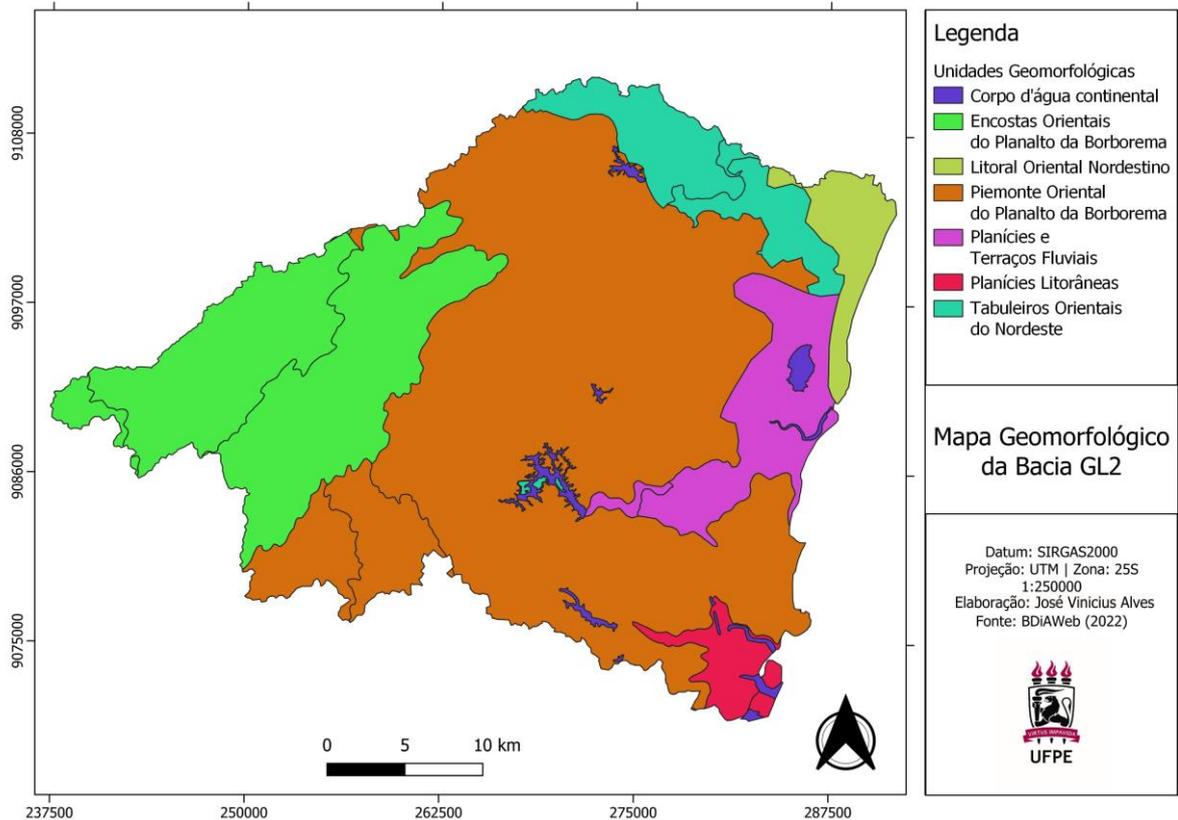


Figura 06 - Mapa Geomorfológico da Bacia GL2.
Fonte: Elaborado pelo autor.

A vegetação é tomada pela classe de Agropecuária, destacando também a influência marinha e fluviomarina. As classes de Floresta Ombrófila Aberta e Densa e Floresta Estacional Semidecidual e Formação Pioneira completam a vegetação. (Figura 07)

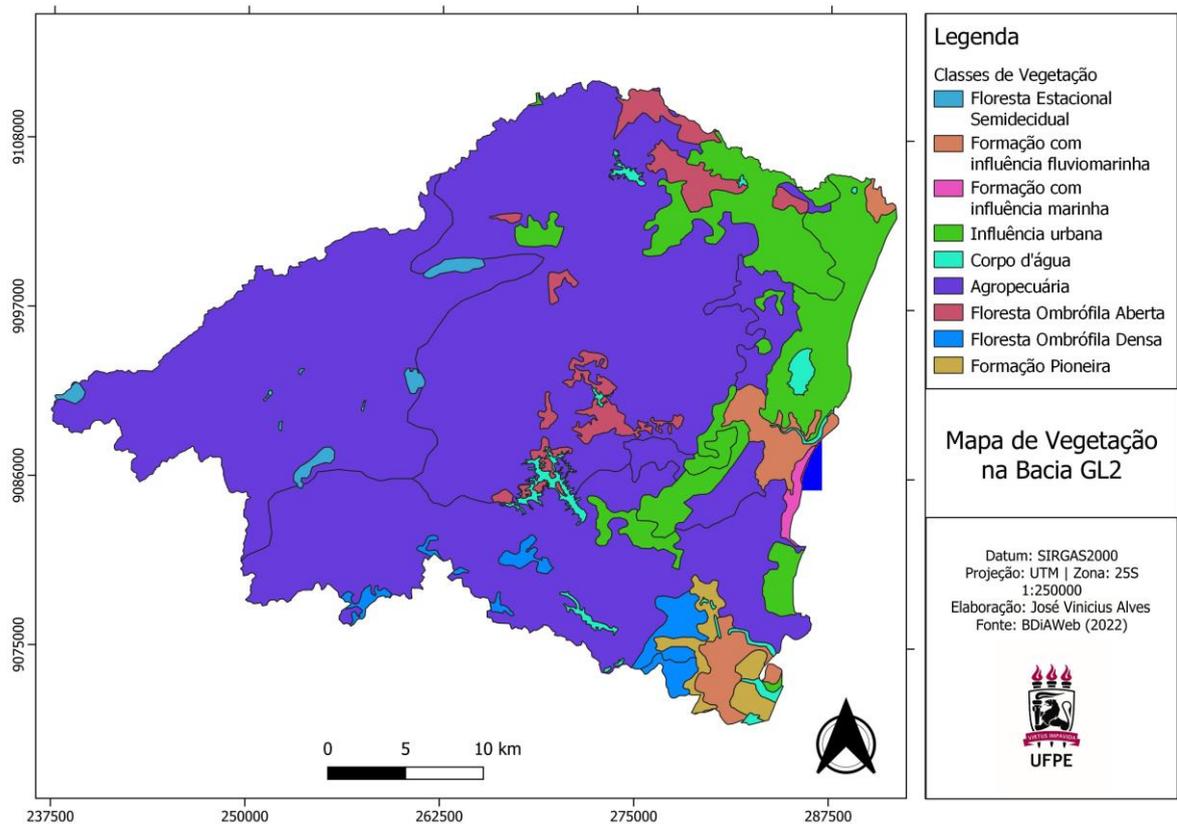


Figura 07 – Mapa de Cobertura Vegetal da Bacia GL2.
 Fonte: Elaborado pelo autor.

Para maior aprofundamento na caracterização física da bacia GL2, é recomendado os trabalhos de Teixeira e Galvêncio (2010) e Costa e Rodrigues (2022).

4.2 Materiais

4.2.1 Google Earth Engine

A plataforma Google Earth Engine foi utilizada para a criação do script, seleção das imagens após os filtros e implementação dos algoritmos de processamento e classificação. Esta plataforma oferece acesso a dados já pré-processados e simplifica tanto a abordagem supervisionada quanto a não-supervisionada.

4.2.2 Sentinel-2A

Neste estudo, foram utilizadas imagens obtidas pelo sensor Multispectral Instrument (MSI), que está a bordo dos satélites Sentinel-2A do Programa Espacial Europeu Copernicus. Esse sensor é notável pelas bandas e suas resoluções - descritas no Quadro 01 - sendo até 3 vezes mais nítido que as bandas do Landsat. (ESA, 2023).

Quadro 01 – Bandas Sentinel-2A

Bandas	Resoluções			Radiométrica (bits)
	Espacial (m)	Comprimento de Onda Central (nm)	Largura de banda (nm)	
B1 (aerosols)	60	442.7	21	12
B2 (blue)	10	492.4	66	
B3 (green)	10	599.6	36	
B4 (red)	10	664.6	31	
B5 (red-edge 1)	20	704.1	15	
B6 (red-edge 2)	20	740.5	15	
B7 (red-edge 3)	20	782.8	20	
B8 (NIR)	10	832.8	106	
B8A (red-edge 4)	20	864.7	21	
B9 (water vapor)	60	945.1	20	
B10 (cirrus)	60	1373.5	31	
B11 (SWIR 1)	20	1613.7	91	
B12 (SWIR 2)	20	2202.4	175	

Fonte: Adaptado de ESA (2023)

4.2.3 PYTHON

A linguagem de programação Python é uma escolha amplamente adotada em ciência de dados devido à sua natureza de alto nível. Com o auxílio do Python, é possível realizar análises e automatizar processos voltados para as geociências de maneira muito mais ágil e eficaz. (REY, 2023)

No Python, uma biblioteca é um conjunto de módulos e funções pré-escritos que podem ser importados em seu programa para fornecer funcionalidades específicas. As bibliotecas são essenciais para a linguagem Python, pois expandem

suas capacidades, permitindo que você realize uma ampla variedade de tarefas sem a necessidade de escrever o código do absoluto zero. (TUAMA, 2020)

Neste trabalho, foram importadas as bibliotecas numpy e rasterio, que são especialmente adequadas para lidar com ciência de dados e, especificamente, dados geoespaciais, respectivamente.

4.2.4 QGIS

O SIG utilizado depois da exportação das imagens classificadas e pós classificadas, para análises finais e revisão do mapeamento foi o QGIS, projetado pela Open Source Geospatial Foundation (OSGeo). QGIS é uma das ferramentas mais utilizadas para análises espaciais, principalmente por ser um software livre, com seu código fonte aberto.

4.3 Métodos

Os processos executados neste trabalho para realizar o mapeamento de uso e cobertura da terra estão esquematicamente representados na Figura 08. As etapas apresentadas no fluxograma são descritas nos tópicos subsequentes.

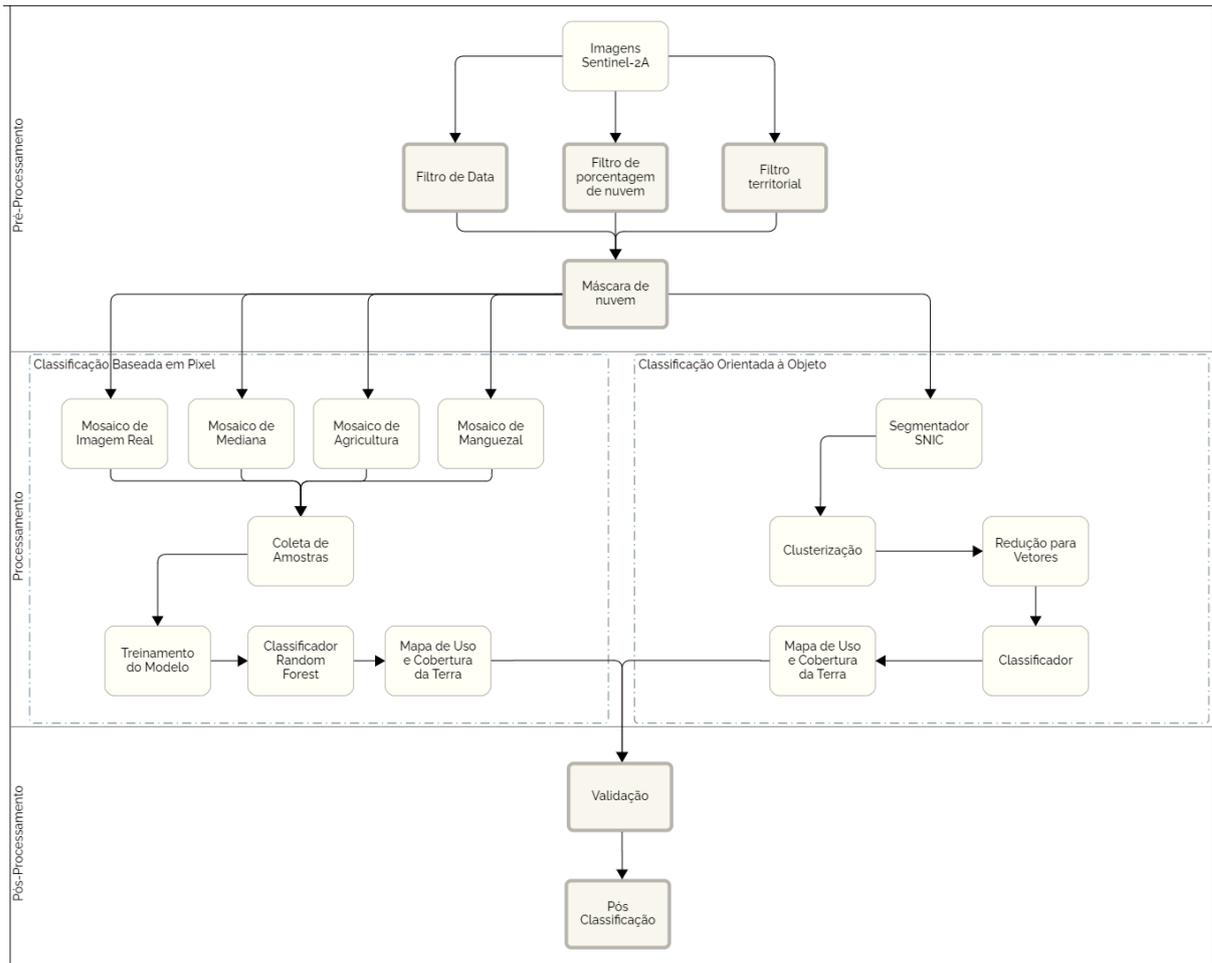


Figura 08 – Fluxograma Metodológico
 Fonte: Elabora pelo autor.

4.3.1 Pré-processamento

No pré-processamento das imagens do Sentinel, foram aplicados diversos filtros para garantir a qualidade e relevância dos dados. Inicialmente, foi estabelecido um filtro temporal, abrangendo o período de 01 de janeiro de 2021 até 31 de dezembro de 2021, com o objetivo de realizar uma análise anual específica para o ano de 2021.

Além disso, foi empregado um filtro de cobertura de nuvens, selecionando apenas as imagens em que a porcentagem de cobertura de nuvens fosse inferior a 30%. Esse critério é essencial para garantir que as informações contidas nas imagens não estejam obstruídas por nuvens, o que poderia prejudicar a análise.

Por último, foi aplicado um filtro territorial, que se baseou na importação do shapefile da bacia fornecido pela Agência Nacional de Águas (ANA). Esse filtro

limitou a seleção das imagens ao território específico da bacia, garantindo que apenas os dados relevantes para a área de estudo fossem considerados no processo de análise. Esses filtros foram implementados com o intuito de assegurar que as imagens Sentinel utilizadas no estudo fossem adequadas, tanto em termos temporais quanto espaciais, para a análise pretendida.

Após a aplicação dos procedimentos mencionados, foi utilizada uma máscara de nuvens automática disponibilizada por Principe (2019) em seu repositório no GitHub. Além disso, foi incorporado o cálculo do Índice de Vegetação por Diferença Normalizada (NDVI), conforme a equação abaixo:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Essa etapa permitiu a quantificação da vegetação presente nas imagens, usando as bandas de infravermelho próximo (NIR) e vermelho (Red), auxiliando assim na análise das características e saúde da cobertura vegetal na área de estudo.

4.3.2 Classificação Supervisionada

O primeiro passo na classificação supervisionada envolveu a criação de quatro mosaicos para o ano de 2021 utilizando a plataforma Google Earth Engine. Cada um desses mosaicos desempenhou um papel crucial na coleta das amostras necessárias para o processo de mapeamento.

O mosaico True Color ou mosaico RGB foi desenvolvido para auxiliar na coleta das amostras urbanas e de solo exposto, conforme a Figura 09. A combinação de bandas é B4, B3 e B2.



Figura 09 – Mosaico “True Color”
Fonte: Elaborado pelo autor

O mosaico False Color foi concebido para identificar com mais precisão os corpos d'água presentes na área, conforme a Figura 10. A combinação de bandas é B8, B4 e B3.

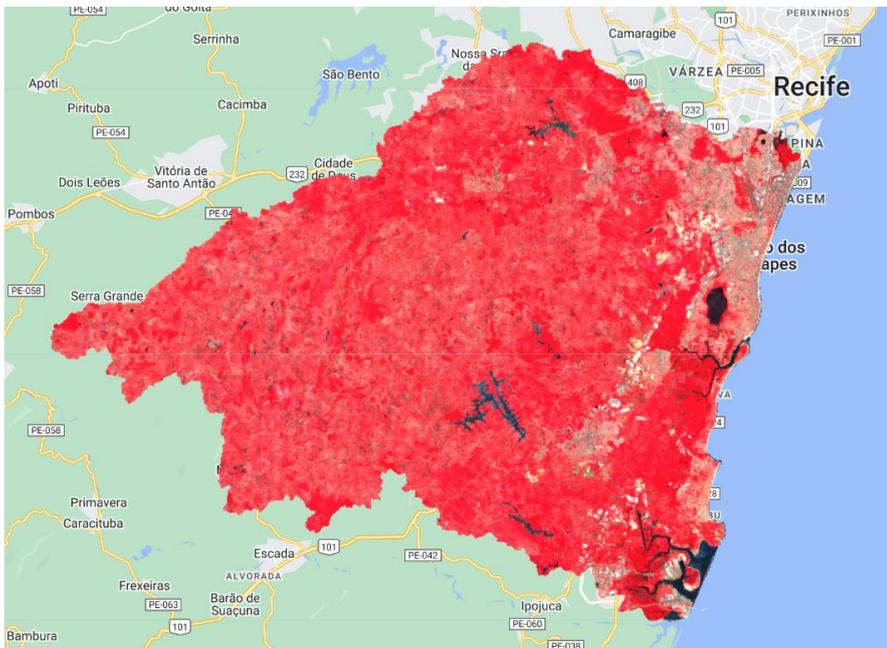


Figura 10 – Mosaico “False Color”
Fonte: Elaborado pelo autor

Todos os mosaicos seguiram as etapas de pré-processamento, com a única diferença sendo as bandas selecionadas para a visualização em cada caso. Isso permitiu a análise de diferentes informações e aspectos das imagens em cada mosaico, ampliando a compreensão da região estudada.

Para o processo de treinamento do modelo, foram coletadas amostras por meio de pontos que representavam as categorias de informações utilizadas para mapear a região. Essas categorias foram classificadas da seguinte forma: Corpos Hídricos (1), Solo Exposto (2), Superfícies Urbanas (3), Pastagem (4), Mangue (5), Formação Florestal (6), Cana de açúcar (7) e Mosaico de Usos (8), conforme apresentado na Tabela 01. Cada ponto de amostra foi cuidadosamente escolhido de modo a garantir que a categoria desejada pudesse ser observada de forma clara e distintiva, a fim de fornecer ao modelo de mapeamento informações precisas e abrangentes sobre a região em questão.

Tabela 01 – Coleta de Amostras

Classe	Treinamento	Validação	Total
1	601	258	859
2	865	371	1235
3	3051	1307	4358
4	809	347	1156
5	2007	860	2867
6	2166	928	3094
7	3595	1541	5136
8	2675	1147	3822

Fonte: Elaborado pelo autor.

Para definir cada categoria, foram adotadas as definições estabelecidas pela coleção 8 do Mapbiomas, como disposto no Quadro 02.

Quadro 02 – Definição das Classes de Uso

Nº da Classe	Nome da Classe	Descrição da Classe
1	Corpos Hídricos	Rios, lagos, represas, reservatórios e outros corpos d'água.
2	Solo Exposto	Áreas onde o solo natural está visível e não coberto por vegetação, corpos d'água, construções ou outros tipos de cobertura.
3	Superfícies Urbanas	Áreas com significativa densidade de edificações e vias, incluindo áreas livres de construções e infraestrutura.
4	Pastagem	Áreas de pastagem predominantemente plantadas, diretamente relacionadas à atividade agropecuária.
5	Mangue	Formações florestais, densas, sempre-verdes, frequentemente inundadas pela maré e associadas ao ecossistema costeiro de Manguezal
6	Formação Florestal	Floresta Ombrófila Densa, Aberta e Mista e Floresta Estacional Semi-Decidual, Floresta Estacional Decidual e Formação Pioneira Arbórea.
7	Cana-de-Açúcar	Áreas cultivadas com a cultura da cana-de-açúcar.
8	Mosaico de Usos	Áreas de uso agropecuário onde não foi possível distinguir entre pastagem e agricultura.

Fonte: Adaptado de MAPBIOMAS (2022)

O conjunto total de amostras coletadas foi separado em dois conjuntos de dados distintos e aleatórios: um conjunto de treinamento, que compreendia 70% das amostras, e um conjunto de validação, que abrangia os 30% restantes.

Utilizando o classificador Random Forest, foram empregadas 100 árvores de decisão no processo de ramificação, utilizando as propriedades das seguintes bandas: B2, B3, B4, B8, B11, B12 e NDVI. Para o treinamento, foi utilizada uma porção aleatória equivalente a 70% das amostras previamente separadas. Os demais parâmetros permaneceram em suas configurações padrão, conforme indicado no Quadro 03, e, com essas configurações, o mapa de classificação foi gerado.

Quadro 03 – Parâmetros Random Forest no GEE

Parâmetro	Tipo	Função
numberOfTrees	Inteiro	O número de árvores de decisão a serem criadas.
variablesPerSplit	Inteiro, padrão: <i>null</i>	O número de variáveis por divisão. Se não for especificado, utiliza a raiz quadrada do número de variáveis.
minLeafPopulation	Inteiro, padrão: 1	Crie apenas nós cujo conjunto de treinamento contenha pelo menos esta quantidade de pontos.
bagFraction	Decimal, padrão: 0.5	A fração do insumo para o saco por árvore.
maxNodes	Inteiro, padrão: <i>null</i>	O número máximo de nós folha em cada árvore. Se não for especificado, o padrão é sem limite.
seed	Inteiro, padrão: 0	A semente da randomização.

Fonte: Adaptado de GOOGLE EARTH ENGINE (2023)

4.3.3 Classificação não-supervisionada

Na abordagem de classificação não-supervisionada, após a conclusão do pré-processamento das imagens, elas foram segmentadas utilizando o algoritmo SNIC. Primeiramente, o valor das sementes (seeds) foi gerado para determinar o tamanho dos clusters, empregando o algoritmo seedGrid. Nesse processo, foi utilizado o valor 5 como referência, enquanto os demais parâmetros mantiveram suas configurações padrão, conforme especificados no Quadro 04.

Quadro 04 – Parâmetros “seedGrid” no GEE

Parâmetro	Tipo	Função
size	Inteiro, padrão: 5	O espaçamento de localização inicial de superpixel, em pixels.
gridType	String, padrão: "square"	Tipo de grade. Pode ser 'square' ou 'hex'.

Fonte: Adaptado de GOOGLE EARTH ENGINE (2023)

Após a definição dos valores das sementes, o algoritmo SNIC procedeu com a segmentação e a formação de clusters na imagem, preparando-a para posterior

redução para o formato de vetores. Os demais parâmetros permaneceram inalterados, seguindo as configurações padrão descritas no Quadro 05.

Quadro 05 – Parâmetros “SNIC” no GEE

Parâmetro	Tipo	Função
image	Imagem	A imagem de entrada para clustering.
size	Inteiro, padrão: 5	O espaçamento de localização inicial de superpixel, em pixels. Se a imagem 'sementes' for fornecida, nenhuma grade será produzida.
compactness	Decimal, padrão: 1	Fator de compactação. Valores maiores fazem com que os clusters sejam mais compactos (quadrados). Definir isso como 0 desativa a ponderação de distância espacial.
connectivity	Inteiro, padrão: 8	Conectividade. Ou 4 ou 8.
neighborhoodSize	Inteiro, padrão: <i>null</i>	Tamanho da vizinhança do bloco (para evitar artefatos nos limites do bloco). O padrão é tamanho 2 * tamanho.
seeds	Imagem, padrão: <i>null</i>	Se fornecido, quaisquer pixels com valor diferente de zero serão usados como locais iniciais. Os pixels que se tocam (conforme especificado por 'conectividade') são considerados pertencentes ao mesmo cluster.

Fonte: Adaptado de GOOGLE EARTH ENGINE (2023)

Dessa maneira, tornou-se possível realizar uma análise visual dos clusters gerados pelo algoritmo SNIC, como ilustrado na Figura 13.

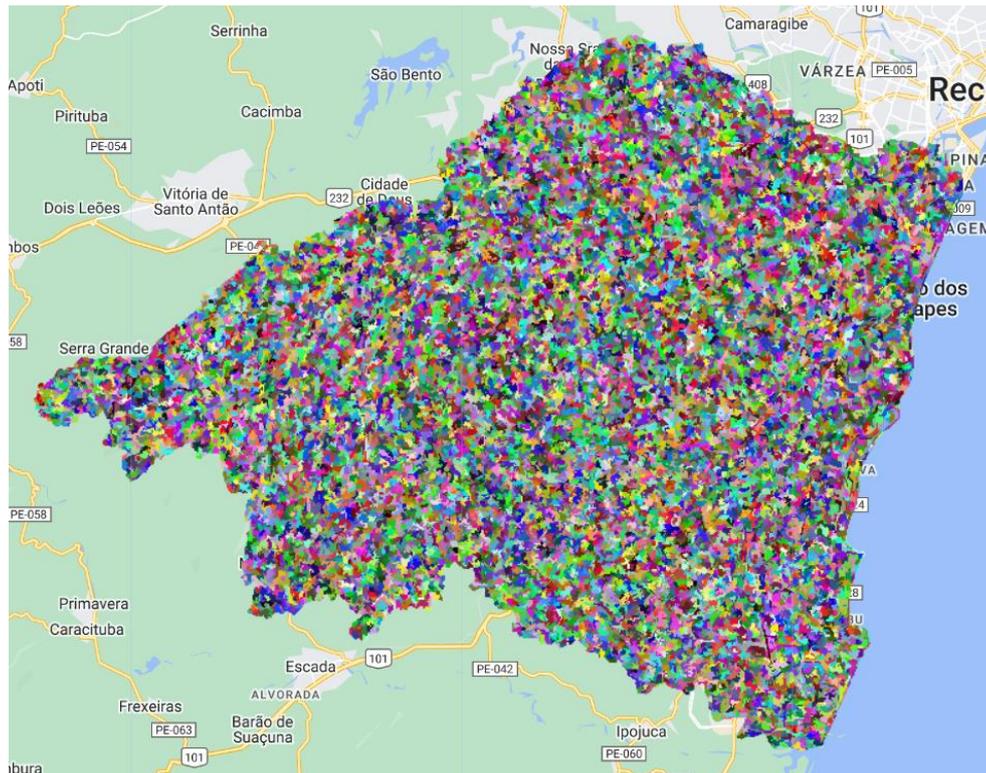


Figura 13 – Mosaico de Clusters
Fonte: Elaborado pelo autor.

Para realizar a conversão da imagem segmentada criada pelo algoritmo SNIC em vetores, empregou-se o algoritmo "reduceToVectors". Este algoritmo efetua a redução da imagem, transformando-a em regiões homogêneas, com base na clusterização previamente realizada. Para calcular as características das regiões, utilizou-se o método "reducer.mean", que calcula a média ponderada das bandas da imagem, visando facilitar a classificação subsequente. Todos os demais parâmetros foram mantidos com as configurações padrão, conforme indicado no Quadro 06.

Quadro 06 – Parâmetros “reduceToVectors” no GEE

Parâmetro	Tipo	Função
image	Imagem	A imagem de entrada. Espera-se que a primeira banda seja do tipo inteiro; pixels adjacentes estarão no mesmo segmento se tiverem o mesmo valor nesta banda.
reducer	Redutor, padrão: <i>null</i>	O redutor a ser aplicado. Suas entradas serão obtidas das bandas da imagem após eliminar a primeira banda. O padrão é Redutor.countEvery()
geometry	Geometria, padrão: <i>null</i>	A região sobre a qual reduzir os dados. O padrão é a área ocupada pela primeira banda da imagem.
scale	Decimal, padrão: <i>null</i>	Uma escala nominal em metros da projeção para trabalhar.
geometryType	String, padrão: "polygon"	Como escolher a geometria de cada feature gerada; um de 'polígono' (um polígono que envolve os pixels no segmento), 'bb' (um retângulo que delimita os pixels) ou 'centróide' (o centróide dos pixels).
eightConnected	Boolean, padrão: <i>true</i>	Se for verdade, os pixels conectados diagonalmente são considerados adjacentes; caso contrário, apenas os pixels que compartilham uma borda serão.
maxPixels	Longo, padrão: 10000000	O número máximo de pixels a serem reduzidos.
tileScale	Decimal, padrão: 1	Um fator de escala usado para reduzir o tamanho do bloco de agregação; usar um tileScale maior (por exemplo, 2 ou 4) pode permitir cálculos que ficam sem memória com o padrão.

Fonte: Adaptado de GOOGLE EARTH ENGINE (2023)

No processo de classificação, optou-se por empregar o algoritmo K-means. Com base no número predefinido de classes, o K-means agrupou as regiões homogêneas para produzir a classificação desejada. Para esse propósito, foram escolhidas as mesmas bandas utilizadas na classificação por Random Forest, que incluem B2, B3, B4, B8, B11, B12 e NDVI. Todos os parâmetros do algoritmo foram mantidos em suas configurações padrão, conforme apresentado no Quadro 07.

Quadro 07 – Parâmetros K-means no GEE

Parâmetro	Tipo	Função
image	Imagem	A imagem de entrada para clustering.
numClusters	Inteiro, padrão: 8	Número de clusters.
numIterations	Inteiro, padrão: 20	Número de iterações
neighborhoodSize	Inteiro, padrão: 0	Tamanho do bairro. A quantidade de extensão de cada bloco (sobreposição) ao calcular os clusters. Esta opção é mutuamente exclusiva com gridSize.
gridSize	Inteiro, padrão: <i>null</i>	Tamanho da célula da grade. Se for maior que 0, kMeans será executado independentemente em células deste tamanho. Isso tem o efeito de limitar o tamanho de qualquer cluster para ser gridSize ou menor. Esta opção é mutuamente exclusiva com bairroSize.
forceConvergence	Boleano, padrão: <i>false</i>	Se verdadeiro, um erro será gerado se a convergência não for alcançada antes de numIterations.
uniqueLabels	Boleano, padrão: <i>true</i>	Se for verdade, os clusters recebem IDs exclusivos. Caso contrário, eles se repetem por bloco ou célula da grade.

Fonte: Adaptado de GOOGLE EARTH ENGINE (2023)

4.3.4 Validação

Após a geração dos mapas por meio da classificação supervisionada e da classificação não-supervisionada, os 30% das amostras restantes, selecionadas aleatoriamente, foram reservadas para conduzir várias avaliações. Essas avaliações incluíram a construção da Matriz de Erro, bem como o cálculo da Acurácia Global e do Índice Kappa.

A Matriz de Erro desempenhou um papel fundamental como o principal método de validação desta pesquisa. A partir dela, foram derivados os índices estatísticos necessários para todas as outras análises presentes neste trabalho.

4.3.5 Pós-classificação

De acordo com Inpe (2006), é viável conduzir uma etapa de pós-classificação no mapeamento de imagens de alta resolução. O propósito dessa etapa é

harmonizar os temas, ou seja, remover pontos isolados que tenham sido classificados de maneira diferente de sua vizinhança. Assim, um script em Python foi criado para reclassificar pixels isolados de uma classe para outra, com o intuito de gerar uma imagem classificada com aparência menos ruidosa.

Por meio do algoritmo, cada pixel da imagem foi tratado individualmente e isolado em uma matriz 3x3 dedicada. Dentro dessa matriz, foi efetuada uma verificação para determinar se todos os pixels vizinhos a ele pertenciam à mesma classe. O Quadro 08 apresenta exemplos que ilustram o resultado do algoritmo em um caso em que a correção foi realizada e em outro caso em que não foi.

Quadro 08 – Exemplos de correções realizadas e não realizadas na fase de pós-classificação.

<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>7</td><td>7</td><td>7</td></tr> <tr><td>7</td><td>8</td><td>7</td></tr> <tr><td>7</td><td>7</td><td>7</td></tr> </table>	7	7	7	7	8	7	7	7	7	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>7</td><td>7</td><td>8</td></tr> <tr><td>7</td><td>4</td><td>8</td></tr> <tr><td>8</td><td>7</td><td>8</td></tr> </table>	7	7	8	7	4	8	8	7	8	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>7</td><td>7</td><td>4</td></tr> <tr><td>4</td><td>4</td><td>8</td></tr> <tr><td>8</td><td>7</td><td>8</td></tr> </table>	7	7	4	4	4	8	8	7	8
7	7	7																											
7	8	7																											
7	7	7																											
7	7	8																											
7	4	8																											
8	7	8																											
7	7	4																											
4	4	8																											
8	7	8																											
A	B	C																											

Fonte: Elaborado pelo autor.

No exemplo A, um pixel da classe Mosaico de Usos (8) está isolado no centro de sua matriz 3x3 e está cercado em todas as direções apenas por pixels da classe Cana-de-açúcar (7). Neste cenário, o pixel isolado da classe Mosaico de Usos (8) seria corrigido para a classe Cana-de-açúcar (7).

Já no exemplo B, um pixel da classe Pastagem (4) está isolado no centro de sua matriz 3x3, mas dentro dessa matriz, há pixels tanto da classe Mosaico de Usos (8) quanto da classe Cana-de-açúcar (7). Nesta situação, o pixel isolado da classe Pastagem (4) permaneceria com o seu valor original.

E no exemplo C, um pixel da classe Pastagem (4) não está isolado no centro de sua matriz 3x3, pois outro pixel da classe Pastagem (4) também está presente junto a ele dentro da matriz. Em circunstâncias assim, o pixel central permaneceria com o seu valor original.

4.3.6 Finalização

Como mencionado anteriormente, o Google Earth Engine é uma plataforma em nuvem altamente poderosa, equipada com uma ampla gama de ferramentas e recursos para o processamento, análise e visualização de dados geoespaciais em larga escala. No entanto, apesar de sua notável capacidade de processamento, ainda existem algumas limitações práticas a serem consideradas.

Durante o desenvolvimento desta pesquisa, deparou-se com o erro "Computed value is too large". Esse erro costuma ocorrer quando uma operação ou consulta específica ultrapassa os recursos disponíveis na plataforma, como capacidade de memória ou poder de processamento. Isso pode acontecer quando lidamos com volumes significativos de dados geoespaciais, como imagens de satélite de altíssima resolução, grandes conjuntos de geometrias ou operações complexas que envolvem cálculos extensos ou agregações em grande escala. (GOOGLE EARTH ENGINE, 2023)

Para resolver esse problema, a plataforma sugere algumas abordagens, como reduzir a área de interesse, diminuir a resolução das imagens usando o parâmetro "scale" ou adicionar parâmetros que auxiliem na divisão do processamento.

Ao longo do processo de redução da imagem em vetores, um estágio crítico para a classificação não-supervisionada proposta neste estudo, o erro mencionado anteriormente surgiu. Para lidar com essa questão, foram aplicadas algumas soluções. Foi necessário ajustar os parâmetros 'maxPixels' e 'tileScale' para aumentar a quantidade de pixels que poderiam ser processados e dividir a área total em blocos menores, a fim de facilitar o processamento paralelo. Apesar de definir esses valores no limite máximo permitido, ainda assim, não foi possível realizar a classificação em toda a bacia com sucesso usando o segmentador SNIC e o classificador K-means.

Portanto, a solução viável encontrada para efetuar a comparação entre as duas classificações foi limitar a área de interesse de forma a se adequar às restrições da plataforma.

Desta forma, esta seção será dividida em duas partes para análise: a primeira abordará a classificação abrangendo toda a bacia utilizando o classificador Random Forest, juntamente com sua validação; a segunda parte será dedicada à

comparação entre as classificações supervisionada e não-supervisionada, em uma área de interesse reduzida em ambos os casos, e suas respectivas validações.

O principal critério para a seleção da área reduzida foi a identificação daquela região onde se apresentavam um maior número de classes a serem mapeadas. Isso se deve ao fato de que não foi possível encontrar uma área reduzida que englobasse todas as classes disponíveis na pesquisa.

5 RESULTADOS

5.1 Classificação Supervisionada

A imagem apresentada na Figura 14 exibe o mapa de classificação obtido por meio do algoritmo Random Forest, sem a inclusão da fase de pós-classificação.

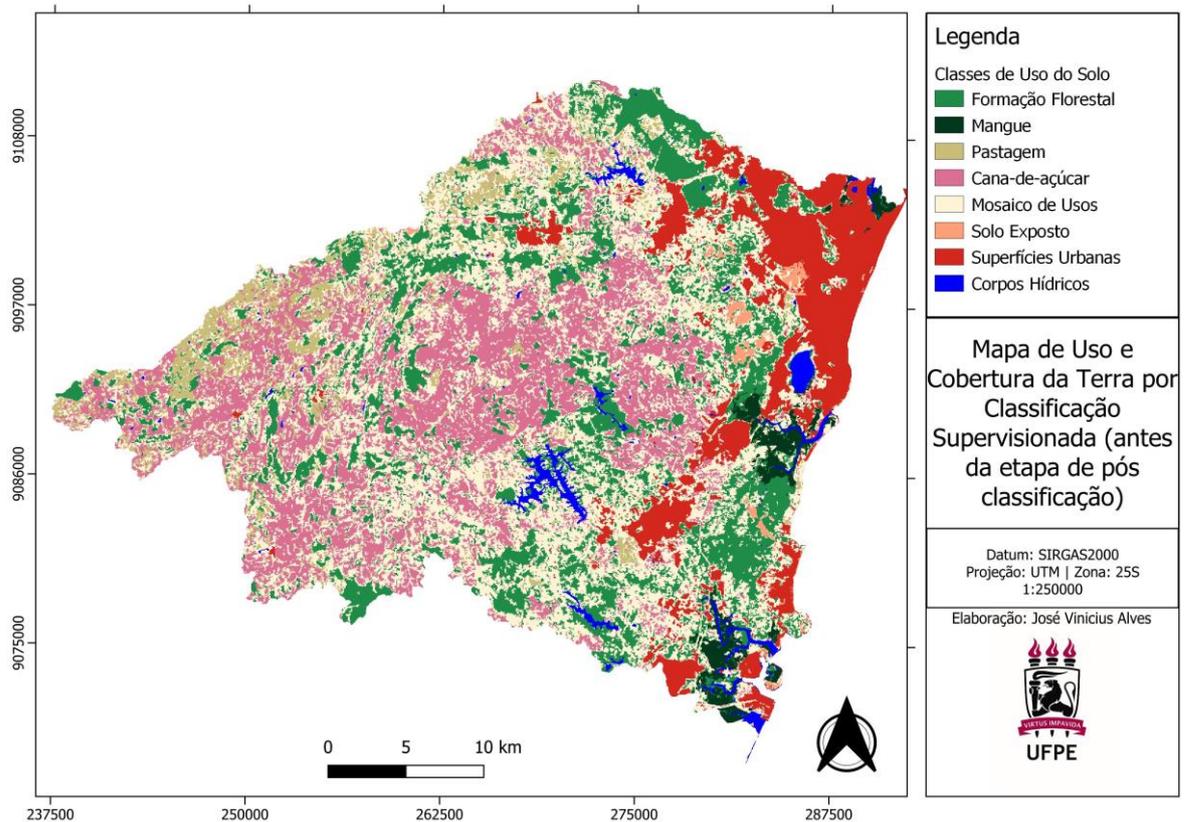


Figura 14 – Mapa gerado pelo GEE de Uso e Cobertura da Terra por Classificação Supervisionada. Fonte: Elaborado pelo autor

A Tabela 02 mostra a matriz de erro gerada a partir da classificação feita anteriormente.

Tabela 02 – Matriz de erro da Classificação Supervisionada gerada pelo GEE.

	1	2	3	4	5	6	7	8	
1	256	0	0	0	2	0	0	0	258
2	0	367	4	0	0	0	0	0	371
3	0	3	1304	0	0	0	0	0	1307
4	0	0	0	336	0	0	10	1	347
5	1	0	0	0	857	2	0	0	860
6	0	0	0	0	7	921	0	0	928
7	0	0	0	0	0	0	1527	14	1541
8	0	0	0	0	0	0	23	1124	1147

Fonte: Elaborado pelo autor.

Analisando a matriz de erro, é possível inferir que, das 6759 amostras de validação, 66 foram classificadas erroneamente, resultando em um total de 0,97% das amostras. A maior confusão ocorreu entre as classes Pastagem (4), Cana-de-açúcar (7) e Mosaico de Usos (8). Na classe Pastagem (4), 2,88% foram classificados como Cana-de-açúcar (7) e 0,28% foram classificados como Mosaico de Usos (8), totalizando uma porcentagem de erro para essa classe de 3,16%. Na classe Mosaico de Usos (8), 2% foram classificados como Cana-de-açúcar (7), que, por sua vez, teve 0,90% classificados como a classe anterior (8).

Outros erros incluíram: na classe Corpos Hídricos (1), 0,77% foram classificados como Mangue (5); na classe Solo Exposto (2), 1,07% foram classificados como Urbano (3); na classe Urbano (3), 0,22% foram classificados como Solo Exposto (2); na classe Mangue (5), ocorreram erros de 0,23% como Formação Florestal (6) e 0,11% como Corpos Hídricos (1), totalizando uma porcentagem de erro para essa classe de 0,34%; e na classe Formação Florestal (6), 0,75% foram classificados como Mangue (5).

Todas as confusões detectadas no desempenho do algoritmo eram previsíveis devido à semelhança entre as classes em relação às suas características espectrais e à base de amostras coletadas. Por exemplo, as classes de pastagens e áreas agrícolas apresentam características espectrais semelhantes ao longo de grande parte do ano, o que torna desafiada a tarefa de distingui-las de forma precisa. (ACCIOLY, 2016)

Após a análise realizada com a matriz de erro, foram calculados os seguintes valores de métricas de avaliação: Acurácia Global e Índice Kappa, resultando em:

- Acurácia Global: 99,12%
- Índice Kappa: 0,98

O valor da Acurácia Global de 99,12% representa um desempenho excepcional no processo de classificação, superando significativamente o limiar mínimo aceitável de 85% estabelecido por Anderson (1971). Isso significa que o algoritmo foi capaz de classificar com precisão a grande maioria dos pixels na imagem, refletindo uma alta concordância entre as classes atribuídas pelo modelo e as classes reais presentes no terreno.

Além disso, o Índice Kappa atingiu um valor de 0,98, o que também indica um desempenho notável. O Índice Kappa é uma métrica que leva em consideração o acordo entre as classificações do modelo e as observações reais, levando em conta a possibilidade de acordos aleatórios. Um valor de 0,9 sugere um alto grau de concordância que vai além do que seria esperado ao acaso, corroborando a qualidade do modelo.

Após a validação do mapa gerado, foi conduzido um processo de refinamento com o auxílio do script em Python, resultando na criação de um mapa corrigido, e incluindo a geração de informações relevantes para a posterior análise.

A imagem original continha um total de 25.485.822 pixels, dos quais 9.585 pixels foram identificados e corrigidos como isolados, de acordo com a metodologia aplicada. Esses pixels corrigidos representam uma porcentagem de 0,04% do total de pixels na imagem. A Tabela 03 fornece informações detalhadas sobre a quantidade de pixels corrigidos para cada classe e a classe original a que eles pertenciam. Horizontalmente, é possível ver quantos pixels cada classe perdeu e para qual classe eles foram destinados. Já verticalmente, quantos pixels cada classe recebeu e qual era a classe originária. Essa análise permite compreender as mudanças nas classes entre o mapa gerado inicialmente pelo GEE e o mapa resultante após a pós-classificação.

Tabela 03 – Relação de Pixels corrigidos na fase de pós-classificação.

	1	2	3	4	5	6	7	8	Total
1	-	0	23	0	187	0	0	0	210
2	0	-	329	164	0	0	0	0	493
3	37	854	-	21	9	245	47	34	1247
4	0	78	142	-	0	13	303	387	923
5	208	0	0	0	-	270	0	0	478
6	0	0	114	54	146	-	74	31	419
7	0	0	411	954	0	185	-	843	2393
8	0	0	116	1268	0	354	1684	-	3422
Total	245	932	1135	2461	342	1067	2108	1295	9585

Fonte: Elaborado pelo autor.

A Tabela 03 revela um padrão notavelmente semelhante à tabela da matriz de erro, com as mesmas classes frequentemente envolvidas em correções. Em geral, as classes mais corrigidas foram Mosaico de Usos (8), Cana-de-Açúcar (7) e Urbano (3). É interessante notar que essas três classes também foram aquelas com o maior número de amostras coletadas inicialmente. Por outro lado, a classe Corpos Hídricos (1) se destaca como a menos afetada pelas correções, o que está alinhado com o fato de ser a classe com menos amostras coletadas inicialmente.

Em relação à classe Mosaico de Usos (8), um total de 3.422 pixels isolados foram identificados, representando uma porcentagem de % de toda a classe. Os maiores ajustes para essa classe envolveram a classe Cana-de-açúcar (7), com 1.684 pixels corrigidos. A classe Cana-de-açúcar (7), por sua vez, teve 2.393 pixels corrigidos, o que equivale a % de toda a classe, sendo a classe Pastagem (4) a principal beneficiária dessas correções. A classe Urbano (3) teve 1.247 pixels corrigidos, o que representa % de toda a classe. Desses, 854 foram realocados para a classe Solo (2). Por outro lado, a classe Corpos Hídricos (1) teve apenas 210 pixels corrigidos, totalizando % de toda a classe. Sua maior relação de correções foi com a classe Mangue (5).

A classe Pastagem (4) se destaca na tabela devido ao número significativo de correções, considerando a proporção das amostras coletadas para essa classe. Foram corrigidos 2.461 pixels, o que resulta em um aumento de % na sua representação na imagem. A maioria desses pixels foi realocada das classes Mosaico de Usos (8) e Cana-de-açúcar (7), totalizando 2.222 pixels. Além disso, a

classe Corpos Hídricos (1) recebeu 245 pixels após a correção, a maioria deles vindo da classe Mangue (5). Por outro lado, a classe Mangue (5) é notável pelo baixo número de pixels recebidos após a pós-classificação, totalizando apenas 342 pixels. A maioria desses pixels foi proveniente das classes Corpos Hídricos (1) e Floresta (6).

Após a etapa de pós-classificação, o mapa final que representa o uso e a cobertura da terra na bacia GL2 está ilustrado na Figura 15.

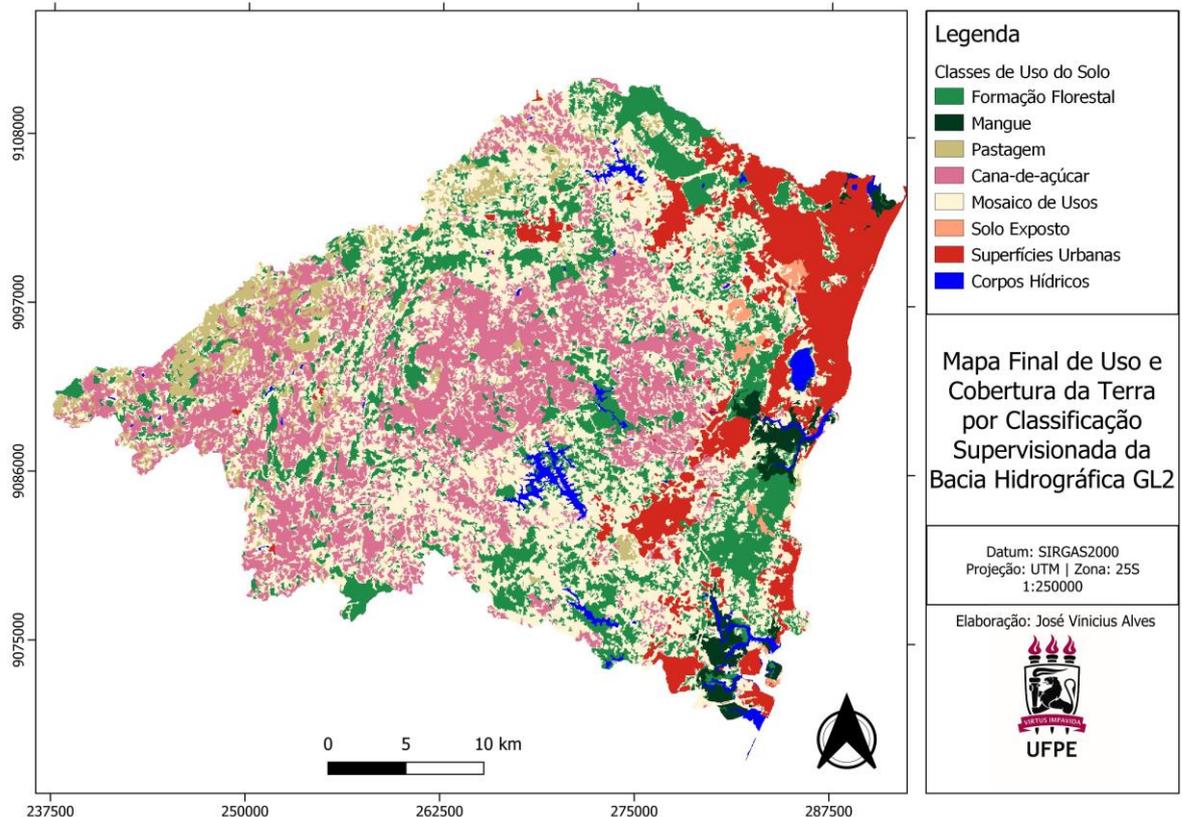


Figura 15 – Mapa Final de Uso e Cobertura da Terra por Classificação Supervisionada

Fonte: Elaborado pelo autor

A Tabela 04 apresenta os valores em quilômetros quadrados (km²) para cada classe mapeada na bacia GL2, juntamente com suas respectivas porcentagens. É evidente que as classes relacionadas à agropecuária se destacam nessa representação. A soma das áreas ocupadas pelas classes Pastagem (4), Cana-de-açúcar (7) e Mosaico de Usos (8) atinge 835,57 km², o que corresponde a impressionantes 68,23% de toda a região, demonstrando seu aspecto mais relevante.

Tabela 04 – Área (km²) das classes no Mapa Final de Classificação Supervisionada.

Classe	Área (km ²)	Porcentagem (%)
1	22.45	1.83
2	8.05	0.66
3	139.16	11.36
4	60.89	4.97
5	24.00	1.96
6	195.36	15.95
7	309.69	25.29
8	464.99	37.97

Fonte: Elaborado pelo autor.

5.2 Comparativo das classificações

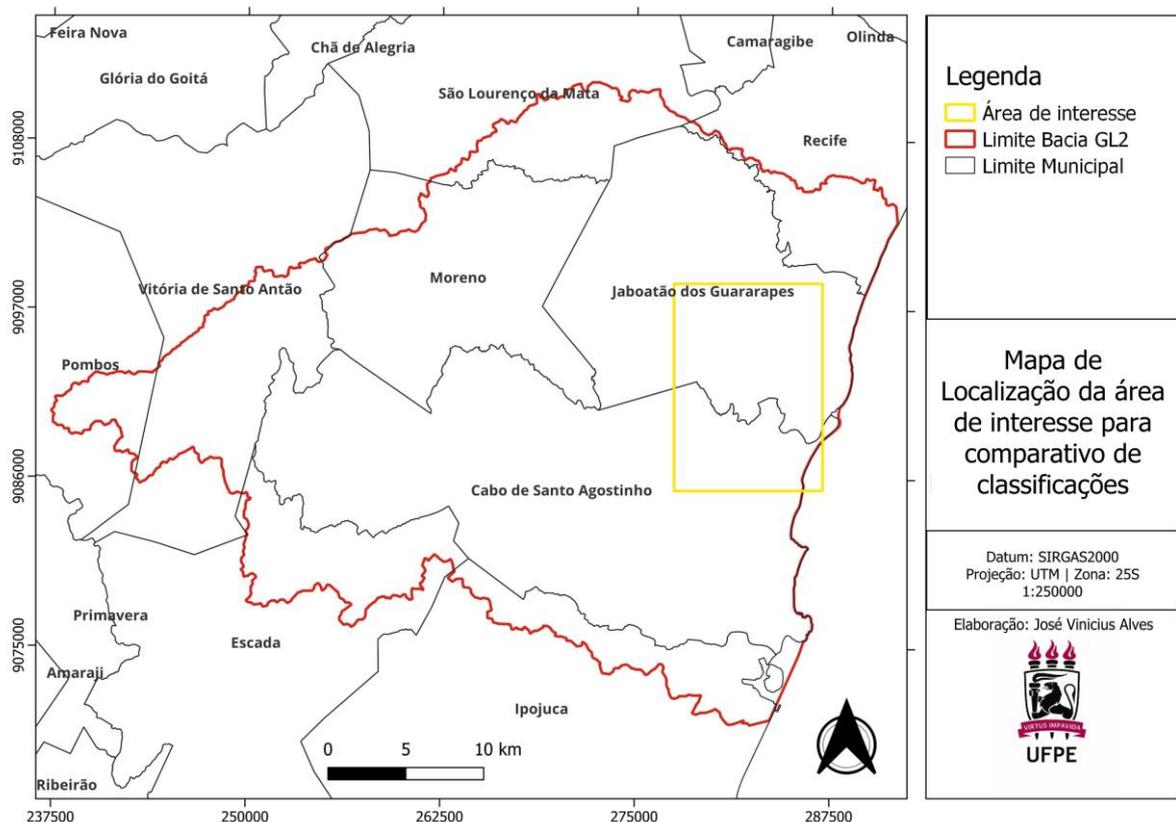


Figura 16 – Mapa de Localização da área de interesse para comparativo de classificações.
 Fonte: Elaborado pelo autor.

Na comparação entre a classificação supervisionada e não-supervisionada, a área delimitada para análise, conforme mostrado na Figura 16, não incluiu a classe Pastagem, que não estava presente nas etapas de amostragem e na classificação

anterior. Como resultado, a classe Mosaico de Usos agora será tratada como a classe (4) nesta fase do estudo.

A Figura 17 mostra os mapas gerados por cada classificação no GEE e depois submetidos ao processo de pós-classificação.

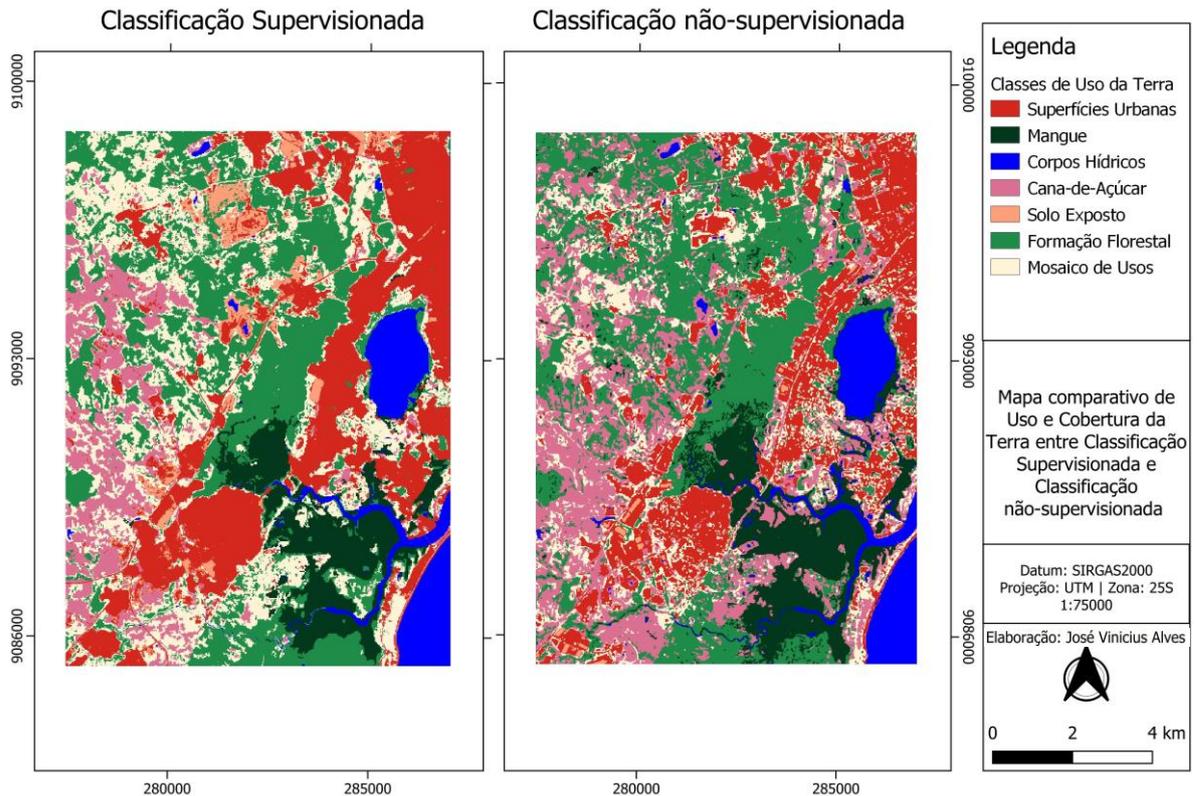


Figura 17 – Mapa comparativo de Uso e Cobertura da Terra entre Classificação Supervisionada e não-supervisionada. Fonte: Elaborado pelo autor.

O Quadro 09 apresenta os valores das áreas em quilômetros quadrados para cada uma das classificações realizadas, juntamente com a porcentagem que cada classe representa em relação à área recortada. Além disso, a tabela também destaca as discrepâncias entre as duas classificações.

Quadro 09 – Área (km²) das classes na comparação entre a classificação supervisionada e não-supervisionada.

Classe	Random Forest		Kmeans		Discrepância	
	Área (km ²)	Porcentagem (%)	Área (km ²)	Porcentagem (%)	Área (km ²)	Porcentagem (%)
1	7.99	6.18	8.02	6.24	0.03	0.06
2	4.65	3.60	0.77	0.60	3.88	3.00
3	30.68	23.74	23.98	18.66	6.70	5.08
5	10.58	8.19	13.99	10.89	3.41	2.70
6	32.04	24.78	35.53	27.64	3.49	2.85
7	12.25	9.47	27.89	21.70	15.65	12.23
8	31.08	24.04	18.36	14.28	12.72	9.76

Fonte: Elaborado pelo autor.

A maior discrepância ocorre na classe Cana-de-açúcar, onde a classificação não-supervisionada apresenta 15,65 km² a mais do que a classificação supervisionada, representando 12,23% da área total. A menor discrepância é observada na classe Corpos Hídricos (1), onde apenas 0,03 km², o que equivale a 0,06% da área, foi modificado entre as duas classificações.

É observado que na classificação por Random Forest, a combinação das classes Solo (2) e Urbano (3) abrange uma área muito maior, totalizando 35,33 km², ou seja, 27,34% da área total. Isso contrasta com a classificação por Kmeans, que abrange 24,75 km², representando 19,26% do total. Essa discrepância é significativa, com uma diferença acumulada de 10,58 km², correspondendo a 8,08% do território total. Além disso, a classe Mosaico de Usos (4) também apresentou um acréscimo de 12,72 km² na classificação por Random Forest em comparação com a classificação por K-means.

As Tabelas 05 e 06 exibem as matrizes de confusão geradas para cada uma das classificações, e na Quadro 10 os valores calculados para a Acurácia Global e o Índice Kappa posteriormente.

Tabela 05 – Matriz de erro da Classificação não-supervisionada na área recortada.

	1	2	3	4	5	6	7	
1	57	0	2	0	2	0	0	61
2	0	21	17	5	0	1	0	44
3	0	5	170	9	0	4	3	191
4	0	0	38	121	0	26	13	198
5	3	0	0	0	112	6	11	132
6	0	0	17	22	6	130	9	184
7	0	0	5	57	0	8	108	178

Fonte: Elaborado pelo autor.

Tabela 06 – Matriz de erro da Classificação Supervisionada na área recortada.

	1	2	3	4	5	6	7	
1	60	0	0	0	1	0	0	61
2	0	31	8	5	0	0	0	44
3	0	5	182	3	0	1	0	191
4	0	0	26	147	0	17	8	198
5	4	0	0	0	124	4	0	132
6	1	0	0	5	3	169	6	184
7	0	0	7	16	0	6	149	178

Fonte: Elaborado pelo autor

Quadro 10 – Relação de Acurácia entre a Classificação Supervisionada e não-supervisionada

	Random Forest	Kmeans
Acurácia Global	94.26%	89.78%
Índice Kappa	0.91	0.87

Fonte: Elaborado pelo autor.

Tanto a Acurácia Global quanto o Índice Kappa indicam que a classificação supervisionada superou a classificação não-supervisionada. A Acurácia Global alcançou 94,26% com o classificador Random Forest, o que representa um aumento de 4,48% em relação ao classificador K-means. O Índice Kappa também favoreceu a classificação não-supervisionada, aumentando 0,4 em relação à classificação não-

supervisionada. No entanto, é importante ressaltar que ambos os classificadores demonstraram alta eficiência, com valores de validação consistentes. Ambos atingiram pelo menos 85% de Acurácia Global e valores acima de 0,8 para o Índice Kappa. Além disso, ao analisar as matrizes de confusão, é possível identificar um padrão de erros semelhantes aos observados nas matrizes e tabelas de áreas discutidas anteriormente.

6 CONCLUSÃO

A pesquisa realizada teve como objetivo central a análise abrangente e comparativa de diferentes métodos de classificação de uso e cobertura da terra na Bacia Hidrográfica do Grupo de Pequenos Rios Litorâneos (GL2). Utilizando a poderosa plataforma Google Earth Engine como ferramenta de processamento e análise, buscamos avançar na compreensão e na aplicação prática dessa temática.

Os resultados alcançados na classificação supervisionada superaram as expectativas, demonstrando excelentes desempenhos com base nas validações realizadas, que incluíram a análise da matriz de erro, acurácia global e índice Kappa. De forma destacada, observamos que as atividades agropecuárias ainda predominam na bacia, totalizando aproximadamente 70% de sua cobertura.

A utilização da plataforma Google Earth Engine, embora tenha grande potencial para o mapeamento de grandes áreas, apresentou limitações na classificação não supervisionada, exigindo a realização de um recorte da área para fins de comparação. Nesse sentido, os resultados revelaram que o método supervisionado, empregando o algoritmo Random Forest de aprendizado de máquina, obteve um desempenho superior em relação à classificação não supervisionada, utilizando os algoritmos de clusterização e K-means.

Ao concluir este estudo, esperamos que os resultados aqui obtidos sirvam como alicerce sólido para futuras investigações e estudos na área. Acreditamos que a ampliação do conhecimento e da aplicação prática do tema contribuirá significativamente para o manejo sustentável dos recursos naturais e a tomada de decisões informadas em relação ao uso e ocupação da terra em bacias hidrográficas como o GL2.

REFERÊNCIAS

ACCIOLY, L. J. de O.; CAVALCANTI JUNIOR, E. de A.; SILVA, A. B. **Relação entre o uso atual e o potencial agroecológico das terras da bacia do Rio Pajeú (PE)**. 2016. Disponível em: <http://www.alice.cnptia.embrapa.br/alice/handle/doc/1059986> Acesso em: 16 set. 2023.

ANDERSON, James R. Land-Use Classification Schemes. **Photogrammetric Engineering**, [s.l.], v. 37, n. 4, p.379-387, abr. 1971.

ANELLO, E. K-Means Clustering in R Tutorial. **Data Camp**. 2023. Disponível em: <https://www.datacamp.com/tutorial/k-means-clustering-r> Acesso em 15 set. 2023.

BERNARDES, T. **Caracterização do ambiente agrícola do Complexo Serra Negra por meio de sensoriamento remoto e sistemas de informação geográfica**. Dissertação (Mestrado). Universidade Federal de Lavras, Minas Gerais, 2006, 119p. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/3483?mode=full>. Acesso em: 20 Ago. 2023.

BRASIL. **Lei Nº 9.433, de 8 de janeiro de 1997**. Institui a Política Nacional de Recursos Hídricos. Brasília, DF: Diário Oficial da União, 1997.

BRAZ, A. M.; BRAZ, A. M.; BONI, P. V.; GARCIA, P. H. M.; PINTO, A. L.; OLIVEIRA, I. J. de. Uso, Cobertura e Manejo da Terra: Contribuições teórico-metodológicas e subsídios à conservação ambiental na Bacia Hidrográfica do Córrego Ribeirãozinho, Selvíria/MS. **Caminhos de Geografia**, Uberlândia, v. 18, n. 64, p. 310–327, 2017. Disponível em: <https://seer.ufu.br/index.php/caminhosdegeografia/article/view/40936>. Acesso em: 07 set. 2023.

Breiman, L. Random Forests. **Machine Learning** 45, 5–32 (2001).

CÂMARA, G. **Sistemas de Informação Geográfica para Aplicações Ambientais e Cadastrais: uma visão geral**, 1998. Disponível em: <http://www.dpi.inpe.br/gilberto/papers/analise.pdf> Acesso: 07 set. 2023.

CAMPOS, W. W.; GASPAR, J.; LAGE, M. de O.; KAWASHIMA, R. S.; GIANNOTTI, M. A.; QUINTANILHA, J. A. Avaliação de Classificadores de Imagem de Satélite a partir do uso de uma técnica de votação. **Revista Brasileira de Cartografia**, [S. l.], v. 68, n. 8, 2016. DOI: 10.14393/rbcv68n8-44385. Disponível em: <https://seer.ufu.br/index.php/revistabrasileiracartografia/article/view/44385>. Acesso em: 30 set. 2023.

COSTA, F. E.; RODRIGUES, R. Z. S. Caracterização Física com uso de geotecnologia na bacia hidrográfica do grupo de pequenos rios litorâneos (gl2). **JHRS**, v.12, n.3, p.88-98. 2022.

CHU, D. Remote sensing of Land Use and Land Cover in Mountain Region: A Comprehensive Study at the Central Tibetan Plateau. **Springer Singapore**, 2019. 227p.

D'alge, J. C. L. **Fundamentos de cartografia para geoprocessamento**. 2003. Disponível em: http://www.dpi.inpe.br/~julio/arquivos/carto_geoprocessamento.pdf Acesso em 07 set. 2023.

ESA. **Sentinel-2**. 2023. Disponível em: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2 Acesso em 05 set. 2023.

FERREIRA, C. C. M. **Zoneamento agroclimático para implantação de sistemas agroflorestais com eucaliptos, em Minas Gerais**. Viçosa, MG: Universidade Federal de Viçosa, 1997. 158p.

FINKLER, R. Unidade 1: a bacia hidrográfica. **Planejamento, manejo e gestão de bacias**, 2013. Disponível em: https://planejamento.mppr.mp.br/arquivos/File/bacias_hidrograficas/planejamento_manejo_e_gestao_unidade_1.pdf Acesso: 07 set. 2023.

PRINCIPE, R. E. geetools-code-editor. **GitHub**. 2019. Disponível em: <https://github.com/fitoprincipe/geetools-code-editor> Acesso em 24 ago. 2023.

FLORENZANO, T. G. **Iniciação em Sensoriamento Remoto**. 3. Ed. São Paulo, Oficina de Textos, 2007. 128p.

FLORENZANO, T. G. **Imagens de satélite para estudos ambientais**. São Paulo: Oficina de textos, 2002.

FOLEY, J. A. et al. Global Consequences of Land Use. **Science**, v.309, n.5734, p. 570-574. 2005.

FOODY, G. M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. **Remote Sensing of Environment**, v.239, 111630, 2020.

FURUYA, M. T. G., FURUYA D. E. G. Aplicação de Machine Learning na identificação de e-mails como spam. **Colloquium Exactarum**, v. 12, n3, p. 31 –38. 2020.

GANEM, K. A. **Classificação da cobertura do solo na Caatinga a partir de imagens do Landsat-8 e da ferramenta Google Earth Engine: uma comparação entre dados com e sem correção atmosférica**. 2017. 182f. Dissertação (Mestrado em Geociências). Instituto de Geociências. Universidade de Brasília. Brasília, 2017.

GARCEZ, G.A; ALVAREZ, L.N. **Hidrologia**. 2 ed. São Paulo, Edgar Blücher Ltda., 2002. 291 p.

GOOGLE EARTH ENGINE. Guides and Tutorials. **Developers Google Earth Engine**. 2023. Disponível em: <https://developers.google.com/earth-engine/tutorials/tutorials> Acesso em 12 set. 2023.

GORELICK, N. **Google Earth Engine**, 2013. Disponível em:
https://projects.listic.univ-smb.fr/seminaires/EarthEngine_LISTIC.21062016.pdf
Acesso: 25 ago. 2023.

GORELICK, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **ScienceDirect**, v.202, p.18-27. 2017.

GUEDES, A. R. M.; GUIMARÃES, V. L. **Sistema de Reconhecimento baseado em Random Forest para Caracteres de Captchas**, 2013. Disponível em:
<http://www.decom.ufop.br/menotti/rp142/trab/trab1-dp2-artigo.pdf> Acesso: 25 ago. 2023.

INPE. **Classificação de Imagens**. 2006. Disponível em:
<https://www.dpi.inpe.br/spring/portugues/tutorial/classific.html#:~:text=O%20que%20%C3%A9%20p%C3%B3s%2Dclassifica%C3%A7%C3%A3o,classificada%20com%20apar%C3%Aancia%20menos%20ruidosa> Acesso em 11 set. 2023.

LANDIS, J.; KOCH, G. The measurement of observer agreement for categorical data, Washington, USA. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

LILLESAND, M. T.; KIEFER, F. W. **Remote Sensing and Image Interpretation**. 3 ed. John Wiley & Sons, 1994.

LEITE, M. E.; DE ALMEIDA, M. I. S. de A. S.; VELOSO, G. A.; FERREIRA, M. F. F. Sensoriamento Remoto aplicado ao Mapeamento da Dinâmica do Uso do Solo na Bacia do Rio Pacuí, no Norte de Minas Gerais, nos anos de 1989, 1999 e 2009. **Revista do Departamento de Geografia**, [S. l.], v. 23, p. 217-231, 2012. DOI: 10.7154/RDG.2012.0023.0009. Disponível em:
<https://www.revistas.usp.br/rdg/article/view/47211>. Acesso em: 08 set. 2023.

LELIS, F. S. et al. Dimensões do Uso e Cobertura da Terra nas Mesorregiões do Estado do Pará. **Revista Espacios**, v.37, n.5, p.5. 2016.

MAPBIOMAS. Legenda Coleção 8 – Descrição Detalhada. **MapBiomias**. 2022.

Disponível em: <https://brasil.mapbiomas.org/wp-content/uploads/sites/4/2023/09/Legenda-Colecao-8-Descricao-Detalhada-PDF-PT-3-1.pdf>

Acesso em 09 set. 2023.

MENESES, P. R.; ALMEIDA, T. **Introdução ao Processamento de Imagens de Sensoriamento Remoto**. Brasília, 2012, 276p.

PAL, M. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, 26:1, 217-222. 2022.

PONZONI, F. J.; SHIMABUKURO, Y. **Sensoriamento Remoto no estudo da Vegetação**. 1 ed. São José dos Campos: Parêntese, v.1, p.127. 2007.

NOVO, E. M. L. M. **Sensoriamento Remoto: Princípios e Aplicações**. 4 ed. São Paulo, Edgar Blücher Ltda., 2010. 388p.

QUEIROZ, V. B. **Análise Comparativa entre os algoritmos Máxima Verossimilhança e Random Forest**. 2021. 105 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Cartográfica e Agrimensura) - Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife.

OLIVEIRA, C. E. **Identificação da passada de um exoesqueleto utilizando algoritmo de classificação não supervisionado k-means**. 2022. 123 p. Dissertação (Mestrado em Engenharia Mecânica na Área de Projetos Mecânicos) - Faculdade de Engenharia, Universidade Estadual Paulista, Bauru, 2022.

Pernambuco. Secretaria de Ciência, Tecnologia e Meio Ambiente. **Plano Estadual de Recursos Hídricos de Pernambuco**. 1998. Disponível em: <http://www.sema.mt.gov.br/PERH>. Acesso em: 14 Ago. 2023.

REY, S.; ARREBAS-BEL, D.; WOLF, L. J. **Geographic Data Science with Python**. CRC Press, 2023. 410 p.

ROCHA, L. S. F. **Aplicação de Algoritmos não supervisionados (k-means e clara) para distinção de classes de uso e cobertura vegetal utilizando o satélite Sentinel-2**. 2021. 134 f. Trabalho de Conclusão de Curso (Graduação em Geografia) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

SANTOS, A. L. F., BARBOSA, F. P. Sistema de Informações Geográficas (SIG) – A importância do uso da ferramenta SIG para projetos de gasodutos. **Revista Científica Multidisciplinar Núcleo do Conhecimento**. Ano 03, Ed. 08, Vol. 13, pp.88-101. 2018.

SANTOS, M. L. F. **Identificação de áreas prioritárias para conservação da Caatinga na bacia do alto curso do Rio Paraíba/PB através do uso das geotecnologias**. 2018. 132 p. Dissertação (Mestrado em Gerenciamento Ambiental) - Programa de Pós-Graduação em Desenvolvimento e Meio Ambiente, Universidade Federal da Paraíba, João Pessoa, 2018.

SILVA, R. V. **Avaliação de Classificadores não-supervisionados, isodata e k-means, para o uso da terra na bacia do rio das ondas**. 2009. 37 f. Trabalho de Conclusão de Curso (Graduação em Geografia) - Instituto de Ciências Humanas, Universidade de Brasília, Brasília.

SCHIAVETTI, A.; CAMARGO, A. F. M. **Conceitos de Bacias Hidrográficas: Teorias e Aplicações**. Ilhéus, Editus, 2002. 293p.

TEIXEIRA J. L.; GALVÍNCIO J. D. Uso de Geotecnologias para a Caracterização Física Espacial da Bacia Hidrográfica do Grupo de Pequenos Rios Litorâneos (GL2) / PE. **RGBF**, v. 3, n. 2, p. 132-138, 2010.

TUAMA, D. O. What are Libraries in Python? **Code Institute**. 2020. Disponível em: <https://codeinstitute.net/global/blog/what-are-libraries-in-python/> Acesso em 15 set. 2023.

TUCCI, C. E. M. (org.). **Hidrologia**: ciência e aplicação. Porto Alegre: Editora da Universidade: ABHR, 1997. 943 p.

TUNDISI, J. G. Novas perspectivas para a gestão de recursos hídricos. **Revista USP**, [S. l.], n. 70, p. 24-35, 2006. DOI: 10.11606/issn.2316-9036.v0i70p24-35. Disponível em: <https://www.revistas.usp.br/revusp/article/view/13529>. Acesso em: 08 set. 2023.

VIEIRA, F. D.; OLIVEIRA, S. R. DE M.; PAIVA, S. R.. Metodologia baseada em técnicas de mineração de dados para suporte à certificação de raças de ovinos. **Engenharia Agrícola**, v. 35, n. 6, p. 1172–1186, nov. 2015.

XIAN, G., CRANE, M. An analysis of urban thermal characteristics and associated land cover in Tampa Bay and Las Vegas using Landsat satellite data. **Remote Sens. Environ**, v.104, p.147–156. 2006.

XIE, S., LIU, L., ZHANG, X., YANG, J., CHEN, X., & GAO, Y. Automatic land-cover mapping using landsat time-series data based on google earth engine. **Remote Sensing**, v. 11, n.24, p. 3023. 2019.