



Universidade Federal Fluminense  
Instituto De Geociências  
Departamento De Geologia E Geofísica  
Programa De Pós-Graduação Em Dinâmica Dos Oceanos E Da Terra

**JAIRSON ALBERTO SAMI**

TEMA

**Aplicação de algoritmos de Machine Learning para Classificação da Cobertura da Terra a partir de série temporal de imagem landsat 8: Estudo de caso do Parque Natural dos Tarrafes de Cacheu, Guiné-Bissau**

NITERÓI-RJ

2025

JAIRSON ALBERTO SAMI

**TEMA**

**Aplicação de algoritmos de Machine Learning para Classificação da Cobertura da Terra a partir de série temporal de imagem landsat 8: Estudo de caso caso Parque Natural dos Tarrafes de Cacheu**

Dissertação apresentada ao Programa de Pós- Graduação em Dinâmica dos Oceanos e da Terra do Departamento de Geologia e Geofísica da Universidade Federal Fluminense, como requisito parcial à obtenção do título de Mestre em Geologia.

**Orientador**

Prof. Dr. Guilherme Borges Fernandez

Dr. Pedro José Farias Fernandes

NITERÓI-RJ

2025

Ficha catalográfica automática - SDC/BIG  
Gerada com informações fornecidas pelo autor

S187a Sami, Jairson Alberto  
Aplicação de algoritmos de Machine Learning para  
Classificação da Cobertura da Terra a partir de série  
temporal de imagem landsat 8: Estudo de caso caso Parque  
Natural dos Tarrafes de Cacheu / Jairson Alberto Sami. - 2025.  
68 f.

Orientador: Guilherme Borges Fernandez.  
Coorientador: Pedro José Farias Fernandes.  
Dissertação (mestrado)-Universidade Federal Fluminense,  
Instituto de Geociências, Niterói, 2025.

1. Inteligência Artificial. 2. GEOBIA. 3. Análise  
Espacial. 4. Ecossistemas Costeiros. 5. Produção  
intelectual. I. Fernandez, Guilherme Borges, orientadora. II.  
Fernandes, Pedro José Farias, coorientadora. III.  
Universidade Federal Fluminense. Instituto de Geociências.  
IV. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

JAIRSON ALBERTO SAMI

TEMA

**Aplicação de algoritmos de Machine Learning para Classificação da Cobertura da Terra a partir de série temporal de imagem landsat 8: Estudo de caso caso Parque Natural dos Tarrafes de Cacheu**

Dissertação apresentada ao Programa de Pós- Graduação em Dinâmica dos Oceanos e da Terra do Departamento de Geologia e Geofísica da Universidade Federal Fluminense, como requisito parcial à obtenção do título de Mestre em Geologia.

Aprovado em 14 de fevereiro de 2025.

**Banca Examinadora:**

---

Prof. Dr. Guilherme Fernandez - UFF (Orientador)

---

Dr. Pedro José Farias Fernandes - UFF (Coorientador)

---

Examinador 1- Leonardo Guimaraes Miquelutti - UFF

---

Examinador 2- Prof. Dr. Debora da Paz Gomes Brandão Ferraz - UFRJ

---

Examinador 4- Prof. Dr. Rômulo Weckmüller - UFF

## AGRADECIMENTOS

Agradeço a Nachibatchi (Deus), que, com Sua sabedoria infinita, guiou meus passos e me fortaleceu em cada desafio. À minha família, que sempre foi o alicerce onde pude me apoiar. Dedico especialmente esta conquista à memória do meu querido pai e à minha mãe, Esperança Ntab, cuja esperança e amor incondicional são uma chama eterna na minha vida. Estendo esta dedicação às minhas sobrinhas, Nadimari e Quetinha, e, em especial, à Suzete Sami, minha irmã-mãe.

Aos meus amigos e colegas que marcaram esta caminhada, minha gratidão é imensa: ao Mamadou Gning, Mohamed Mboup e Victor da Costa, pelas conversas descontraídas, sinceras e sábias durante toda a nossa estada no RJ; ao Daniel Manenga, por ser uma companhia fiel nos momentos de estudo e nas idas à Cantareira; ao Makuta, que se destaca pela sua generosidade e pela união que promove entre os africanos na UFF; ao Amadu Djau, ao Nhancai Da Silva e ao Fernando Oliveira Jr., cuja amizade tem sido constante desde a minha chegada ao Brasil; à Anarela Badinca, cuja motivação foi fundamental nessa caminhada; à Desejada Nazaré Mendes, que, com sua generosidade silenciosa e presença serena, contribuiu significativamente nessa jornada da vida; e ao meu primo Roque, Armandina, Heliana e Nadine, confidentes de meus desabafos sobre as tempestades da vida.

Aos meus orientadores, Guilherme Borges e Pedro Fernandes, minha eterna gratidão: a Guilherme, por sua orientação precisa e paciência, e a Pedro, meu amigo, com quem compartilhei tantos momentos de reflexão de vida e aprendizado.

À minha ex-orientadora Cristiane Nunes Francisco, cuja orientação transcende os muros da UFF. Ela sempre acreditou no meu potencial e foi a primeira a me guiar no mundo da pesquisa.

E, com o coração pleno de gratidão, agradeço aos profissionais que, como eu, dedicam sua vida ao estudo e à preservação das áreas costeiras. Vocês são os guardiões de nossos mares e manguezais e, sem o seu trabalho incansável, nossos ecossistemas estariam em perigo. Um agradecimento especial à equipe do LAGEF, pelo acolhimento no laboratório onde desenvolvi a pesquisa, e à direção do IBAP e, em particular, aos guarda-parques, que me auxiliaram com tanta dedicação e profissionalismo durante o trabalho de campo.

Ao Governo Brasileiro e à CAPES, que, por meio da bolsa de mestrado, abriram portas para que eu pudesse desenvolver minha pesquisa e contribuir com o conhecimento sobre os nossos ecossistemas costeiros. Sem esse apoio institucional, esta jornada teria sido impossível.

**Resumo:** O objetivo geral deste trabalho é classificar a cobertura da terra no Parque Natural dos Tarrafes de Cacheu (PNTC), conhecido por abrigar o maior bloco contínuo de manguezais na Guiné-Bissau, utilizando algoritmos de Machine Learning (ML) e uma série temporal de imagens Landsat 8/OLI de 2019 a 2023. Os objetivos específicos são: (i) avaliar o desempenho de diferentes algoritmos de Machine Learning na classificação da cobertura da terra; (ii) analisar os atributos extraídos das imagens e sua influência na qualidade da classificação gerada pelo algoritmo com melhor desempenho; e (iii) quantificar a variação das áreas das classes de cobertura da terra. Primeiramente, dados do satélite Landsat 8/OLI foram obtidos através da plataforma Google Earth Engine. Os dados foram filtrados para o mês de abril dos anos de 2019 a 2023, correspondente à estação seca. Na segunda etapa, foram calculados Índices de Vegetação (IVs) como NDVI, NDWI, MNDWI, MMRI e EVI. Na terceira etapa, foi feita a segmentação a partir da imagem NDVI de 2019 e 1000 polígonos gerados pela segmentação foram rotulados nas seguintes classes: Agricultura, Mangue, Superfície Hipersalina, Rio, Outra Vegetação, Vegetação Densa e Área Construída que serviram como dados de treinamento e testes de modelos de ML (70% para treinamento e 30% para testes). Na quarta etapa, houve o pré-processamento dos dados de treinamento e os modelos de ML Random Forest (RF), Support Vector Machine (SVM), K-nearest neighbors (KNN) e Multi-Layer Perceptron (ML) foram treinados e avaliados com base na Acurácia, Precisão, Recall, F1 Score, Cross validation accuracy (CV Accuracy) e Área sob a curva ROC (ROC AUC). Na última etapa, o modelo de melhor desempenho foi utilizado na classificação da cobertura da terra e seu índice Kappa foi Recall 0.7847 calculado a partir de 500 amostras de validação coletadas em campo. Os modelos apresentaram todas as métricas acima de 0.86. O MLP apresentou acurácia, precisão, Recall e F1-Score com valores de 0.984, indicando excelente capacidade de classificação, também apresentando maior CV Accuracy (0.975) e maior ROC AUC (0.997). Entretanto, após o exame das curvas de validação, foi visto que os modelos apresentam diferentes comportamentos em termos de generalização, com Random Forest e SVM exibindo um bom equilíbrio entre as métricas de treinamento e validação, enquanto KNN e MLP mostram tendência ao sobreajuste com certos valores de hiperparâmetros. Para efeito da análise o SVM coloca-se em vantagem na generalização. Pois foi usada para a classificação final, verificou-se que Entre 2019 e 2023, as áreas do parque mostram estabilidade geral, com flutuações ligadas a ciclos sazonais, pressões antrópicas e práticas locais, como plantio de caju e arroz.

**Palavras-Chaves:** *Inteligência Artificial, GEOBIA, Análise Espacial, Ecossistemas Costeiros, Guiné-Bissau*

**Abstract:** The primary goal of this study is to classify land cover in the Cacheu Mangroves Natural Park (PNTC), known for hosting the largest continuous block of mangroves in Guinea-Bissau, using Machine Learning (ML) algorithms and a time series of Landsat 8/OLI images from 2019 to 2023. The specific objectives are: (i) to evaluate the performance of different Machine Learning algorithms in land cover classification; (ii) to analyze the attributes extracted from the images and their influence on the classification quality produced by the best-performing algorithm; and (iii) to quantify the variation in the areas of the land cover classes. Initially, Landsat 8/OLI satellite data were acquired through the Google Earth Engine platform. The data were filtered for April of each year from 2019 to 2023, corresponding to the dry season. In the second stage, vegetation indices (VIs) such as NDVI, NDWI, MNDWI, MMRI, and EVI were calculated. In the third stage, segmentation was performed using the NDVI image of 2019, generating 1,000 polygons that were labeled into the following classes: Agriculture, Mangrove, Hypersaline Surface, River, Other Vegetation, Dense Vegetation, and Built-up Area. These labeled polygons were used as training and testing datasets for the ML models (70% for training and 30% for testing). In the fourth stage, training data underwent preprocessing, and the ML models Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP) were trained and evaluated based on Accuracy, Precision, Recall, F1 Score, Cross-validation Accuracy (CV Accuracy), and Area Under the Receiver Operating Characteristic Curve (ROC AUC). In the final stage, the best-performing model was used for land cover classification, achieving a Kappa index of 0.7847 based on 500 field validation samples. All models showed metrics above 0.86. The MLP model exhibited excellent classification capability, with Accuracy, Precision, Recall, and F1 Score values of 0.984, the highest CV Accuracy (0.975), and the highest ROC AUC (0.997). However, an examination of validation curves revealed differing behaviors among models in terms of generalization. Random Forest and SVM displayed a good balance between training and validation metrics, while KNN and MLP showed tendencies toward overfitting under certain hyperparameter configurations. For analysis purposes, the SVM model was deemed advantageous in terms of generalization and was used for the final classification. The results indicated overall stability in the park's land cover areas between 2019 and 2023, with fluctuations linked to seasonal cycles, anthropogenic pressures, and local practices such as cashew and rice cultivation.

**Keywords:** *Artificial Intelligence, GEOBIA, Spatial Analysis, Coastal Ecosystems, Guinea-Bissau*

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	<b>10</b>
<b>2. ÁREA DE ESTUDO</b>	<b>14</b>
<b>3. MATERIAIS E MÉTODOS</b>	<b>17</b>
3.1 Materiais	17
3.2 Métodos	17
3.2.1 Processamento de imagens de sensoriamento remoto	19
3.2.1.1 Aquisição e processamento de dados no Google Earth Engine	19
3.2.2 Segmentação de Imagens	20
3.2.3 Extração e Avaliação de Atributos	21
3.2.4 Machine Learning	24
3.2.4.1 Pré-Processamento	24
3.2.4.2 Treinamento de modelos de Machine Learning	25
3.2.4.3 Avaliação de modelos de Machine Learning	27
3.2.5 Análise Temporal das Classes de Cobertura da Terra e Validação do Campo	30
3.2.5.2 Série Temporal e Dinâmicas de Mudança	30
3.2.5.2 Validação de Campo e Precisão dos Mapas	31
<b>4. RESULTADOS</b>	<b>32</b>
4.1 Avaliação de Desempenho dos Algoritmos	32
4.2 Impacto dos Atributos na Classificação	37
4.5 Variações Temporais e Validação de Mapa	41
<b>5. DISCUSSÃO</b>	<b>45</b>
<b>5. CONCLUSÃO</b>	<b>48</b>
<b>6. REFERÊNCIAS</b>	<b>49</b>
<b>7. APÊNDICE</b>	<b>62</b>
7.1 Classificação das variáveis segundo o RFE	62
7.2 Registro Fotográfico Do Trabalho De Campo	63

## ÍNDICE FIGURAS

Figura 1: Delimitação e Localização da Área de Estudo	15
Figura 2. Fluxograma das Etapas do Trabalho.	18
Figura 3. Polígonos Segmentados com Base no NDVI de 2019	21
Figura 4. Frequência de Amostras por Classe de Cobertura da Terra	24
Figura 5. Desempenho de Quatro Modelos de ML (SVM, KNN, RF e MLP) Com 72 atributos	32
Figura 6. Curvas de Validação para Quatro Algoritmos de Aprendizado Supervisionado: RF, KNN, SVM e MLP.	35
Figura 7. Comparação de Visual de Modelos de Classificação: RF, KNN, SVM e MLP	37
Figura 8. Ranking de Importância das Variáveis Usando RFE	38
Figura 9. Variação das Métricas de Desempenho com o Número de Atributos nos Modelos de ML (SVM, KNN, RF e MLP)	39
Figura 10. Variação das Áreas Ocupadas por Diferentes Classes de Cobertura do Solo entre 2019 e 2023	41
Figura 11. Análise Espacial das Mudanças na Cobertura da Terra (2019-2023)	43
Figura 12. Matriz de Confusão na Validação do Mapa de Cobertura da Terra.	44
Figura 13. Variação Percentual da Cobertura do Solo entre as Porções Norte e Sul ao Longo do Tempo (2019–2023).	44

## ÍNDICE TABELAS

Tabela 1: Materiais e Ferramentas Utilizados no Processamento de Dados	17
Tabela 2: Chave de Interpretação de Imagens de Satélite para Classes de Cobertura da Terra	23
Tabela 3. Métricas Utilizadas para Avaliação dos Modelos: SVM, KNN, RF e MLP	33

## 1. INTRODUÇÃO

A classificação da cobertura da terra utilizando imagens de satélite e algoritmos de Machine Learning (JASIM; JASIM; AL-HAMEEDAWI, 2024; YAN *et al.*, 2023) tornou-se uma ferramenta crucial para a gestão e monitoramento ambiental (EKEANYANWU *et al.*, 2022; SHALU; GURJEET SINGH, 2023; YAN *et al.*, 2023). Em comparação com os métodos tradicionais, que frequentemente apresentam limitações em termos de precisão e abrangência (KUSSUL *et al.*, 2017; YANG *et al.*, 2024), os algoritmos de Machine Learning oferecem vantagens significativas em termos de precisão e eficiência (BASUKALA *et al.*, 2017).

Tradicionalmente, o processamento de imagens de satélite é realizado através de técnicas tradicionais de Sistemas de Informação Geográfica (GIS), que, embora úteis, possuem limitações em termos de capacidade de processamento e análise de grandes volumes de dados (EKEANYANWU *et al.*, 2022; HERMOSILLA *et al.*, 2022; YANG *et al.*, 2024). No entanto, com o avanço das linguagens de programação e o desenvolvimento de algoritmos de Machine Learning, houve uma transformação significativa nesse campo (NICULESCU; LAM, 2019). Esses algoritmos avançados têm potencializado a capacidade de lidar com grandes volumes de dados de forma mais eficiente e precisa, superando as restrições impostas pelos softwares tradicionais (NICULESCU; LAM, 2019; PARTHA PROTIM ROY; MD. SHAHRIAR ABDULLAH; IQTIAR MD; SIDDIQUE, 2024).

O Machine learning (ML) surgiu nos finais da década de 1950 (RITU; BHAMBRI, 2023), definida como ramo da inteligência artificial em que máquinas aprendem a partir de dados legíveis, identificando padrões e prevendo os resultados (UPRETI, 2022), aprimorando significativamente as aplicações em sensoriamento remoto e GIS (LARY *et al.*, 2016). Há diferentes tipos de aprendizado de máquina, como o aprendizado supervisionado (o algoritmo é treinado a partir de exemplos), o não supervisionado (o algoritmo identifica padrões nos dados a partir de uma métrica de similaridade, sem necessidade de treinamento) (HALDER; GHOSH; GHOSH, 2011), e o aprendizado por reforço (em que o algoritmo executa suas tarefas e aprende com sua própria experiência

(FAN, 2024). Dessa forma, o Machine Learning é aplicado em diversas áreas como visão computacional (KLAIBER; KLOPFER, 2022), processamento de linguagem natural (NECULA; DUMITRIU; GREAVU-ȘERBAN, 2024), processamento de sinais (CALAZANS *et al.*, 2024) e em estudos ambientais utilizando imagens de sensoriamento remoto. Como os algoritmos de Machine Learning são capazes de aprender padrões complexos a partir dos dados, essa técnica tem sido aplicada na bibliografia recente da área de sensoriamento remoto, pois permite uma classificação mais refinada e detalhada da cobertura da terra (EKEANYANWU *et al.*, 2022). Essa evolução tecnológica não só melhora a qualidade dos resultados obtidos, mas também amplia as possibilidades de análise e aplicação em diversas áreas da gestão ambiental, como em mudanças na cobertura do solo, a detecção de desmatamento e a avaliação de impactos ambientais (PARTHA PROTIM ROY; MD. SHAHRIAR ABDULLAH; IQTIAR MD. SIDDIQUE, 2024). Diversos estudos têm destacado a crescente adoção de algoritmos de código aberto como uma alternativa viável à utilização de aplicativos pagos em sensoriamento remoto (EKEANYANWU *et al.*, 2022; PARTHA PROTIM ROY; MD. SHAHRIAR ABDULLAH; IQTIAR MD. SIDDIQUE, 2024). Essa tendência tem impulsionado a aplicação de algoritmos open source em uma ampla gama de tarefas, desde as mais simples até as mais complexas.

Algoritmos e frameworks de Machine Learning estão disponíveis em bibliotecas como Keras, Pytorch, Tensorflow e Scikit-learn. Tais bibliotecas Python contribuíram com a popularização do uso de técnicas avançadas de aprendizado de máquina e de ciência de dados, uma vez que permitem resultados precisos com poucas linhas de código (NGUYEN *et al.*, 2019). Sendo assim, na área de sensoriamento remoto, é utilizado o aprendizado de máquina supervisionado e não supervisionado. A classificação supervisionada baseia-se no treinamento de modelos com dados previamente rotulados, permitindo ao algoritmo aprender a partir de exemplos com classes conhecidas, como na distinção de diferentes tipos de cobertura terrestre em imagens de satélite. Em contraste, a classificação não supervisionada não requer rótulos durante o processo de treinamento, sendo responsável pela identificação autônoma de padrões e agrupamento de dados semelhantes, como na segmentação de áreas vegetativas sem informações prévias (HALDER; GHOSH; GHOSH, 2011; WU *et al.*, 2019). A bibliografia especializada

na área tem mostrado resultados precisos no uso de algoritmos de aprendizado supervisionado para a classificação de imagens de satélite, como o Random Forest, SVM, KNN e redes neurais (A V *et al.*, 2024; OUCHRA; BELANGOUR; ERRAISSI, 2024; ÖZBULDU; ŞEKERLI, 2024).

Apesar de muitos avanços no Machine Learning aplicado ao sensoriamento remoto, duas abordagens principais têm sido utilizadas: a Geographic Object-Based Image Analysis (GEOBIA) e a abordagem Pixel (BLASCHKE, 2010; LANG *et al.*, 2019; NICULESCU; LAM, 2019). Ambas se mostram eficazes em contextos diferentes, muito embora a abordagem pixel possui algumas limitações que podem gerar uma classificação ruidosa (KWENDA; GWETU; FONOU-DOMBEU, 2023). A abordagem baseada em pixel realiza a classificação com base nas características de cada pixel individual (CONTO *et al.*, 2024), o que o torna mais simples e direto, especialmente em casos em que a variabilidade dentro dos objetos geográficos é mínima. Essa abordagem pode ser menos eficaz em contextos em que a resolução espacial e a complexidade dos objetos geográficos são elevadas, pois pode levar a uma classificação menos precisa e a uma maior ocorrência de confusões entre classes semelhantes (POWAR; PANHALKAR; PATIL, 2022).

A abordagem GEOBIA foca na segmentação de imagens em objetos geográficos, em vez de analisar dados pixel por pixel (KWENDA; GWETU; FONOU-DOMBEU, 2023). Ela utiliza critérios de homogeneidade para extrair objetos significativos, considerando textura, cor, forma, espectros e tamanhos (BENZ *et al.*, 2004; BLASCHKE, 2010). O processo inicia-se com cada pixel como um objeto individual, e segmentos maiores são formados pela fusão de objetos adjacentes, baseando-se em três parâmetros principais: escala, compacidade e forma. O parâmetro de escala controla a variação espectral e o tamanho dos objetos segmentados, enquanto o critério de forma pondera a importância relativa entre cor e forma, e o critério de compacidade define o quão compactos são os objetos formados (BENZ *et al.*, 2004; NICULESCU; LAM, 2019). Essa abordagem permite uma análise mais contextualizada e semântica, útil para detectar padrões complexos que não são evidentes na análise pixel a pixel (KWENDA; GWETU; FONOU-

DOMBEU, 2023). A bibliografia demonstra bons resultados no uso do algoritmo de segmentação multirresolução (BAATZ; SCHÄPE, 2000).

A aplicação de técnicas de machine learning ao sensoriamento remoto para a classificação de uso e cobertura da terra por meio GEOBIA tem enfrentado desafios significativos, principalmente no que se refere à seleção dos atributos mais apropriados para melhorar as métricas da classificação (CLEWLEY *et al.*, 2014). A escolha criteriosa desses atributos, que podem incluir variáveis espectrais, texturais, geométricas entre outras, é crucial para o desempenho dos classificadores, uma vez que são esses elementos que permitem ao modelo distinguir com maior precisão as diferentes classes de cobertura terrestre (SHALU; GURJEET SINGH, 2023). A correta definição e extração dos atributos influenciam diretamente a robustez e a confiabilidade dos resultados obtidos (CHAUHAN *et al.*, 2023), principalmente em áreas com dinâmica ambiental complexa e com grande número de classes de cobertura da terra, como os ambientes de manguezais.

A floresta de manguezais, devido à sua natureza dinâmica e vulnerabilidade às ações humanas conjugadas com as alterações climáticas (CINCO-CASTRO; HERRERA-SILVEIRA, 2020; JIA *et al.*, 2019; MOSCHETTO; RIBEIRO; FREITAS, DE, 2021), necessita de monitoramento constante para garantir sua conservação e sustentabilidade. Esses ecossistemas, localizados na interface entre ambientes terrestres e marinhos, enfrentam pressões significativas, como desmatamento, poluição e urbanização (SOUSA *et al.*, 2023; ZHU, B.; LIAO; SHEN, 2021). Para avaliar adequadamente as condições dos manguezais e detectar mudanças precoces, é essencial a implementação de uma série temporal de sensoriamento remoto e análises (ZHU, B.; LIAO; SHEN, 2021). Esse monitoramento contínuo de manguezais para ser eficaz e precisa, além das bandas espectrais, de dados derivados de sensoriamento remoto para permitir a diferenciação adequada de classes de vegetação, como os Índices Espectrais. Há diversos índices espectrais na literatura, porém a bibliografia tem indicado resultados eficazes no uso do NDVI, NDWI, MMRI, EVI, MNDWI (BRUBACHER; GUASSELLI; OLIVEIRA, 2017; DINIZ *et al.*, 2019; KWAN *et al.*, 2020; MCFEETERS, 1996; ZHAO; QU, 2024).

A escolha de algoritmos como Random Forest, SVM, MLP e KNN para análise de desempenho em Machine Learning baseia-se na ampla aplicabilidade e robustez desses métodos em diversas áreas. Estudos recentes destacam que o Random Forest é eficaz em cenários de classificação com alta precisão, embora sua complexidade aumente o tempo de computação (THANH NOI; KAPPAS, 2017). O SVM continua a ser uma técnica preferencial para lidar com dados de alta dimensionalidade e classes desequilibradas devido à sua capacidade de encontrar margens ótimas de separação (THANH NOI; KAPPAS, 2017). As Redes Neurais MLP, amplamente utilizadas para padrões complexos, foram validadas em diferentes aplicações, como a classificação em imagens hiperespectrais, onde mostram maior precisão na generalização em dados de grande escala (HERATH, 2021). Já o KNN é reconhecido por sua simplicidade e adequação em conjuntos de dados menores, sendo comparado sistematicamente com algoritmos mais complexos para assegurar desempenho competitivo em classificação (ZHANG, M.; SHI; XU, 2020).

A Guiné-Bissau possui uma costa rica em áreas de manguezais, fortemente influenciada por um regime de macromaré que varia entre 3 a 7 metros (DODMAN; SÁ, 2005), afetando até 150 km no interior do país (ANEME, 2018). Essa dinâmica costeira, associada ao vasto ecossistema de manguezais no Parque Nacional de Cacheu, integra um contexto único das áreas úmidas da África Ocidental. Esse bioma é vital para a biodiversidade e a subsistência das populações locais, mas verifica-se a lacuna na classificação da cobertura da terra, especialmente com técnicas de Machine Learning, que ainda é pouco explorada na região.

Portanto, para preencher a lacuna científica citada, o presente trabalho tem como objetivo geral classificar a cobertura da terra no Parque Natural dos Tarrafes de Cacheu (PNTC), na Guiné-Bissau, utilizando algoritmos de Machine Learning e uma série temporal de imagens Landsat 8/OLI de 2019 a 2023. Para alcançar este objetivo, estabelecemos os seguintes objetivos específicos: (i) avaliar o desempenho dos algoritmos de Machine Learning, incluindo Random Forest, SVM, Redes Neurais e KNN, na classificação da cobertura da terra; (ii) analisar os atributos extraídos das imagens, tais como NDVI, NDWI, MMRI, EVI, MNDWI e bandas espectrais (1 - 7), e sua influência

na qualidade da classificação gerada pelo algoritmo com melhor desempenho; e (iii) quantificar a variação das áreas das classes de cobertura da terra ao longo da série temporal de imagens analisadas.

## 2. ÁREA DE ESTUDO

A área de estudo é o Parque Natural dos Tarrafes de Cacheu, situado na região de Cacheu, na Guiné-Bissau, África Ocidental, Figura 1.



**Figura 1:** Delimitação e Localização da Área de Estudo

O Parque Natural dos Tarrafes de Cacheu instituído no ano 2000 (GUINÉ BISSAU, 2000), cobre uma área total de 88.615 hectares e é notável por abrigar o maior bloco contínuo de manguezais da África Ocidental (IBAP, 2024). A área é dividida em dois grandes blocos, situados ao norte e ao sul (GUINÉ BISSAU, 2000), cada um com distintas formas de ocupação do solo e práticas de conservação, influenciadas pelas

diversas práticas culturais das populações locais (JUNIOR *et al.*, 2019). O parque encontra-se entre as latitudes 12° 20' N e 12° 40' N e as longitudes 15° 45' W e 16° 15' W.

Dividido em três zonas principais: a Zonas de Preservação, que abriga a maior diversidade biológica e inclui áreas extensas de manguezais protegidos para conservação, onde todas as atividades são proibidas, exceto visitas públicas, observações científicas e práticas agrícolas, silvícolas, piscatorias e pecuárias restritas às comunidades residentes, previstas no regulamento interno do parque; a Zonas de Desenvolvimento Sustentável, voltada para atividades econômicas que promovem a sustentabilidade e beneficiam as comunidades locais, como a agricultura e a pesca sustentável; e a Zona Tampão, que serve como área de transição entre a Zona de Preservação e a Zona de Desenvolvimento Sustentável, onde qualquer atividade potencialmente prejudicial ao meio ambiente necessita de autorização da administração do parque (BISSAU-GUINÉ, 2008; JUNIOR *et al.*, 2019).

A vasta extensão de manguezais e outros ecossistemas úmidos, como pântanos e florestas ribeirinhas, conferem ao parque uma importância ecológica significativa, sustentando uma rica biodiversidade e fornecendo habitats essenciais para inúmeras espécies (JUNIOR *et al.*, 2019). O clima é caracterizado como tropical úmido, com chuvas abundantes nos meses de julho a agosto (WEATHER, 2024). A população local é de aproximadamente 7.120 habitantes, com 830 vivendo na porção sul. O parque é dividido em 27 aldeias, sendo 17 localizadas na região sul (IBAP, 2024). A economia da região é baseada principalmente na agricultura de subsistência e na pesca, embora a gestão sustentável desses recursos seja cada vez mais crítica devido à crescente pressão econômica (JUNIOR *et al.*, 2019).

Nos últimos anos, o avanço da monocultura de cajueiro (SECA; PEREIRA; SILVA, 2021) e outros interesses econômicos, como a pesca ilegal, o abate de árvores de grande porte e a exploração de palmeiras para a construção civil, extração de sal, produção de lenha, carvão e defumagem dos peixes têm representado desafios significativos para a conservação do parque (JUNIOR *et al.*, 2019). Devido à sua localização na zona

fronteiriça com Ziguinchor, sul do Senegal, uma área economicamente ativa (SONKO; DEAC, 2020), o parque enfrenta uma pressão considerável (IBAP, 2024; JUNIOR *et al.*, 2019), intensificada pela chegada de nativos do território senegalense que ali se instalam devido à instabilidade política na região do Senegal (MULA, 2010), e aumentam a pressão sobre exploração dos recursos naturais. Esses fatores alimentam o mercado ilegal tanto no Senegal quanto na Guiné-Bissau, ameaçando a tranquilidade dos habitantes locais e gerando preocupações para as autoridades ambientais (JUNIOR *et al.*, 2019). A gestão eficaz e a vigilância contínua são essenciais para proteger esses ecossistemas frágeis e garantir o desenvolvimento sustentável da região e do PNTC.

### 3. MATERIAIS E MÉTODOS

#### 3.1 Materiais

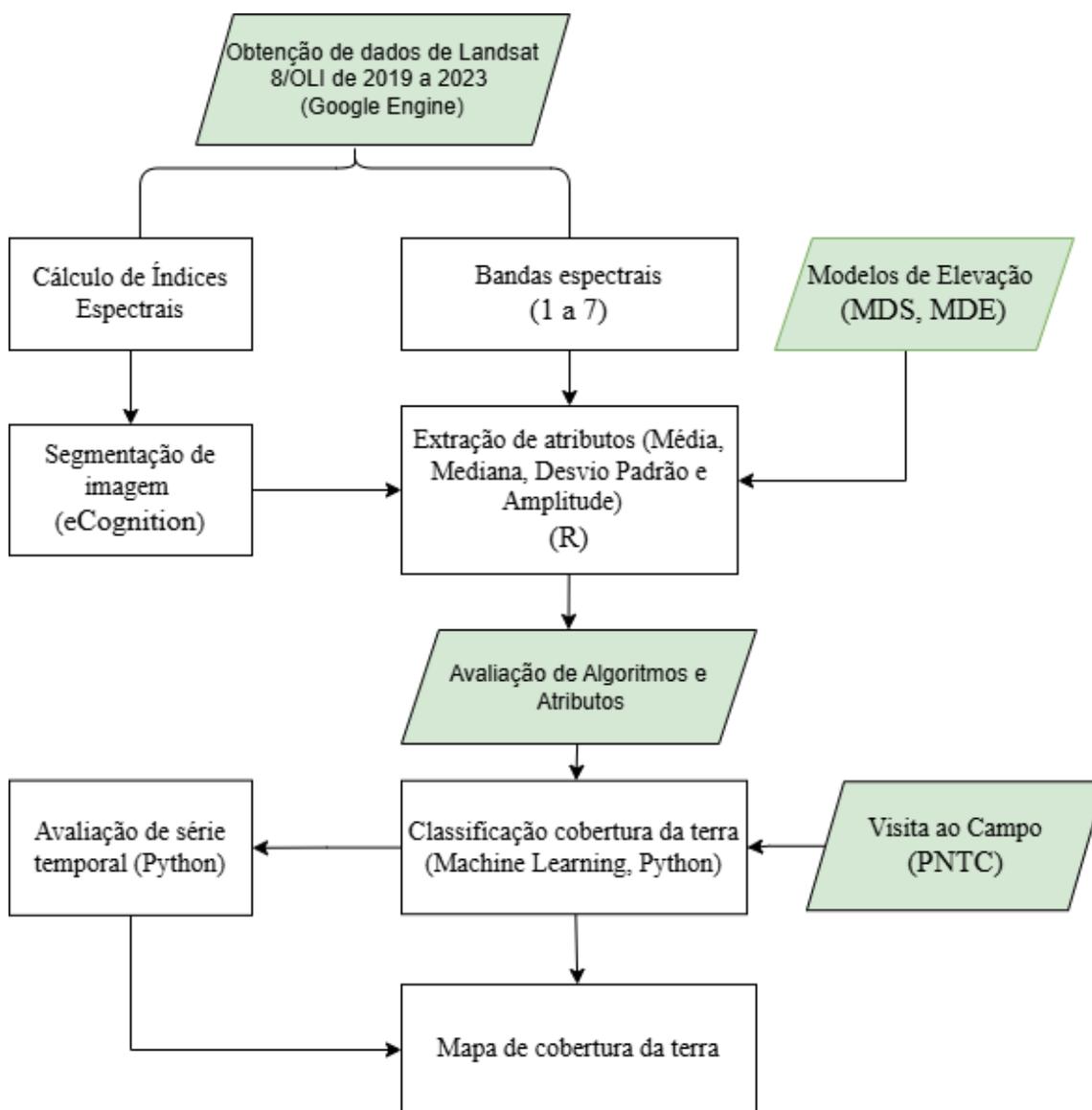
Os materiais utilizados neste trabalho são apresentados na Tabela 1, abaixo. Eles incluem dados de imagens de satélite e modelos digitais de elevações processados no ambiente Google Earth Engine, com as respectivas fontes e características detalhadas.

**Tabela 1:** *Materiais e Ferramentas Utilizados no Processamento de Dados*

<b>Dado</b>	<b>Características</b>	<b>Ferramenta</b>	<b>Versão</b>	<b>Fonte</b>
Imagens de Satélite Landsat 8/OLI	- Período: 2019 a 2023 - Bandas: (1 a 7) - Resolução: 30 metros - Filtragem: Abril (estação seca)	Google Earth Engine	N/A	NASA/ <u>USGS</u>
Modelos Digitais de Elevação do Copernicus	MDS com resolução espacial de 30 metros	Google Earth Engine	N/A	<u>ESA</u>

### 3.2 Métodos

Os procedimentos metodológicos consistem em etapas que visam avaliar os atributos e algoritmos de machine learning por meio de GEOBIA e classificação da cobertura da terra e utilizando uma série temporal de 2019 a 2023, Figura 2.



**Figura 2.** Fluxograma das Etapas do Trabalho.

A primeira etapa consistiu na aquisição e processamento de imagens de satélite pelo Google Earth Engine, incluindo o cálculo de índices espectrais, bandas espectrais

e modelos de elevação. A segunda etapa envolveu a segmentação das imagens e a extração de atributos. A terceira etapa corresponde à avaliação dos atributos e dos algoritmos de machine learning. A quarta etapa teve como objetivo a realização de trabalho de campo para validação da classificação, e a quinta etapa foi dedicada à geração de uma série temporal da cobertura da terra.

### **3.2.1 Processamento de imagens de sensoriamento remoto**

#### **3.2.1.1 Aquisição e processamento de dados no Google Earth Engine**

Os dados de satélite Landsat 8/OLI foram obtidos através da plataforma Google Earth Engine, que combina dados de sensoriamento remoto com serviço de computação de alto desempenho, tornando o processamento de imagens de satélite mais rápido e simples (GOOGLE, 2024). Os dados foram filtrados para a média do mês de abril dos anos de 2019 a 2023, período correspondente à estação seca, caracterizada pelo estresse hídrico, o que facilita a discriminação das classes de cobertura da terra selecionadas na área de estudo uma vez que a prática agrícola ocorre principalmente durante a estação chuvosa (MONTEIRO; BIDARRA; PIFFER, 2020).

A maior parte do processamento das imagens foi realizada no Google Earth Engine, incluindo operações básicas como cálculos de índices espectrais, mosaico, filtros e recorte, conforme detalhado neste [script](#) no Google Earth Engine.

Cálculo dos índices espectrais:

Índice de Vegetação por Diferença Normalizada (NDVI) – corresponde a razão normalizada entre as bandas NIR e Red (equação 1), a fim de destacar a cobertura vegetal e diferenciá-la das regiões de sombra (ZHAO; QU, 2024).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad \text{Equação 1}$$

Índice de Água por Diferença Normalizada (NDWI) - corresponde a razão normalizada entre as bandas NIR e Green (equação 2), a fim de auxiliar no delineamento de feições de corpos de água (MCFEETERS, 1996).

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad \text{Equação 2}$$

Índice de Água por Diferença Normalizada Modificado (MNDWI) - corresponde a razão normalizada entre as bandas Green e SWIR (equação 3), para facilitar a delimitação de áreas inundadas (BRUBACHER; GUASSELLI; OLIVEIRA, 2017).

$$MNDWI = \frac{GREEN - SWIR}{GREEN + SWIR} \quad \text{Equação 3}$$

Índice Modular de Reconhecimento de Mangue (MMRI) corresponde a razão normalizada entre os MNDWI e NDVI (Equação 4), que facilita a separação da floresta de mangue das demais formações florestais (DINIZ *et al.*, 2019).

$$MMRI = \frac{|MNDWI| + |NDVI|}{|MNDWI| - |NDVI|} \quad \text{Equação 4}$$

Índice de Vegetação Aprimorado (EVI), ou Índice de Vegetação Aprimorado, é um índice desenvolvido para melhorar a detecção de vegetação em áreas densamente vegetadas e reduzir algumas das limitações associadas ao NDVI (KWAN *et al.*, 2020). É calculado pela equação 5:

$$EVI = 2.5 * \frac{(NIR - RED)}{(NIR + 6.0 * RED - 7.6 * BLUE + 1)}. \quad \text{Equação 5}$$

### 3.2.2 Segmentação de Imagens

Para a segmentação das imagens (Figura 3), foi utilizado o Índice de Vegetação por Diferença Normalizada (NDVI) do ano de 2019 como base para segmentar a imagem, pois o NDVI realça a vegetação, permitindo distinguir áreas vegetadas de superfícies não vegetadas (COLARES *et al.*, 2015). O processo foi realizado no software eCognition, versão 9.1.3 (ECOGNITION, 2024), utilizando o algoritmo de *Multiresolution Segmentation*. Os parâmetros configurados foram: escala 150, forma 0.1 e compacidade 0.3, foram testados e selecionados através da interpretação visual como ilustrado na Tabela 2 (Sami, *et al.*, 2023). A segmentação de imagem desempenha um papel fundamental na análise de imagens de satélite, dividindo a imagem em segmentos significativos e homogêneos, com base em características como cor, intensidade ou

textura. Esse método é projetado para identificar e agrupar pixels que compartilham propriedades semelhantes, facilitando a classificação e interpretação de diferentes classes de cobertura da terra (BLASCHKE, 2010; LANG *et al.*, 2019).

Illustration of the Segmentation of the Southern Part of the PNTC



**Figura 3.** Polígonos Segmentados com Base no NDVI de 2019

### 3.2.3 Extração e Avaliação de Atributos

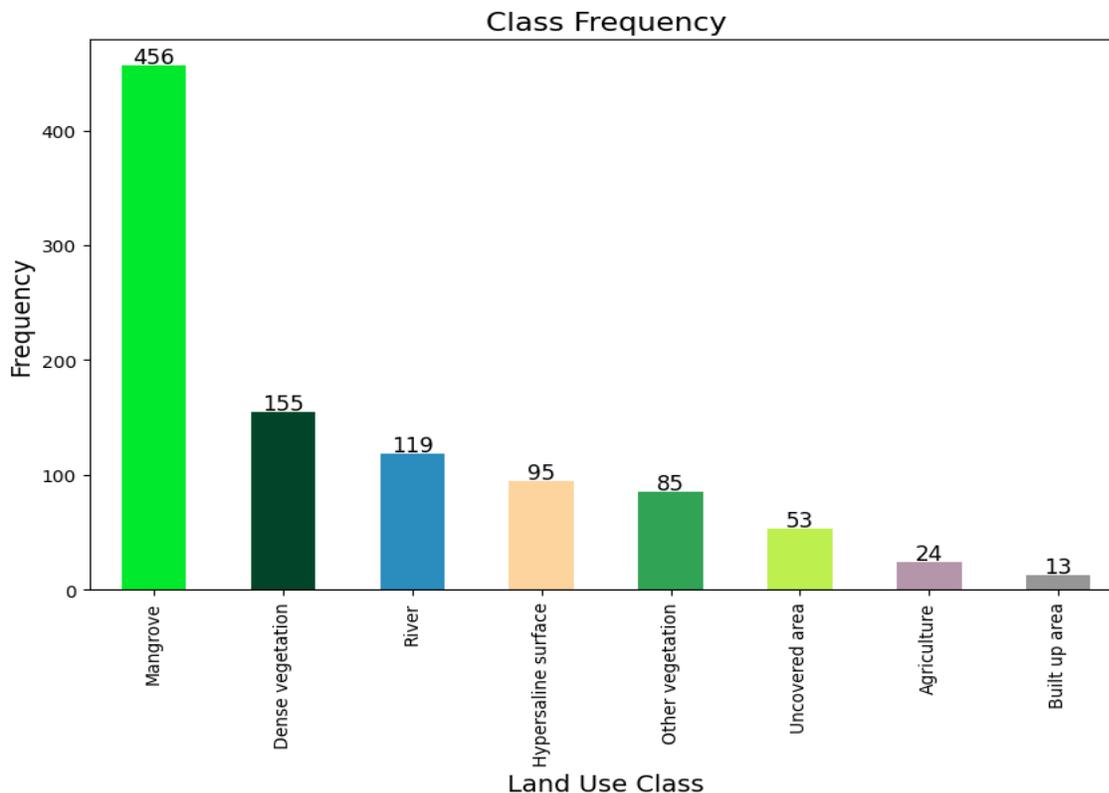
Posteriormente, foi gerado um shapefile com os segmentos identificados totalizando 17201 polígonos que correspondem a área de estudo específica. Em seguida, foram definidas as classes de cobertura da terra: Agricultura, Mangue, Apicum, Rio, Outra vegetação, Vegetação densa, Área construída e Área descoberta, com base na interpretação visual das imagens de satélite (composição colorida) no QGIS e Google Earth. As amostras foram coletadas a partir de 1000 polígonos aleatórios das classes de cobertura da terra definidas.

**Tabela 2:** Chave de Interpretação de Imagens de Satélite para Classes de Cobertura da Terra

Composição Colorida	Classe de Cobertura	Textura e Forma	Observações
	Agricultura	Textura homogênea, com padrão de linhas retas ou regulares	Áreas cultivadas, podendo incluir talhões organizados em áreas perto da linha de costa ou no interior
	Mangue	Textura irregular e densa, padrões em áreas costeiras	Presente em áreas de maré e terrenos salinos próximos a estuários.
	Superfície hipersalino	Textura muito homogênea, suave, contornos difusos	Áreas de salinas naturais ou solo exposto a alta salinidade, próximos de área de mangue.
	Rio	Forma alongada e sinuosa	De cor azul ou turva e, a largura pode variar com a profundidade e volume de água.
	Outra Vegetação	Textura variada com certo padrão, menos densa que vegetação densa.	Inclui plantação de caju ou vegetação mista em áreas menos densas.
	Vegetação Densa	Textura espessa, densa, contornos irregulares	Florestas ou áreas de vegetação densa, com alta biomassa.
	Área Construída	Textura heterogênea, formas retangulares ou circulares	Edificações e pequenos vilarejos
	Área Descoberta	Textura homogênea, sem vegetação ou com pouca cobertura	Áreas sem ou com pouca, vegetação como solo exposto, campos abertos e pastagem

Foram realizados empilhamento das imagens de 2019 a 2023 de todas as bandas espectrais de 1 a 7, índices de vegetação descritos acima (NDVI, NDWI, MMRI, EVI e MNDWI), e MDS (Modelo Digital de Superfície) do Copernicus DEM (ESA, 2024) utilizando a biblioteca raster GDAL (GDAL, 2024) no ambiente R. Esse processo envolveu a sobreposição dessas imagens com 30 metros de resolução espacial ao *shapefile* da amostra, permitindo a extração de valores estatísticos como média, mediana, desvio padrão e amplitude para cada polígono presente nas imagens de satélite. Esses dados foram então compilados em uma tabela de atributos detalhada. Os atributos extraídos são cruciais para a classificação da cobertura da terra, pois cada um deles possui relevância específica dentro do contexto analisado e impacta de maneira distinta os modelos de algoritmos de machine learning utilizados.

Durante a exploração da área de estudo (Figura 4), foram identificados 17201 polígonos, sendo que a tabela de atributos resultante contempla 367 colunas, abrangendo o intervalo temporal de 2019 a 2023. Um aspecto notável é a significativa desproporção na distribuição das classes de cobertura da terra, com os manguezais dominando em relação às outras categorias. A área construída representa uma porção menor, a região é predominantemente caracterizada por manguezais, o que implica em desafios específicos para a classificação e modelagem das diferentes classes de cobertura da terra.



**Figura 4.** Frequência de Amostras por Classe de Cobertura da Terra

### 3.2.4 Machine Learning

#### 3.2.4.1 Pré-Processamento

Para o desenvolvimento do script, utilizamos o ambiente do Google Collaboratory, uma plataforma poderosa que oferece não apenas recursos robustos para processamento, como GPUs e acesso a bibliotecas Python essenciais, mas também uma integração perfeita com serviços como Google Drive e Google Earth Engine, facilitando o gerenciamento e compartilhamento de dados e resultados (GOOGLE, 2024).

Inicialmente, foram empregadas técnicas de pré-processamento de dados para machine learning, que incluíram a filtragem de valores faltantes, a padronização dos dados, balanceamento das classes entre outros (TEOH; RONG, 2022). Observou-se uma grande desproporcionalidade entre as classes de cobertura da terra, o que

influencia significativamente o treinamento dos modelos. Modelos de classificação treinados com conjuntos de dados desequilibrados tendem a favorecer as classes majoritárias, resultando em classificações incorretas para as classes minoritárias (KRAWCZYK, 2016).

Para abordar esse desbalanceamento, testaram-se técnicas de balanceamento de classes, como Oversampling e Undersampling. O Oversampling aumenta a quantidade de amostras da classe minoritária, seja replicando dados existentes ou criando novas amostras sintéticas por meio de técnicas como o SMOTE (Synthetic Minority Over-sampling Technique). Em contraste, o Undersampling reduz a quantidade de amostras da classe majoritária, removendo dados para equilibrar as proporções das classes (CHAWLA *et al.*, 2002; SINGH *et al.*, 2022). Essas abordagens mostram-se eficazes para melhorar o desempenho dos algoritmos de Machine Learning em cenários com disparidade significativa entre as classes, sendo que o SMOTE se destaca por evitar o overfitting ao gerar novas amostras sintéticas em vez de simplesmente replicar as existentes (SINGH *et al.*, 2022; ZHANG, C.; GUO; LU, 2017).

Foi utilizada a técnica de Eliminação Recursiva de Características (RFE), que visa selecionar os atributos mais relevantes para o modelo (SCIKIT-LEARN, 2024). Este método avalia a importância de cada atributo com base em como sua presença melhora a precisão do modelo. Ao filtrar atributos menos relevantes, a abordagem não só ajuda a identificar os melhores atributos para o modelo, mas também reduz o custo computacional, resultando em uma execução mais rápida tanto do treinamento quanto da previsão. Focar apenas nos atributos mais significativos melhora a eficiência do modelo, reduz o tempo de execução e torna o modelo mais eficiente e interpretável (RAMEZAN, 2022).

#### **3.2.4.2 Treinamento de modelos de Machine Learning**

Para a implementação dos algoritmos e aprendizado de máquina foi usada a biblioteca Scikit-learn, no ambiente Google Colaboratory. A escolha dessa biblioteca se baseou na robustez e na ampla adoção pela comunidade acadêmica, permitindo a aplicação dos modelos preditivos com eficiência (BISONG, 2019; SALAMA, 2024). Para

o treinamento e teste dos modelos, o conjunto de dados em universo de 1000 amostras foi dividido em 70% para treinamento e 30% para teste, garantindo uma proporção equilibrada para avaliar a generalização dos algoritmos. Essa divisão foi aplicada para maximizar a quantidade de dados disponíveis para treinamento, sem comprometer a validação (RAMACHANDRA; MONDAL; SETTURU, 2023).

Florestas Aleatórias (RF – *Random Forest*), Máquinas de Vetores de Suporte (SVM – *Support Vector Machines*), K-Vizinhos Mais Próximos (KNN – *K-Nearest Neighbors*) e Perceptron Multicamadas (MLP – *Multilayer Perceptron*). Cada um desses algoritmos foi escolhido com base em suas capacidades únicas para lidar com diferentes tipos de dados e desafios de classificação (CHAWLA *et al.*, 2002). Nota-se que todos os algoritmos em estudo são parametrizados (CHAWLA *et al.*, 2002), com exceção do KNN, que é não paramétrico. Algoritmos parametrizados possuem parâmetros ajustáveis que influenciam seu desempenho e precisam ser configurados para otimizar a precisão do modelo. Por outro lado, o algoritmo não paramétrico não possui parâmetros ajustáveis internos e utiliza regras fixas ou métodos baseados em distância para realizar previsões (ZHANG, C.; GUO; LU, 2017).

### **Random Forest (RF)**

Se destaca ao combinar múltiplas árvores de decisão, utilizando um processo chamado "bagging" (Bootstrap Aggregating) para criar um conjunto diversificado de árvores. Cada árvore é treinada com uma amostra aleatória dos dados e um subconjunto aleatório de atributos, o que não só melhora a precisão da classificação, mas também ajuda a evitar problemas de sobreajuste (overfitting). Isso torna o Random Forest ideal para conjuntos de dados complexos, onde a variabilidade dos dados pode levar a modelos menos robustos. A combinação das previsões de várias árvores permite obter uma estimativa mais confiável e estável da classe final (DURAN *et al.*, 2024; SCHONLAU, 2023).

**SVM (Support Vector Machine)** conhecido por sua eficácia em espaços de alta dimensionalidade, o SVM utiliza hiperplanos para separar as classes. O objetivo é encontrar o hiperplano que maximiza a margem entre as classes, o que proporciona uma

classificação precisa mesmo em conjuntos de dados complexos e não-lineares. Para lidar com problemas não-lineares, o SVM pode usar diferentes funções de kernel, como o kernel radial (RBF), que transforma os dados em um espaço de dimensionalidade mais alta para encontrar uma separação mais eficaz (CORTES; VAPNIK, 1995; SUYKENS; VANDEWALLE, 1999).

**KNN (K-Nearest Neighbors)** baseado no princípio da proximidade, o KNN classifica os dados com base na similaridade com seus vizinhos mais próximos. Para prever a classe de um novo ponto, o KNN calcula a distância entre o ponto e os pontos de treinamento e atribui a classe mais comum entre os K vizinhos mais próximos. Essa abordagem é intuitiva e eficaz para conjuntos de dados onde a estrutura das classes não é linear ou facilmente separável. No entanto, o desempenho do KNN pode ser afetado por dados desbalanceados ou pela escolha inadequada do valor de K, exigindo uma cuidadosa seleção de parâmetros (RAJA SAKTI ARIEF; SYAHRIL; SUHERMAN, 2023).

**MLP (Multi-Layer Perceptron)** conhecido como um tipo de rede neural artificial, o MLP é composto por múltiplas camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio em uma camada está conectado a todos os neurônios da camada seguinte, e as conexões têm pesos ajustáveis. O MLP é capaz de aprender representações complexas dos dados através do treinamento, onde as funções de ativação não-lineares permitem capturar relações mais sutis entre os atributos. Isso o torna especialmente útil em tarefas onde a natureza dos dados é altamente não-linear, como em problemas de reconhecimento de padrões e previsão complexa (ELANSARI; OUANAN; BOURRAY, 2024).

### 3.2.4.3 Avaliação de modelos de Machine Learning

Além de implementar os modelos, realizamos uma análise detalhada para entender como diferentes atributos e parâmetros influenciam as métricas de desempenho de cada algoritmo. Foram realizados diversos testes combinando atributos e classe da cobertura da terra para identificar aqueles que proporcionam as melhores métricas, utilizando validação cruzada. Validação Cruzada é uma técnica de avaliação em aprendizado de máquina que divide o conjunto de dados em vários subconjuntos, ou

"folds", para testar a generalização do modelo. O modelo é treinado  $K$  vezes, cada vez usando  $K-1$  folds para treinamento e o fold restante para teste. Os resultados são agregados para estimar o desempenho do modelo, proporcionando uma avaliação mais robusta e imparcial. Tipos comuns de validação cruzada incluem  $K$ -Fold, Leave-One-Out (LOO) e a estratificada, que garante uma distribuição proporcional das classes (MONTESINOS LÓPEZ; MONTESINOS LÓPEZ; CROSSA, 2022). Além disso, esse processo envolve a configuração de hiperparâmetros variados para os RF, SVM, KNN e RN, realizando uma busca exaustiva para identificar os parâmetros que resultam nas melhores métricas para cada algoritmo (NERSISYAN *et al.*, 2022). Nesse caso específico foi utilizado o Random Forest Classifier do Scikit-learn (BISONG, 2019).

Foram analisados diversos aspectos para cada modelo de ML que permite a escolha do melhor das melhores métricas. As métricas avaliadas incluem:

**Acurácia:** A porcentagem de previsões corretas realizadas pelo modelo em relação ao total de previsões feitas. É calculada (equação 6) como o número de previsões corretas dividido pelo número total de previsões (CHICCO; JURMAN, 2020).

$$Accuracy = \frac{(Number\ of\ correct\ predictions)}{(Total\ number\ of\ predictions)} \times 100\% \quad \text{Equação (6)}$$

**Precisão:** A proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo. Mede a exatidão das previsões positivas e é calculada (Equação 7) como o número de verdadeiros positivos dividido pelo número de verdadeiros positivos mais falsos positivos (CHICCO; JURMAN, 2020).

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad \text{Equação 7}$$

**Recall:** A proporção de instâncias positivas corretamente identificadas pelo modelo em relação ao total de instâncias positivas reais. Mede a capacidade do modelo de capturar todas as instâncias positivas e é calculado (Equação 8) como o número de verdadeiros positivos dividido pelo número de verdadeiros positivos mais falsos negativos (CHICCO; JURMAN, 2020; RICHARDSON *et al.*, 2024).

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \quad \text{Equação 8}$$

F1 Score: A média harmônica entre a precisão e o recall (Equação 9). É uma métrica que equilibra a precisão e o recall e é especialmente útil quando há uma desigualdade entre as classes (CHICCO; JURMAN, 2020; RICHARDSON *et al.*, 2024).

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{Equação 9}$$

Acurácia de Validação Cruzada: A média da acurácia obtida através de validação cruzada, onde o conjunto de dados é dividido em múltiplos subconjuntos (ou folds), e o modelo é treinado e testado em diferentes combinações desses subconjuntos. Fornece uma estimativa da performance do modelo em dados não vistos (MONTESINOS LÓPEZ; MONTESINOS LÓPEZ; CROSSA, 2022)

Desvio Padrão da Validação Cruzada: A medida da variação da acurácia obtida em diferentes iterações de validação cruzada. Indica a consistência da performance do modelo ao longo dos diferentes subconjuntos de dados (MONTESINOS LÓPEZ; MONTESINOS LÓPEZ; CROSSA, 2022).

Tempo de Execução: O tempo total necessário para treinar e testar o modelo. Inclui o tempo de computação para processar os dados e realizar previsões, indica o custo computacional, quanto mais demorado for para treinar um modelo maior é o custo computacional (GUIMARÃES *et al.*, 2023).

Melhores Parâmetros: Os parâmetros do modelo que proporcionaram o melhor desempenho durante o treinamento e a validação. Estes são ajustados para otimizar a performance do modelo (JAMALEDDYN; AYACHI; BINIZ, 2023).

ROC AUC: A Área Sob a Curva (AUC) da Curva Característica de Operação do Receptor (ROC – *Receiver Operating Characteristic*) mede a capacidade do modelo de distinguir entre as classes. Essa métrica varia de 0 a 1, onde 1 indica um modelo com desempenho perfeito, 0,5 representa um desempenho equivalente ao acaso, e valores

próximos de 0 indicam uma inversão completa na classificação, ou seja, o modelo está errando sistematicamente. (GOOGLE, 2024; RICHARDSON *et al.*, 2024).

Por outro lado, foram verificadas as questões centrais de aprendizado de máquina para a tomada de decisões na escolha dos modelos de aprendizado de máquina. O overfitting (sobreajuste) e o underfitting (subajuste) são aspectos críticos que impactam diretamente o desempenho e a generalização dos modelos (RAJA SAKTI ARIEF; SYAHRIL; SUHERMAN, 2023). O overfitting ocorre quando o modelo aprende não apenas os padrões relevantes dos dados de treinamento, mas também os ruídos, resultando em alta acurácia nesse conjunto, contudo, com desempenho insatisfatório em dados novos, ou seja, com baixa capacidade de generalização, (MONTESINOS LÓPEZ; MONTESINOS LÓPEZ; CROSSA, 2022). Em contraste, o underfitting se caracteriza pela incapacidade do modelo em captar padrões relevantes nos dados, sendo excessivamente simplista e levando a uma baixa acurácia tanto nos conjuntos de treinamento quanto nos de teste (MONTESINOS LÓPEZ; MONTESINOS LÓPEZ; CROSSA, 2022).

### **3.2.5 Análise Temporal das Classes de Cobertura da Terra e Validação do Campo**

#### **3.2.5.2 Série Temporal e Dinâmicas de Mudança**

Nossa abordagem para analisar as variações ao longo do tempo nas classes de cobertura da terra começou com a seleção do melhor algoritmo treinando com dados de 2019, ou seja, o que possui as melhores métricas, os quais extrapolamos para outros anos subsequentes. Isso nos permitiu mergulhar profundamente nas mudanças que ocorreram nas diferentes classes de cobertura da terra ao longo dos anos.

Ao discutir com os técnicos do Instituto da Biodiversidade e das Áreas Protegidas Guiné-Bissau (IBAP) e a equipe da direção do Parque de Tarrafe de Cacheu e membros importantes das comunidades locais, ganhamos conhecimentos sobre como essas classes evoluíram e os fatores principais que impulsionaram essas mudanças, baseado também em estudos relevantes sobre a área. Realizar trabalhos de campo proporciona

conhecimentos transversais sobre as dinâmicas e mudanças na cobertura da terra (SÁNCHEZ-AZOFEIFA *et al.*, 2003).

Para quantificar e analisar as mudanças na cobertura da terra ao longo da série temporal de 2019 a 2023, foram calculadas as variações temporais nas classes de cobertura, a taxa de variação anual média dessas classes e a variação percentual na cobertura da terra entre as porções Norte e Sul. As variações temporais foram determinadas pela diferença absoluta entre as áreas das classes em diferentes períodos, enquanto a taxa de variação considerou a proporção anual de mudança em relação à área inicial. Já a variação percentual entre Norte e Sul foi calculada com base na diferença percentual de Norte menos Sul. Essa abordagem proporcionou uma compreensão mais detalhada das dinâmicas de mudança na paisagem e sobre os padrões de uso e cobertura da terra tanto do bloco Norte como do bloco Sul do PNTC.

#### **3.2.5.2 Validação de Campo e Precisão dos Mapas**

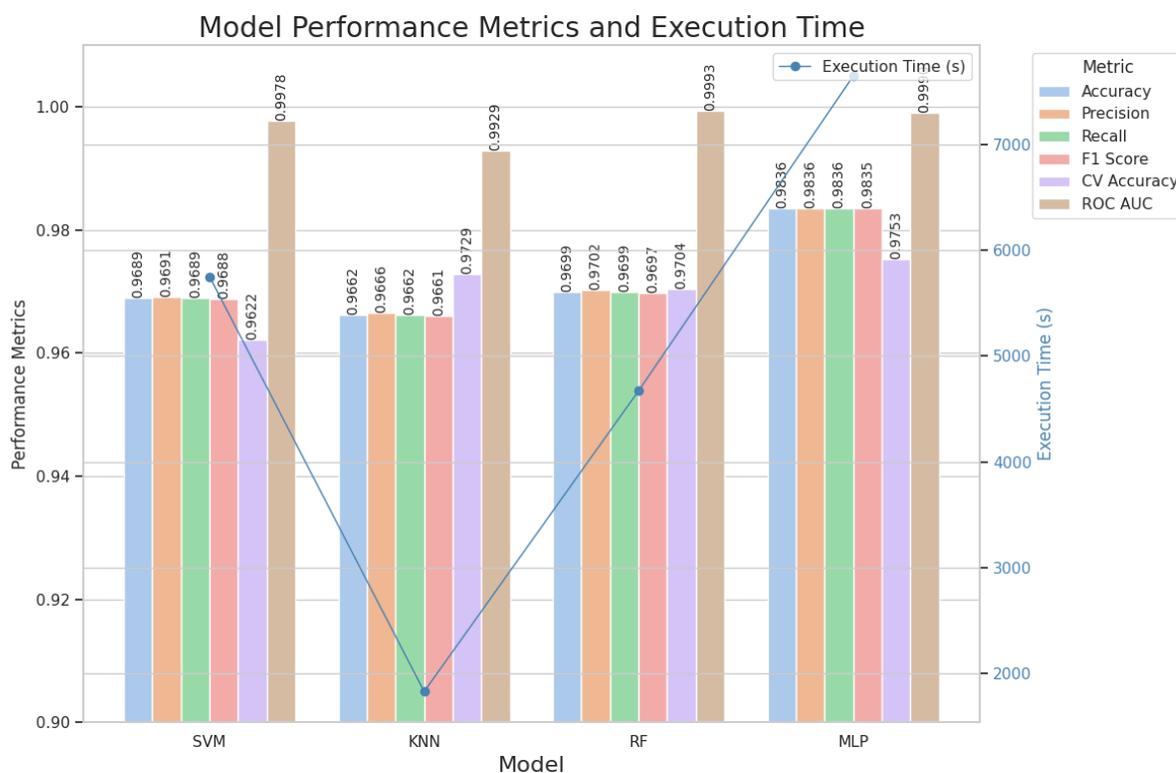
Entre os dias 27 de abril e 5 de maio de 2024, foi realizado trabalho de campo com foco na validação dos mapas temáticos gerados a partir de imagens de satélite, por meio da comparação dos padrões identificados nas imagens com observações diretas em campo. Foram visitados mais de 30 locais estrategicamente selecionados, abrangendo tanto áreas aquáticas quanto terrestres, para verificar as classes de cobertura da terra e confirmar a precisão dos dados da classificação do melhor algoritmo selecionado, utilizando como referência a verdade de campo, ou seja, pontos de controle visitados. Na validação foi aplicado o Índice Kappa, uma métrica amplamente reconhecida em estudos de sensoriamento remoto para medir o grau de concordância entre os mapas classificados e os dados de campo (CARVALHO *et al.*, 2024). A pontuação Kappa varia entre -1 e 1. Pontuações acima de 0,8 são geralmente consideradas boas concordância; zero ou inferior significa que não há concordância, ou seja, se for igual a zero, pode ser considerada como classificação aleatória (CARVALHO *et al.*, 2024). Complementando as análises, foram coletadas mais de 500 fotografias georreferenciadas, que posteriormente foram revisadas para aprimorar a interpretação das classes de uso e cobertura da terra. Essa abordagem integrada garantiu não apenas

a confiabilidade e precisão dos mapas temáticos, mas também uma compreensão mais aprofundada das dinâmicas de mudança na área de estudo.

## 4. RESULTADOS

### 4.1. Avaliação de Desempenho dos Algoritmos

A análise dos resultados dos algoritmos de aprendizado de máquina (SVM, KNN, RF e MLP) aplicados à classificação da cobertura do solo no PNTC destaca diferenças em desempenho, eficiência (Figura 5) e generalização (Figura 6), utilizando métricas como acurácia, precisão e ROC AUC (Tabela 3). Além disso, foram identificados padrões relevantes de variação temporal e espacial, contribuindo para entender a dinâmica ambiental e os desafios de manejo sustentável na região.



**Figura 5.** Desempenho de Quatro Modelos de ML (SVM, KNN, RF e MLP) Com 72 atributos

**Tabela 3.** Métricas Utilizadas para Avaliação dos Modelos: SVM, KNN, RF e MLP**Model Parameter Table**

Model	Parameter	Value
SVM	C	10
SVM	coef0	0.0
SVM	degree	2
SVM	gamma	scale
SVM	kernel	rbf
SVM	shrinking	True
KNN	algorithm	auto
KNN	leaf size	10
KNN	metric	manhattan
KNN	n_neighbors	3
KNN	p	1
KNN	weights	distance
RF	criterion	entropy
RF	max_depth	30
RF	max_features	sqrt
RF	max_samples	0.75
RF	min_samples_leaf	1
RF	min_samples_split	2
RF	n_estimators	500
MLP	activation	tanh
MLP	alpha	0.001
MLP	hidden_layer_sizes	(100,)
MLP	learning_rate	invscaling
MLP	learning_rate_init	0.01
MLP	solver	adam

Entre os modelos, o MLP apresenta o melhor desempenho em termos de acurácia, precisão, recall e F1-Score, todos com valores de 0.984, indicando excelente capacidade de classificação. Ele também tem a maior Acurácia na Validação Cruzada (0.975) e o valor mais elevado de ROC AUC (0.997), o que sugere um bom equilíbrio entre a sensibilidade e a especificidade. No entanto, o tempo de execução do MLP é o mais alto entre os modelos, com 7643.725 segundos, o que pode ser uma desvantagem em contextos em que o tempo de processamento é um fator crítico.

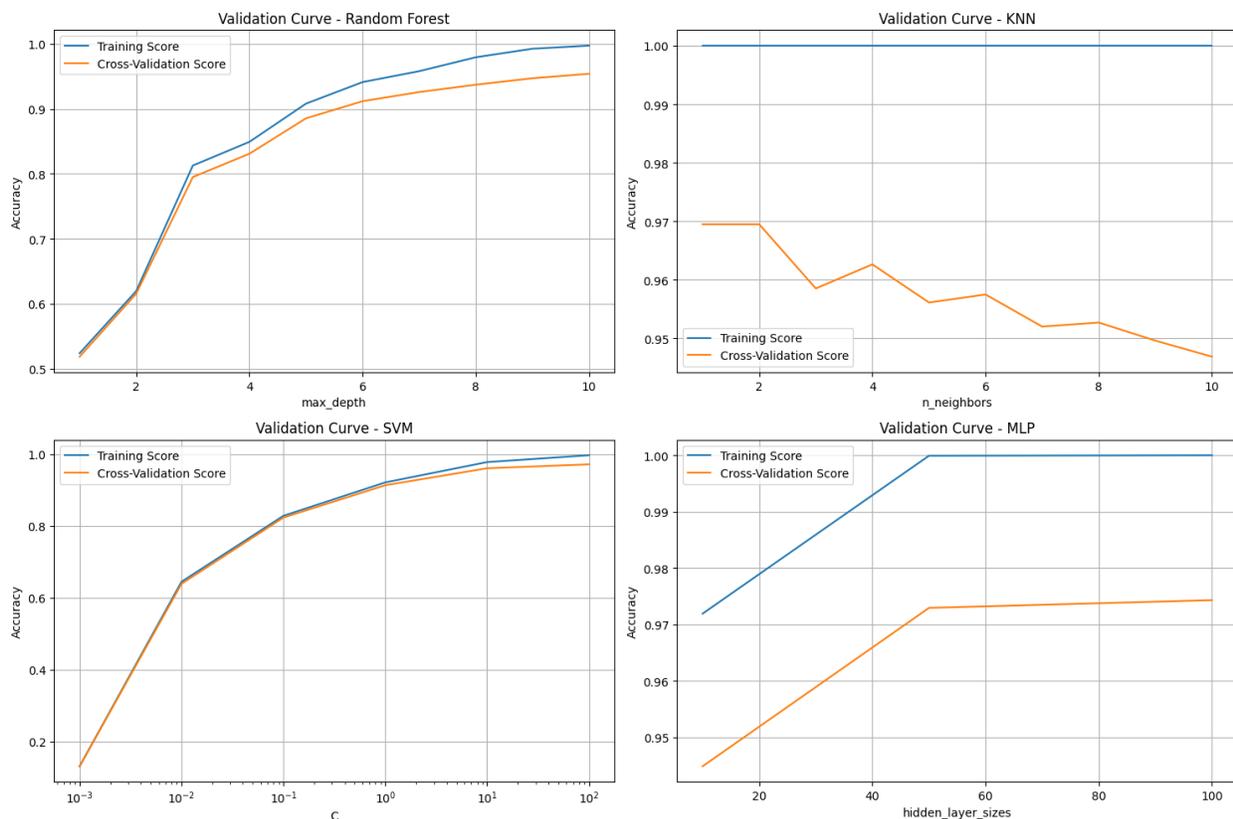
O Random Forest também apresenta excelente desempenho com uma acurácia de 0.970, precisão de 0.970, recall de 0.970 e uma acurácia da validação cruzada de 0.970, além de uma boa pontuação no ROC AUC (0.996). Seu tempo de execução é significativamente menor que o do MLP, mas ainda assim relativamente alto, com 4671.654 segundos.

O modelo SVM também se destaca, com métricas ligeiramente inferiores às do Random Forest, mas ainda muito competitivas. Sua acurácia é de 0.969, com precisão

e recall também em 0.969. Seu ROC AUC é ligeiramente superior, em 0.998, o que o coloca entre os modelos mais precisos na classificação de classes. Contudo, o tempo de execução do SVM é de 5748.419 segundos, que ainda é consideravelmente alto.

Por outro lado, o KNN apresenta uma acurácia levemente inferior (0.966), mas ainda competitiva. Suas métricas de precisão e recall também estão em 0.966, com um valor de ROC AUC de 0.993. Seu principal ponto positivo é o tempo de execução significativamente menor, 1829.507 segundos, o que pode torná-lo uma escolha viável em situações em que o tempo de processamento é mais crítico, ainda que o modelo tenha desempenho levemente inferior aos demais.

Apesar desses modelos apresentarem métricas elevadas, é importante ressaltar que esses valores não são as únicas variáveis a serem consideradas ao selecionar um modelo. Nesse caso, também usamos as curvas de validação (Figura 6), que fornecem uma visão mais aprofundada sobre o comportamento de cada modelo em relação a diferentes hiperparâmetros. A curva de validação ajuda a identificar não apenas o desempenho geral, mas também possíveis sinais de sobreajuste ou subajuste, permitindo uma avaliação mais completa da capacidade do modelo de generalizar bem para novos dados.



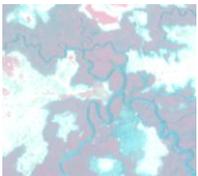
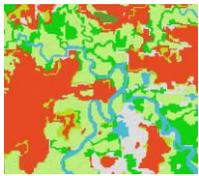
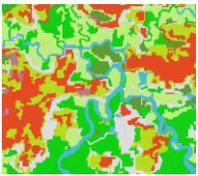
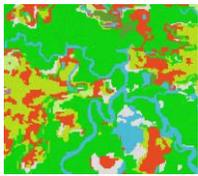
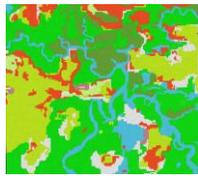
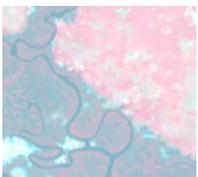
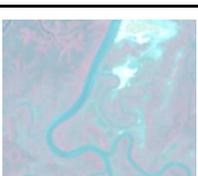
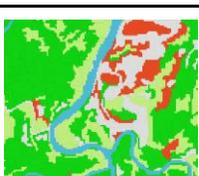
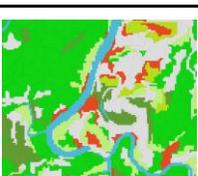
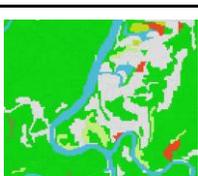
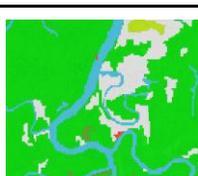
**Figura 6.** Curvas de Validação para Quatro Algoritmos de Aprendizado Supervisionado: RF, KNN, SVM e MLP.

Em cada um dos subgráficos, é possível observar a comparação entre o *Training Score* e o *Cross-Validation Score* em função de um hiperparâmetro específico de cada modelo. No caso do Random Forest, a acurácia no treinamento e na validação cruzada aumenta à medida que a profundidade máxima das árvores (*max\_depth*) aumenta até cerca de 8, onde a performance muda a taxa de crescimento. No entanto, a ligeira discrepância entre as duas curvas sugere um início de sobreajuste (*overfitting*) para profundidades maiores.

O gráfico de KNN indica que a acurácia de treinamento se mantém perfeita independentemente do número de vizinhos (*n\_neighbors*), mas a acurácia da validação cruzada diminui de forma significativa à medida que esse parâmetro aumenta, sugerindo que o modelo se ajusta bem aos dados de treinamento, mas apresenta dificuldades para generalizar, especialmente para valores maiores de *n\_neighbors*. Para o SVM, a curva

de validação exibe uma melhora tanto na acurácia do treinamento quanto na validação conforme o parâmetro de regularização ( $C$ ) aumenta, estabilizando-se em torno de  $C \approx 10^0$ . A proximidade entre as curvas indica que o modelo se generaliza bem.

Por fim, o gráfico do MLP mostra que a acurácia no treinamento atinge 100% rapidamente à medida que o número de neurônios na camada oculta (*hidden\_layer\_sizes*) aumenta, enquanto a acurácia na validação cruzada continua subindo de forma mais gradual. Isso sugere que, embora o modelo se ajuste perfeitamente aos dados de treinamento, ainda há uma leve discrepância em termos de generalização. Em resumo, os modelos analisados apresentam diferentes comportamentos em termos de generalização, com Random Forest e SVM exibindo um bom equilíbrio entre as métricas de treinamento e validação, enquanto KNN e MLP mostram tendência ao sobreajuste com certos valores de hiperparâmetros. Para efeito da análise o SVM coloca-se em vantagem na generalização.

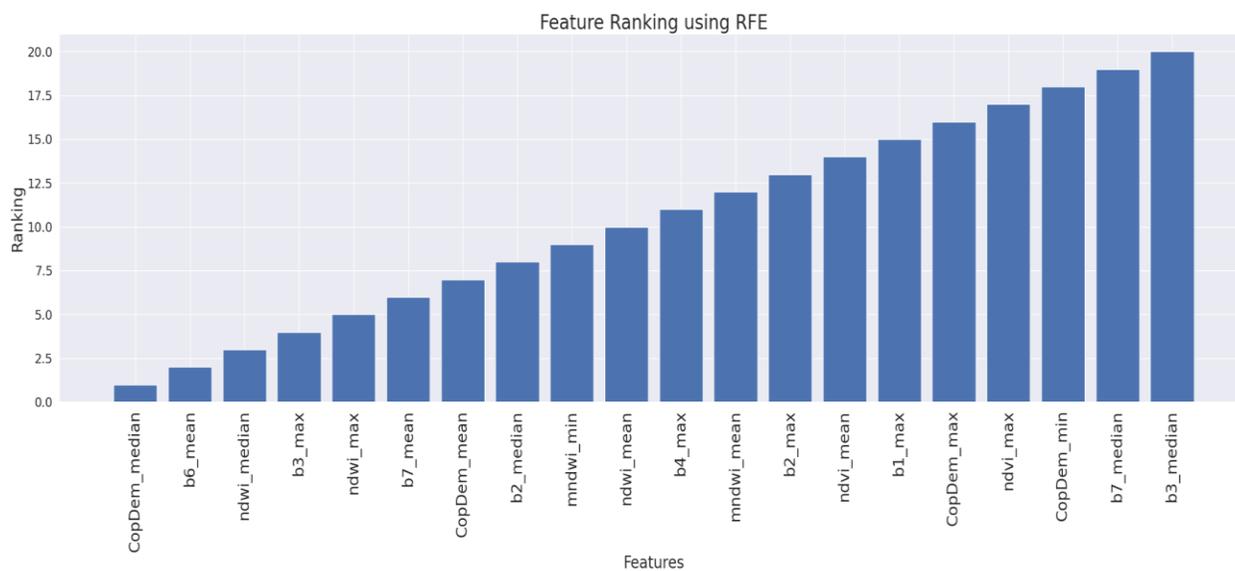
Area composição colorida	Segmentação da área	RF	KNN	SVM	MLP		
							
							
							
							
Hypersaline surface	uncovered area	mangrove	other vegetation	River	dense vegetation	agriculture	built up area

**Figura 7.** Comparação de Visual de Modelos de Classificação: RF, KNN, SVM e MLP

Na análise visual (Figura 7) da classificação dos algoritmos aplicados à cobertura do solo, observa-se variações na precisão e na qualidade da classificação para as diferentes categorias analisadas. O SVM apresentou uma distribuição mais uniforme das classes, evidenciando uma menor presença de ruído em relação aos outros algoritmos, especialmente quando comparado ao KNN, que apresenta maior granularidade em áreas homogêneas, o que pode indicar instabilidade na definição das fronteiras entre as classes. Por outro lado, o RF, apesar de produzir métricas robustas no treinamento, visualmente verifica-se ligeira sensibilidade em confundir área descoberta com as outras classes. O MLP, embora apresente resultados visualmente bons, ainda apresenta leve mistura entre classes, especialmente em regiões com características espectrais semelhantes, o que pode ser associado à variabilidade das classes de cobertura da terra em uma área pequena. Resultados visuais indicam que o SVM possui maior robustez na classificação de áreas complexas, provavelmente devido à sua capacidade de maximizar margens entre classes, minimizando erros de classificação em áreas críticas.

## 4.2 Impacto dos Atributos na Classificação

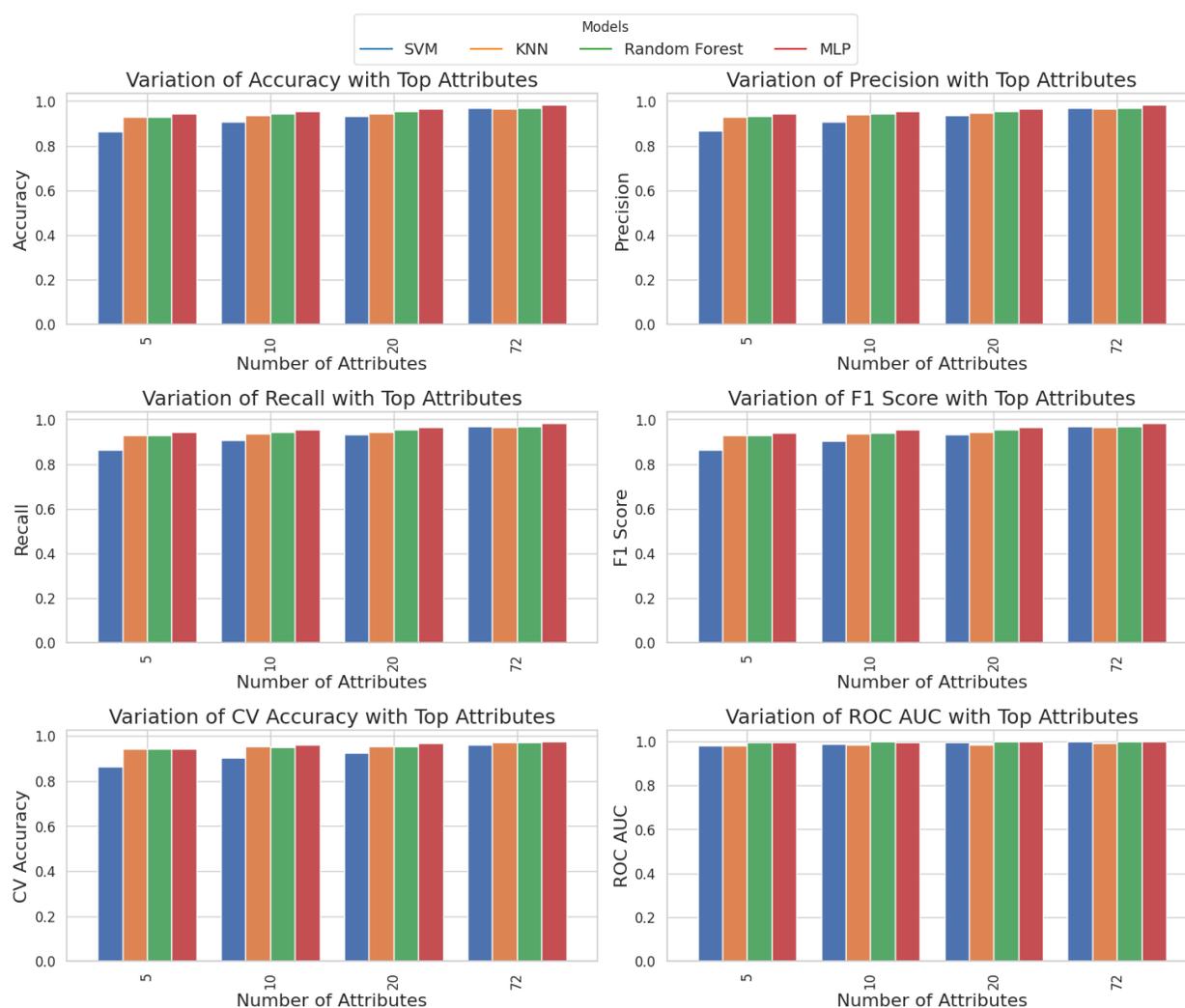
O gráfico (Figura 8) exibe o ranking dos 20 atributos mais importantes obtidos por meio do método Recursive Feature Elimination (RFE).



**Figura 8.** Ranking de Importância das Variáveis Usando RFE

Nesse contexto, as variáveis com menores posições apresentam maior relevância para o modelo. As variáveis localizadas à esquerda, como Mediana do Copernicus Dem, média do b6 e mediana do NDWI, destacam-se como as mais importantes, enquanto aquelas à direita, como mediana do b3 e mediana do b7, apresentam menor influência no desempenho do modelo

Esses resultados sugerem que os atributos mais relevantes contêm informações essenciais para as métricas do modelo, enquanto as menos importantes podem ser candidatas à exclusão, visando reduzir a complexidade sem comprometer a qualidade preditiva. Para validar essa hipótese, realizamos testes com diferentes conjuntos de atributos, contendo 5, 10, 20 e 72, a fim de avaliar o comportamento das métricas em função do número de variáveis selecionadas para cada modelo, como se vê na Figura 9.



**Figura 9.** *Variação das Métricas de Desempenho com o Número de Atributos nos Modelos de ML (SVM, KNN, RF e MLP)*

Logo de cara, na Figura 9, nota-se que, à medida que o número de atributos aumenta, todos os modelos exibem melhorias em quase todas as métricas, o que sugere que a inclusão de atributos adicionais aprimora a capacidade preditiva e discriminativa dos modelos.

Especificamente, o gráfico de accuracy mostra um aumento geral para todos os modelos com o acréscimo de atributos, sendo o MLP o modelo com maior precisão, seguido de pelo Random Forest. O SVM, que inicialmente apresenta uma precisão inferior, melhora consideravelmente à medida que mais atributos são adicionados, alcançando um desempenho competitivo. Na precisão, observa-se uma tendência similar, com o MLP mantendo o melhor desempenho. O SVM, novamente, inicia em um patamar mais baixo, mas melhora gradualmente, indicando uma dificuldade inicial em minimizar falsos positivos. O MLP e o Random Forest, por sua vez, já apresentam valores altos de precisão mesmo com poucos atributos, sugerindo maior robustez na detecção correta das classes.

O gráfico de recall reforça essa tendência de melhoria com o aumento de atributos, com o MLP liderando o desempenho, seguido pelo Random Forest. A sensibilidade do SVM aumenta de forma mais acentuada conforme o número de atributos cresce, especialmente com 72 atributos, o que indica uma redução dos falsos negativos e um aumento de seu recall. Já a pontuação F1 Score, que é uma média harmônica entre precisão accuracy e recall, reflete uma tendência similar: o MLP continua a exibir o melhor desempenho, mostrando que mantém um equilíbrio ótimo entre precisão e sensibilidade. O SVM, apesar de começar com um desempenho inferior, exibe uma melhoria consistente conforme mais atributos são introduzidos.

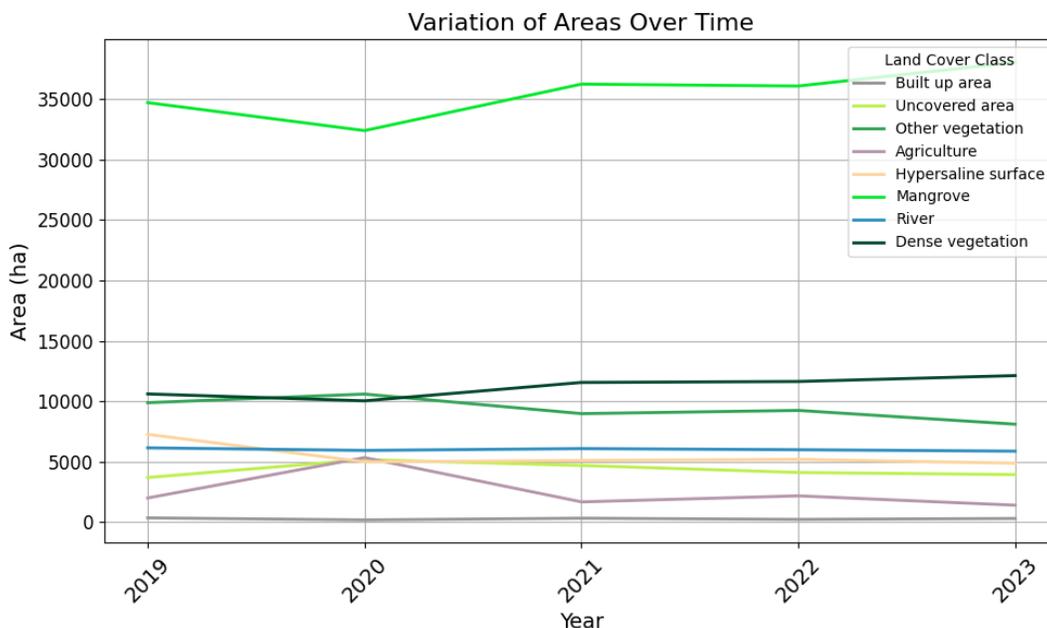
Na validação cruzada, a precisão (CV Accuracy) segue uma tendência onde todos começam em 94, com exceção de SVM. Verificou-se quase a mesma estabilidade entre KNN e Random Forest. O MLP mostra a maior estabilidade e precisão conforme o número de atributos aumenta. Esse resultado reforça a robustez desses modelos em

diferentes subconjuntos de dados. No caso da área sob a curva ROC, que é uma métrica importante para avaliar a capacidade de discriminação entre classes, o Random Forest atinge quase o valor máximo de 1, seguido pelo MLP, sugerindo um desempenho excelente na separação das classes. A curva do SVM melhora com o aumento de atributos, enquanto o KNN apresenta uma variação não linear ao redor dos 20 atributos, mas se estabiliza com 72 atributos.

De forma geral, pode-se verificar que todas as métricas estão acima de 0.86, tendo como máximo 0.99 o que indica bons resultados. O SVM, embora inicialmente apresente desempenho inferior em comparação com outros modelos, exibe um crescimento substancial, indicando que se beneficia amplamente do aumento de atributos. O KNN, por outro lado, demonstra menor estabilidade na métrica ROC AUC, o que pode sugerir uma sensibilidade a ruídos ou uma limitação em sua capacidade discriminativa com menor quantidade de atributos.

#### 4.5. Variações Temporais e Validação de Mapa

O gráfico da Figura 10 mostra a variação das áreas ocupadas por diferentes classes de cobertura do solo entre 2019 e 2023.

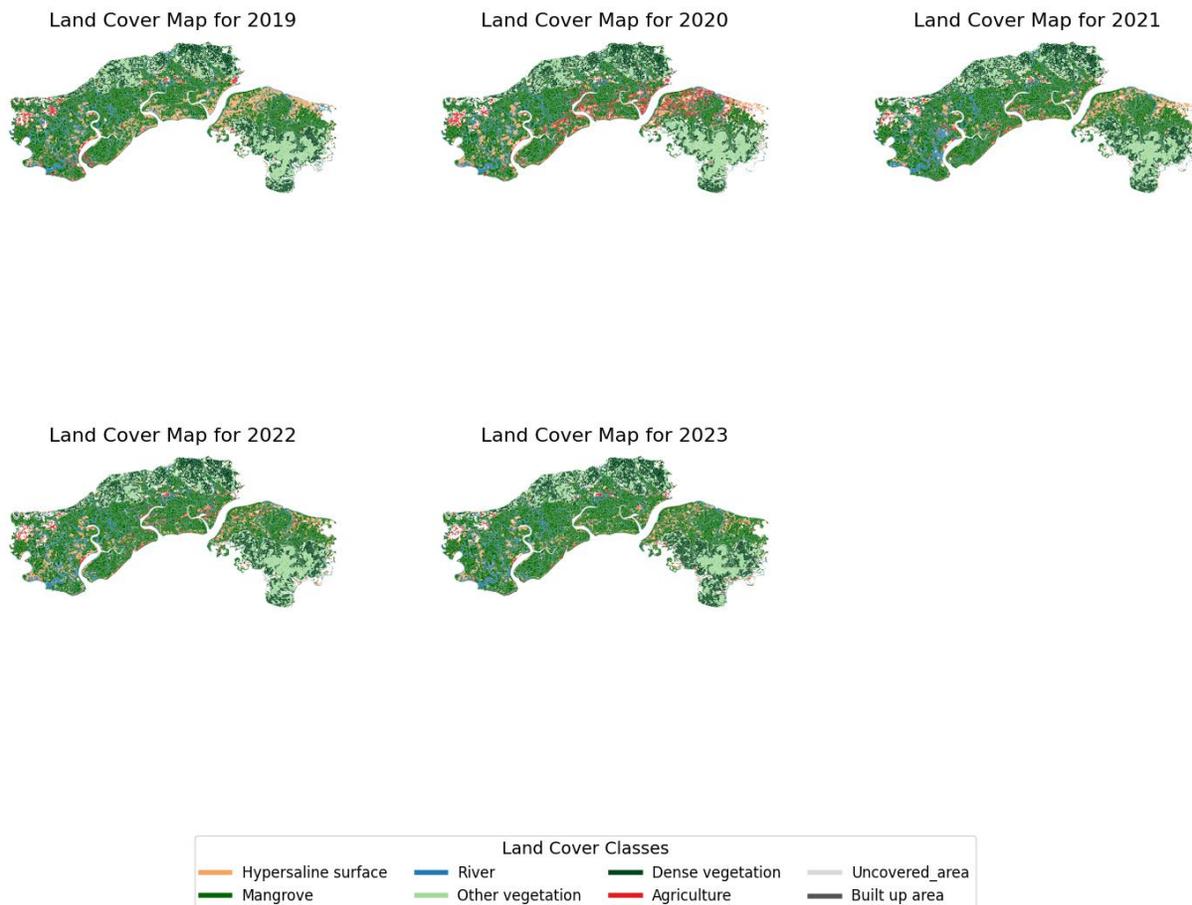


**Figura 10.** *Varição das Áreas Ocupadas por Diferentes Classes de Cobertura do Solo entre 2019 e 2023*

A vegetação densa e outras vegetações mantêm-se estável em torno de 10.000 hectares. A vegetação densa apresenta leve decréscimo em 2020 e leve aumento em 2021, já outras vegetações mostram flutuações, com redução gradual após uma leve baixa em 2021. A classe mangue, com maior extensão, apresenta leve tendência de declínio em 2020, e com leve aumento nos anos subsequentes, o declínio reflete as alterações ambientais, embora exibam leves fluxos de conversão para áreas agrícolas. Essas transições podem estar associadas à retirada da floresta de mangue para a prática do cultivo de arroz em áreas anteriormente ocupadas por manguezais. Além disso, observa-se a presença de ruídos que podem ser atribuídos a possíveis erros de classificação. A agricultura varia moderadamente, atingindo um pico em 2020 e declinando até 2023, sendo possível atribuir a resposta a ciclos sazonais de plantio de arroz e a resposta espectral semelhante com a superfície hipersalina. A classe superfície hipersalina mostra leve oscilação, com cerca de 5.000 hectares, mostram um decréscimo sutil ao longo do período analisado. No entanto, pode ser parcialmente atribuída a erros de classificação. A classe Rio mantém-se estável, sem variações perceptíveis, enquanto a Categoria área construída é praticamente nula, indicando baixa urbanização em todo o parque o que foi muito perceptível no trabalho de campo. A área descoberta apresenta pequenas variações, com um pico em 2020 e leve redução posterior, essa classe é associada a pastagem de gado e cultivo de arroz na área não úmida, ou seja, arroz de panpan que tem ciclos sazonais de anos para o abandono do campo.

As oscilações observadas sugerem eficácia nas práticas de manejo, garantindo estabilidade para quase todas as classes de cobertura do solo, sendo que pequenas flutuações podem resultar de erros sutis de classificação.

### Land Cover Changes (2019-2023)

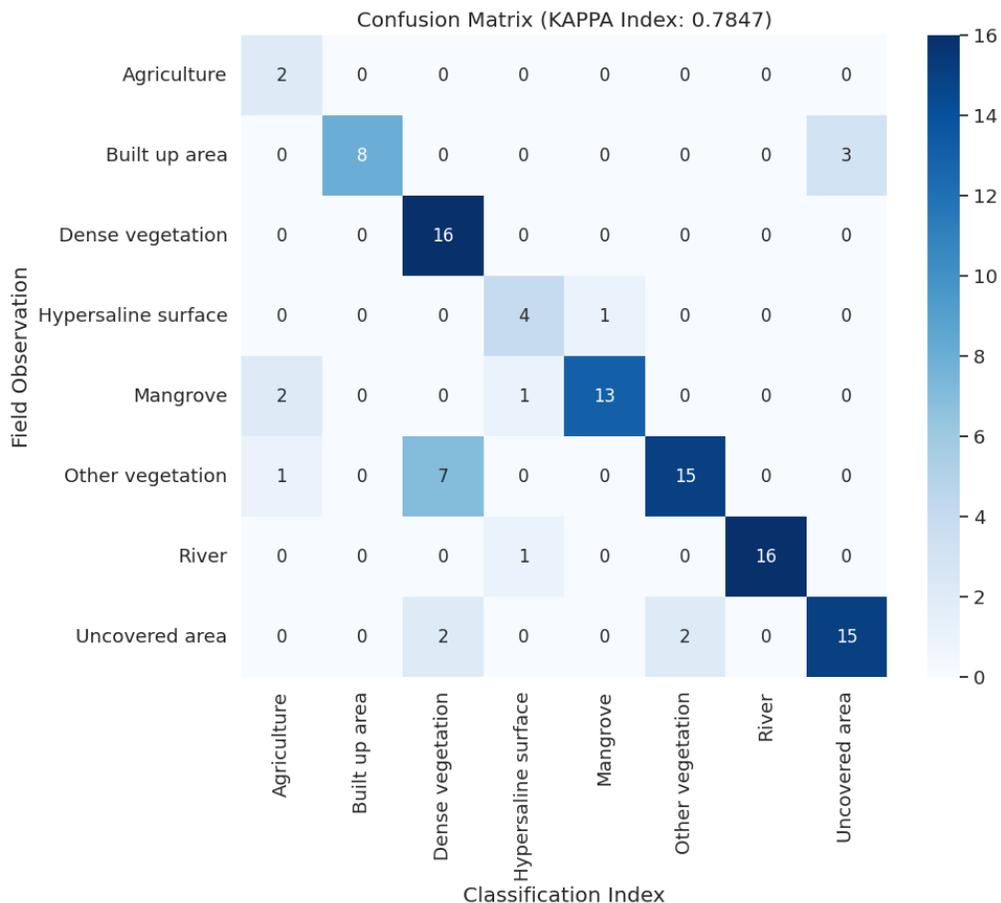


**Figura 11.** *Análise Espacial das Mudanças na Cobertura da Terra (2019-2023)*

A classificação apresentada na Figura 11 ilustra as mudanças espaciais nas classes de cobertura do solo entre 2019 e 2023. Observa-se uma divisão espacial marcante entre as porções norte e sul do PNTC.

A porção norte caracteriza-se pela predominância de vegetação natural, como vegetação densa e manguezais, com mudanças mais sutis ao longo do período, sugerindo menor impacto humano direto. Já a porção sul apresenta variações significativas associadas à expansão de outras vegetações, ou seja, plantio de caju, com aumento de áreas construídas e redução de cobertura vegetal densa e uma alta incidência de superfícies hipersalinas. Essas dinâmicas refletem interações complexas

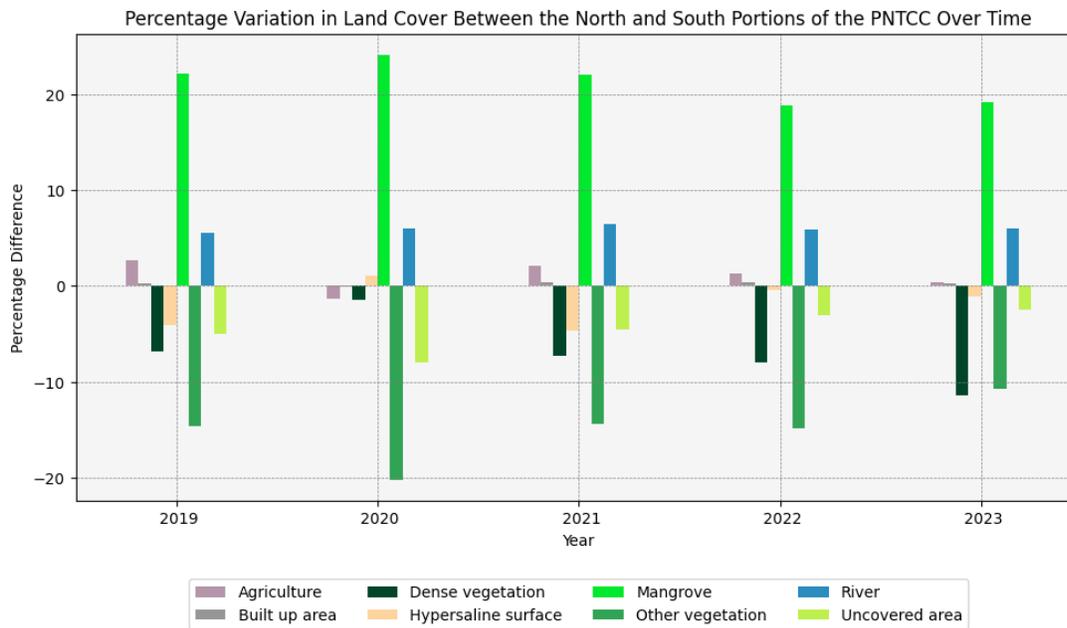
entre fatores ambientais e antrópicos, destacando a necessidade de estratégias de planejamento territorial e conservação ambiental para equilibrar o desenvolvimento humano e a preservação ecológica.



**Figura 12.** Matriz de Confusão na Validação do Mapa de Cobertura da Terra.

Na validação do mapa de cobertura da terra, como mostra a Figura 12, alcançou-se o índice Kappa de 0,7847, indicando uma concordância razoável entre o mapa classificado e os levantamentos de campo realizados. A matriz de confusão mostra que classes como Vegetação Densa e Rio foram bem classificadas, enquanto erros ligeiros ocorreram em Área descoberta e Outra Vegetação. Confere-se que a classificação de cobertura da terra teve uma boa consonância com dados que verificamos em campo.

A variação percentual na cobertura do solo entre as porções Norte e Sul, ao longo do período de 2019 a 2023 (Figura 13).



**Figura 13.** *Variação Percentual da Cobertura do Solo entre as Porções Norte e Sul ao Longo do Tempo (2019–2023).*

A análise foi baseada na diferença percentual entre as áreas classificadas nas porções Norte e Sul da região de estudo, considerando cada classe de uso e cobertura do solo ao longo da série temporal. As classes de mangue e outra vegetação apresentaram as maiores discrepâncias espaciais. Observou-se uma predominância consistente de manguezais no Norte e de outra vegetação no Sul durante todo o período analisado. Além disso, a porção Sul mostrou maior presença de áreas descobertas em comparação ao Norte. Essas variações sugerem que os padrões de transformação da paisagem diferem entre as duas regiões, possivelmente refletindo influências ambientais e atividades humanas distintas.

## 5. DISCUSSÃO

A aplicação de algoritmos de ML, como SVM e MLP, RF e KNN desempenha um papel crucial no monitoramento de ecossistemas sensíveis, como os manguezais do Parque Natural dos Tarrafes de Cacheu. Estudos recentes demonstram que as técnicas de ML superam os métodos tradicionais ao capturar com eficácia padrões complexos e relações não lineares em dados ambientais (SHAIK; SHAIK; PRIYA, 2024). O nosso

estudo corrobora com estudos recentes que mostram a capacidade de SVM e RF em lidar com alta dimensionalidade dos dados e mitigar o sobreajuste (MAHVASH MOHAMMADI; HEZARKHANI, 2020; RAJA SAKTI ARIEF; SYAHRIL; SUHERMAN, 2023) tornando-os uma escolha atraente para a classificação de zonas úmidas e áreas com alta complexidade (CHERIAN; K, 2024). Também pode-se verificar que o MLP é muito suscetível a sobreajuste por causa da sua grande complexidade e exige maior volume de dados para potencializar a sua capacidade de generalizar quando submetido aos dados não visto no treinamento, como aponta (ZHU, Z., 2024). Outrossim podemos constatar que o MLP às vezes pode apresentar boa performance das métricas de avaliação, mas com uma capacidade insatisfatória de generalização, o que deve ser levado em consideração na sua escolha (ZHU, Z., 2024). O mesmo pode ser constatado para o KNN devido a sua simplicidade que torna difícil de lidar com dados de alta dimensionalidade (RAO, 2023).

Um aspecto crucial do estudo foi o uso de técnicas avançadas de seleção de variáveis, como o Recursive Feature Elimination (RFE), para reduzir redundâncias e aumentar a eficiência computacional. A eficácia do RFE foi amplamente testada em diferentes contextos (PRATAMA; LATIPAH; SARI, 2022; RAMEZAN, 2022). Por exemplo, autores como (RAMEZAN, 2022) avaliaram a transferibilidade de conjuntos de características derivados do RFE utilizando imagens multiespectrais do Sentinel-2A para a classificação de uso e cobertura do solo. Descobriu-se que os conjuntos de características otimizados pelo RFE apresentam melhor desempenho dentro do mesmo conjunto de dados e podem ser transferidos, com diferentes graus de precisão, para outros conjuntos de dados com características remotamente semelhantes. Esses resultados sugerem que, em tarefas de monitoramento ambiental, o RFE pode ser empregado para selecionar conjuntos de características robustos e potencialmente transferíveis para outras regiões com condições ambientais similares, embora ajustes possam ser necessários para otimizar a precisão.

Além disso, o uso do SMOTE (Synthetic Minority Over-sampling Technique) para equilibrar as classes em nosso estudo demonstrou melhorias significativas nas métricas de classificação, corroborando pesquisas na área de sensoriamento remoto (FONSECA;

DOUZAS; BACAO, 2021; NGO *et al.*, 2022). Esse método também se alinha a estudos na área da saúde (KOSOLWATTANA *et al.*, 2023), que apontam que amostras sintéticas, embora eficazes, podem apresentar limitações, como ambiguidade, baixa qualidade e dificuldade de separação em relação à classe majoritária.

A seleção de atributos utilizando o RFE também evidenciou impacto positivo na precisão de modelos de machine learning (ML), como apontado em estudos de revisão de literatura (PRIYATNO; WIDIYANINGTYAS, 2024). O nosso estudo alcançou uma acurácia mínima de 86% com apenas cinco atributos, resultado considerado aceitável para aplicações de sensoriamento remoto em áreas úmidas. Esses achados corroboram pesquisas (DAS *et al.*, 2022) que demonstram que a seleção e a redução de atributos em sensoriamento remoto não apenas aumentam a precisão das métricas, mas também reduzem os custos computacionais associados. Já na validação do mapa, nosso estudo obteve um índice Kappa de 0,78, considerado aceitável. Em comparação, o estudo conduzido por Ramachandra *et al.*, (2023), apresentou um índice Kappa 0,66 para SVM e 0,85 para RF (RAMACHANDRA; MONDAL; SETTURU, 2023).

O sensoriamento remoto aplicado a machine learning (ML) de baixo custo econômico apresenta grande relevância para o monitoramento e a conservação ambiental em áreas remotas com recursos limitados (NGO *et al.*, 2022). Conforme evidenciado por Fonseca *et al.*, (2022), cerca de 70% dos parques da Guiné-Bissau carecem de planos de manejo e apresentam um número reduzido de funcionários. O PNTC, devido à sua diversidade ecológica e à vasta extensão de florestas de mangue, representa interesses diversos em relação ao uso da terra, como apontado pelos estudos IBAP (2018) e Junior *et al.* (2019). Entre 2019 e 2023, observou-se que as dinâmicas ecológicas da região foram significativamente influenciadas por práticas humanas e alterações ambientais. Apesar de os manguezais demonstrarem certa estabilidade, com flutuações que sugerem esforços de conservação nos anos analisados, o estudo de (JUNIOR *et al.*, 2018) indica que, nos anos 2000, a área total de manguezais era de aproximadamente 50.000 hectares. Ainda segundo o referido estudo, os manguezais na Zona Sul apresentam maior conservação em comparação à Zona Norte, o que corrobora as observações realizadas durante o trabalho de campo.

Os dados deste estudo sugerem que a porção Sul apresenta maior atividade antrópica, com predominância de classes associadas às atividades humanas, em contraste com a porção Norte. Essa conclusão diverge parcialmente do estudo de Fonseca *et al.*, (2022), que aponta maior concentração de plantações de caju na porção Norte. No entanto, quando analisadas as classes de cobertura proporcionalmente, os resultados divergem dos apresentados no referido estudo. Essas dinâmicas estão alinhadas com pesquisas sobre mudanças no uso da terra em áreas protegidas da Guiné-Bissau, que destacam a crescente pressão de atividades antrópicas em diferentes níveis como o cultivo de caju, pesca ilegal, caça e práticas agrícolas (BERSACOLA *et al.*, 2018; HOCKINGS; SOUSA, 2013; JUNIOR *et al.*, 2018; 2019).

## 5. CONCLUSÃO

Este estudo evidenciou o potencial dos algoritmos de Machine Learning na classificação da cobertura da terra, utilizando imagens Landsat 8 no Parque Natural dos Tarrafes de Cacheu (PNTC). Reforçou-se a necessidade de métodos modernos para o monitoramento ambiental em regiões com recursos financeiros limitados. Os resultados demonstraram que técnicas de ML, como SVM, RF, KNN e MLP, apresentam desempenho elevado, com todas as métricas avaliadas acima de 86% utilizando apenas cinco atributos, comprovando sua eficácia na identificação de padrões complexos em ecossistemas dinâmicos como os manguezais, com baixo custo computacional e esforço reduzido.

A análise temporal apontou uma relativa estabilidade dos manguezais, mesmo diante das pressões antrópicas, ressaltando a necessidade de esforços contínuos de conservação. Além disso, a seleção criteriosa de atributos e a validação dos mapas garantiram resultados confiáveis e aplicáveis ao manejo sustentável do PNTC, contribuindo significativamente para o desenvolvimento de políticas de gestão ambiental baseadas em evidências.

Por fim, este trabalho reafirma a viabilidade do uso de algoritmos de código aberto e dados acessíveis, oferecendo uma metodologia replicável para outras áreas costeiras da

África Ocidental, promovendo o equilíbrio entre a preservação ambiental e o desenvolvimento socioeconômico.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

A V, S. *et al.* **Automated Forest Types Classification using Ensemble Machine Learning in Hyperspectral Satellite Images**. Em: *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*. [s.l.]: IEEE, 2024. ISBN: 979-8-3503-6066-0, DOI: 10.1109/IACIS61494.2024.10721870.

ANEME. **Enquadramento Perspetivas De Desenvolvimento Levantamento E Caracterização Das Empresas Comerciais E Industriais, Guiné-Bissau**. Lisboa: [s.n.], 2018. Disponível em: <[https://www.aneme.pt/site/wp-content/uploads/2018/07/ESTUDO\\_guineBissau-1.pdf](https://www.aneme.pt/site/wp-content/uploads/2018/07/ESTUDO_guineBissau-1.pdf)>. Acesso em: 26/nov./24.

BAATZ, M.; SCHÄPE, A. **Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation**. [s.l.]: [s.n.], [s.d.]. Disponível em: <<https://www.semanticscholar.org/paper/Multiresolution-Segmentation-:-an-optimization-for-Baatz-Sch%C3%A4pe/364cc1ff514a2e11d21a101dc072575e5487d17e>>. Acesso em: 13/dez./2000.

BASUKALA, A. K. *et al.* **Towards improved land use mapping of irrigated croplands: performance assessment of different image classification algorithms and approaches**. *European Journal of Remote Sensing*, [s.l.], v. 50, nº 1, p. 187–201, 2017. ISSN: 2279-7254, DOI: 10.1080/22797254.2017.1308235.

BENZ, U. C. *et al.* **Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information**. *ISPRS Journal of Photogrammetry and Remote Sensing*, [s.l.], v. 58, nº 3–4, p. 239–258, 2004. ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2003.10.002.

BERSACOLA, E. *et al.* **Primate occurrence across a human-impacted landscape in Guinea-Bissau and neighbouring regions in West Africa: using a systematic**

**literature review to highlight the next conservation steps.** *PeerJ*, [s.l.], v. 6, p. e4847, 2018. ISSN: 2167-8359, DOI: 10.7717/peerj.4847.

BISONG, E. **Introduction to Scikit-learn.** *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA: Apress, 2019. p. 215–229. DOI: 10.1007/978-1-4842-4470-8\_18.

BISSAU-GUINÉ. **Plano de Gestão Parque Natural dos Tarrafes do Rio Cacheu-PNTC.** Bissau: [s.n.], 2008. Disponível em: <<https://ibapgbissau.org/wp-content/uploads/2022/06/Plano-de-Gestao-de-PNTC.pdf>>. Acesso em: 26/nov./24.

BLASCHKE, T. **Object based image analysis for remote sensing.** *ISPRS Journal of Photogrammetry and Remote Sensing*, [s.l.], v. 65, nº 1, p. 2–16, 2010. ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2009.06.004.

BRUBACHER, J. P.; GUASSELLI, L. A.; OLIVEIRA, G. G. De. **Delimitação de áreas inundáveis a partir de Modified Normalized Difference Water Index (MNDWI) no Município de Esteio (RS, Brasil).** *Pesquisas em Geociências*, [s.l.], v. 44, nº 2, p. 367, 2017. ISSN: 1807-9806, DOI: 10.22456/1807-9806.78279.

CALAZANS, M. A. A. *et al.* **Machine Learning and Graph Signal Processing Applied to Healthcare: A Review.** *Bioengineering*, [s.l.], v. 11, nº 7, p. 671, 2024. ISSN: 2306-5354, DOI: 10.3390/bioengineering11070671.

CARVALHO, L. F. De *et al.* **Using of high spatial resolution images to evaluate the thematic accuracy of land use and occupation maps with the Kappa index.** *MOJ Ecology & Environmental Sciences*, [s.l.], v. 9, nº 5, p. 220–228, 2024. ISSN: 25732919, DOI: 10.15406/mojes.2024.09.00329.

CHAUHAN, K. *et al.* **Feature Extraction from Image Sensing (Remote): Image Segmentation.** Em: *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. [s.l.]: IEEE, 2023. ISBN: 979-8-3503-3086-1, DOI: 10.1109/ICAC3N60023.2023.10541413.

CHAWLA, N. V. *et al.* **SMOTE: Synthetic Minority Over-sampling Technique.** *Journal of Artificial Intelligence Research*, [s.l.], v. 16, p. 321–357, 2002. ISSN: 1076-9757, DOI: 10.1613/jair.953.

CHERIAN, S. M.; K, R. **Random forest and support vector machine classifiers for coastal wetland characterization using the combination of features derived from optical data and synthetic aperture radar dataset.** *Journal of Water and Climate Change*, [s.l.], v. 15, nº 1, p. 29–49, 2024. ISSN: 2040-2244, DOI: 10.2166/wcc.2023.238.

CHICCO, D.; JURMAN, G. **The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.** *BMC Genomics*, [s.l.], v. 21, nº 1, p. 6, 2020. ISSN: 1471-2164, DOI: 10.1186/s12864-019-6413-7.

CINCO-CASTRO, S.; HERRERA-SILVEIRA, J. **Vulnerability of mangrove ecosystems to climate change effects: The case of the Yucatan Peninsula.** *Ocean & Coastal Management*, [s.l.], v. 192, p. 105196, 2020. ISSN: 09645691, DOI: 10.1016/j.ocecoaman.2020.105196.

CLEWLEY, D. *et al.* **A Python-Based Open Source System for Geographic Object-Based Image Analysis (GEOBIA) Utilizing Raster Attribute Tables.** *Remote Sensing*, [s.l.], v. 6, nº 7, p. 6111–6135, 2014. ISSN: 2072-4292, DOI: 10.3390/rs6076111.

COLARES, I. V. V. *et al.* **Aplicação de GEOBIA para classificação da cobertura da terra no Parque Nacional do Itatiaia.** *Revista Brasileira de Geomática*, [s.l.], v. 3, nº 1, p. 19, 2015. ISSN: 2317-4285, DOI: 10.3895/rbgeo.v3n1.5456.

CONTO, D. De *et al.* **Comparação entre abordagens orientada a objetos geográficos e pixel a pixel para classificação supervisionada de imagens MSI / Sentinel-2.** *Revista de Geociências do Nordeste*, [s.l.], v. 10, nº 1, p. 426–440, 2024. ISSN: 2447-3359, DOI: 10.21680/2447-3359.2024v10n1ID33713.

CORTES, C.; VAPNIK, V. **Support-vector networks.** *Machine Learning*, [s.l.], v. 20, nº 3, p. 273–297, 1995. ISSN: 0885-6125, DOI: 10.1007/BF00994018.

DAS, J. K.; N, V. H. **Reducing Dimension for Remote Sensing Image Dataset**. Em: *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. [s.l.]: IEEE, 2022. ISBN: 978-1-6654-0816-5, DOI: 10.1109/ICACCS54159.2022.9785203.

DINIZ, C. *et al.* **Brazilian Mangrove Status: Three Decades of Satellite Data Analysis**. *Remote Sensing*, [s.l.], v. 11, n° 7, p. 808, 2019. ISSN: 2072-4292, DOI: 10.3390/rs11070808.

DODMAN, T.; SÁ, J. **Monitorização de aves aquáticas no Arquipélago dos Bijagós, Guiné-Bissau**. Bissau: [s.n.], 2005. Disponível em: <<https://didinho.org/Arquivo/MONITORIZACAODEAVESAQUATICASNOSBIJAGOS.pdf>>. Acesso em: 26/nov./24. ISBN: 9058829596.

DURAN, F. *et al.* **PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST CLASSIFIER DAN LIGHTGBM CLASSIFIER UNTUK PREDIKSI PENYAKIT JANTUNG**. *Data Sciences Indonesia (DSI)*, [s.l.], v. 3, n° 2, p. 98–103, 2024. ISSN: 2809-0721, DOI: 10.47709/dsi.v3i2.3831.

ECOGNITION, T. **Trimble eCognition**. <https://geospatial.trimble.com/en/products/software/trimble-ecognition>. 2024. Disponível em: <<https://geospatial.trimble.com/en/products/software/trimble-ecognition>>. Acesso em: 26/nov./24.

EKEANYANWU, C. V. *et al.* **Merging GIS and Machine Learning Techniques: A Paper Review**. *Journal of Geoscience and Environment Protection*, [s.l.], v. 10, n° 09, p. 61–83, 2022. ISSN: 2327-4336, DOI: 10.4236/gep.2022.109004.

ELANSARI, T.; OUANAN, M.; BOURRAY, H. **A novel Mathematical Modeling for Deep Multilayer Perceptron Optimization: Architecture Optimization and Activation Functions Selection**. *Statistics, Optimization & Information Computing*, [s.l.], v. 12, n° 5, p. 1409–1424, 2024. ISSN: 2310-5070, DOI: 10.19139/soic-2310-5070-1990.

ESA. **Copernicus Dem - Agência Espacial Europeia (ESA)**. [s.d.]. Disponível em: <<https://panda.copernicus.eu/#:~:text=Component%20Data%20Access%20PANDA%20Catalogue.%20PANDA%20allows%20anonymous>>. Acesso em: 26/nov./24.

FAN, Z. **An exploration of reinforcement learning and deep reinforcement learning**. *Applied and Computational Engineering*, [s.l.], v. 73, n° 1, p. 154–159, 2024. ISSN: 2755-2721, DOI: 10.54254/2755-2721/73/20240386.

FONSECA, J.; DOUZAS, G.; BACAO, F. **Improving Imbalanced Land Cover Classification with K-Means SMOTE: Detecting and Oversampling Distinctive Minority Spectral Signatures**. *Information*, [s.l.], v. 12, n° 7, p. 266, 2021. ISSN: 2078-2489, DOI: 10.3390/info12070266.

GDAL. **GDAL documentation**. 2024. Disponível em: <<https://gdal.org/en/latest/>>. Acesso em: 26/nov./24.

GOOGLE. **Receiver-operating characteristic curve (ROC)**. *aberta*. 2024. Disponível em: <[https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#receiver-operating\\_characteristic\\_curve\\_roc](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#receiver-operating_characteristic_curve_roc)>. Acesso em: 27/nov./24.

**Google Earth Engine**. [s.d.]. Disponível em: <<https://earthengine.google.com/>>. Acesso em: 26/nov./24.

GUIMARÃES, M. *et al.* **Predicting Model Training Time to Optimize Distributed Machine Learning Applications**. *Electronics*, [s.l.], v. 12, n° 4, p. 871, 2023. ISSN: 2079-9292, DOI: 10.3390/electronics12040871.

HALDER, A.; GHOSH, A.; GHOSH, S. **Supervised and unsupervised landuse map generation from remotely sensed images using ant based systems**. *Applied Soft Computing*, [s.l.], v. 11, n° 8, p. 5770–5781, 2011. ISSN: 15684946, DOI: 10.1016/j.asoc.2011.02.030.

HERATH, H. C. M. **Performance Evaluation of Machine Learning Classifiers for Hyperspectral Images**. Em: *2021 IEEE 21st International Conference on*

*Communication Technology (ICCT)*. [s.l.]: IEEE, 2021. ISBN: 978-1-6654-3206-1, DOI: 10.1109/ICCT52962.2021.9657977.

HERMOSILLA, T. *et al.* **Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes.** *Remote Sensing of Environment*, [s.l.], v. 268, p. 112780, 2022. ISSN: 00344257, DOI: 10.1016/j.rse.2021.112780.

HOCKINGS, K. J.; SOUSA, C. **Human-Chimpanzee Sympatry and Interactions in Cantanhez National Park, Guinea-Bissau: Current Research and Future Directions.** *Primate Conservation*, [s.l.], v. 26, nº 1, p. 57–65, 2013. ISSN: 0898-6207, DOI: 10.1896/052.026.0104.

IBAP. **Instituto Nacional de Meio Ambiente -IBAP, Guiné-Bissau.** [s.d.]. Disponível em: <<https://ibapgbissau.org/pntc-ap/>>. Acesso em: 26/nov./24.

JASIM, B.; JASIM, O.; AL-HAMEEDAWI, A. **Evaluating Land Use Land Cover Classification Based on Machine Learning Algorithms.** *Engineering and Technology Journal*, [s.l.], v. 0, nº 0, p. 1–12, 2024. ISSN: 2412-0758, DOI: 10.30684/etj.2024.144585.1638.

JIA, M. *et al.* **A New Vegetation Index to Detect Periodically Submerged Mangrove Forest Using Single-Tide Sentinel-2 Imagery.** *Remote Sensing*, [s.l.], v. 11, nº 17, p. 2043, 2019. ISSN: 2072-4292, DOI: 10.3390/rs11172043.

JUNIOR, A. C. *et al.* **PANORAMA SOBRE O ESTADO ATUAL DE CONSERVAÇÃO DO MANGUEZAL DA REGIÃO COSTEIRA DE CACHEU/GUINÉ BISSAU.** *InterEspaço: Revista de Geografia e Interdisciplinaridade*, [s.l.], v. 4, nº 14, p. 49, 2018. ISSN: 2446-6549, DOI: 10.18764/2446-6549.v4n14p49-66.

\_\_\_\_\_. **ÁREAS PROTEGIDAS PARA A CONSERVAÇÃO DOS MANGUEZAIS EM GUINÉ-BISSAU: ESTUDO SOBRE A IMPORTÂNCIA DO PARQUE NATURAL DOS TARRAFES DO RIO CACHEU.** *Revista Gestão & Sustentabilidade Ambiental*, [s.l.], v. 8, nº 2, p. 123, 2019. ISSN: 2238-8753, DOI: 10.19177/rgsa.v8e22019123-154.

\_\_\_\_\_. **Características das Áreas Protegidas em Guiné-Bissau: o caso do Parque Natural dos Tarrafes do Rio Cacheu.** *Revista Eletrônica Casa de Makunaima*, [s.l.], v. 1, n° 2, p. 37–55, 2019. ISSN: 2595-5888, DOI: 10.24979/makunaima.v1i2.461.

KLAIBER, M.; KLOPFER, J. **A Systematic Literature Review on SOTA Machine learning-supported Computer Vision Approaches to Image Enhancement.** *Jurnal Ilmu Komputer dan Informasi*, [s.l.], v. 15, n° 1, p. 21–31, 2022. ISSN: 2502-9274, DOI: 10.21609/jiki.v15i1.1017.

KOSOLWATTANA, T. *et al.* **A self-inspected adaptive SMOTE algorithm (SASMOTE) for highly imbalanced data classification in healthcare.** *BioData Mining*, [s.l.], v. 16, n° 1, p. 15, 2023. ISSN: 1756-0381, DOI: 10.1186/s13040-023-00330-4.

KRAWCZYK, B. **Learning from imbalanced data: open challenges and future directions.** *Progress in Artificial Intelligence*, [s.l.], v. 5, n° 4, p. 221–232, 2016. ISSN: 2192-6352, DOI: 10.1007/s13748-016-0094-0.

KUSSUL, N. *et al.* **Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data.** *IEEE Geoscience and Remote Sensing Letters*, [s.l.], v. 14, n° 5, p. 778–782, 2017. ISSN: 1545-598X, DOI: 10.1109/LGRS.2017.2681128.

KWAN, C. *et al.* **An Accurate Vegetation and Non-Vegetation Differentiation Approach Based on Land Cover Classification.** *Remote Sensing*, [s.l.], v. 12, n° 23, p. 3880, 2020. ISSN: 2072-4292, DOI: 10.3390/rs12233880.

KWENDA, C.; GWETU, M. V.; FONOU-DOMBEU, J. V. **A critical survey of GEOBIA methods for forest image detection and classification.** *Geocarto International*, [s.l.], v. 38, n° 1, 2023. ISSN: 1010-6049, DOI: 10.1080/10106049.2023.2256302.

LANG, S. *et al.* **GEOBIA Achievements and Spatial Opportunities in the Era of Big Earth Observation Data.** *ISPRS International Journal of Geo-Information*, [s.l.], v. 8, n° 11, p. 474, 2019. ISSN: 2220-9964, DOI: 10.3390/ijgi8110474.

LARY, D. J. *et al.* **Machine learning in geosciences and remote sensing.** *Geoscience Frontiers*, [s.l.], v. 7, nº 1, p. 3–10, 2016. ISSN: 16749871, DOI: 10.1016/j.gsf.2015.07.003.

MAHVASH MOHAMMADI, N.; HEZARKHANI, A. **A Comparative Study of SVM and RF Methods for Classification of Alteration Zones Using Remotely Sensed Data.** *Journal of Mining and Environment*, [s.l.], v. 11, nº 1, p. 49–61, 2020. ISSN: 22518606, DOI: 10.22044/jme.2019.7956.1664.

MCFEETERS, S. K. **The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features.** *International Journal of Remote Sensing*, [s.l.], v. 17, nº 7, p. 1425–1432, 1996. ISSN: 0143-1161, DOI: 10.1080/01431169608948714.

MONTEIRO, M. A.; BIDARRA, Z. S.; PIFFER, M. **A estratégia top-down da política pública de agricultura familiar na Guiné-Bissau.** *Revista Brasileira de Desenvolvimento Regional*, [s.l.], v. 8, nº 2, p. 55, 2020. ISSN: 2317-5443, DOI: 10.7867/2317-5443.2020v8n2p55-78.

MONTESINOS LÓPEZ, O. A.; MONTESINOS LÓPEZ, A.; CROSSA, J. **Multivariate Statistical Machine Learning Methods for Genomic Prediction.** Cham: Springer International Publishing, 2022. ISBN: 978-3-030-89009-4, DOI: 10.1007/978-3-030-89010-0.

MOSCHETTO, F. A.; RIBEIRO, R. B.; FREITAS, D. M. DE. **Urban expansion, regeneration and socioenvironmental vulnerability in a mangrove ecosystem at the southeast coastal of São Paulo, Brazil.** *Ocean & Coastal Management*, [s.l.], v. 200, p. 105418, 2021. ISSN: 09645691, DOI: 10.1016/j.ocecoaman.2020.105418.

MULA, S. **O peso crescente do fenómeno urbano e o seu impacto na Segurança Alimentar: um estudo de caso na comunidade peri-urbana de Susana, São Domingos (Guiné-Bissau).** - ISCTE - Instituto Universitario de Lisboa, 2010. Disponível em:

<<https://www.proquest.com/openview/08fa13fde08ebd0abd7b329325211dd7/1?cbl=2026366&diss=y&pq-origsite=gscholar>>. Acesso em: 26/nov./24.

NECULA, S.-C.; DUMITRIU, F.; GREAVU-ŞERBAN, V. **A Systematic Literature Review on Using Natural Language Processing in Software Requirements Engineering.** *Electronics*, [s.l.], v. 13, n° 11, p. 2055, 2024. ISSN: 2079-9292, DOI: 10.3390/electronics13112055.

NERSISYAN, S. *et al.* **ExhauFS: exhaustive search-based feature selection for classification and survival regression.** *PeerJ*, [s.l.], v. 10, p. e13200, 2022. ISSN: 2167-8359, DOI: 10.7717/peerj.13200.

NGO, L. H. *et al.* **The composition of time-series images and using the technique SMOTE ENN for balancing datasets in land use/cover mapping.** *Acta Montanistica Slovaca*, [s.l.], n° 27, p. 342–359, 2022. ISSN: 1335-1788, DOI: 10.46544/AMS.v27i2.05.

NGUYEN, G. *et al.* **Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey.** *Artificial Intelligence Review*, [s.l.], v. 52, n° 1, p. 77–124, 2019. ISSN: 0269-2821, DOI: 10.1007/s10462-018-09679-z.

NICULESCU, S.; LAM, C. N. **Geographic Object-Based Image Analysis of Changes in Land Cover in the Coastal Zones of the Red River Delta (Vietnam).** *Journal of Environmental Protection*, [s.l.], v. 10, n° 03, p. 413–430, 2019. ISSN: 2152-2197, DOI: 10.4236/jep.2019.103024.

OUCHRA, H.; BELANGOUR, A.; ERRAISSI, A. **Supervised Machine Learning Algorithms for Land Cover Classification in Casablanca, Morocco.** *Ingénierie des systèmes d'information*, [s.l.], v. 29, n° 1, p. 377–387, 2024. ISSN: 16331311, DOI: 10.18280/isi.290137.

ÖZBULDU, M.; ŞEKERLI, Y. E. **Comparative analysis of different supervised methods for satellite-based land-use classification: a case study of Reyhanlı.** *Mustafa Kemal Üniversitesi Tarım Bilimleri Dergisi*, [s.l.], 2024. ISSN: 2667-7733, DOI: 10.37908/mkutbd.1485236.

PARTHA PROTIM ROY; MD. SHAHRIAR ABDULLAH; IQTIAR MD. SIDDIQUE. **Machine learning empowered geographic information systems: Advancing Spatial analysis and decision making.** *World Journal of Advanced Research and Reviews*,

[s.l.], v. 22, n° 1, p. 1387–1397, 2024. ISSN: 25819615, DOI: 10.30574/wjarr.2024.22.1.1200.

POWAR, S. K.; PANHALKAR, S. S.; PATIL, A. S. **An Evaluation of Pixel-based and Object-based Classification Methods for Land Use Land Cover Analysis Using Geoinformatic Techniques.** *Geomatics and Environmental Engineering*, [s.l.], v. 16, n° 2, p. 61–75, 2022. ISSN: 2300-7095, DOI: 10.7494/geom.2022.16.2.61.

PRATAMA, A. R. I.; LATIPAH, S. A.; SARI, B. N. **OPTIMASI KLASIFIKASI CURAH HUJAN MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN RECURSIVE FEATURE ELIMINATION (RFE).** *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, [s.l.], v. 7, n° 2, p. 314–324, 2022. ISSN: 2540-8984, DOI: 10.29100/jipi.v7i2.2675.

PRIYATNO, A. M.; WIDIYANINGTYAS, T. **A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS.** *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, [s.l.], v. 9, n° 2, p. 196–207, 2024. ISSN: 2527-4864, DOI: 10.33480/jitk.v9i2.5015.

R. **RStudio.** 2024. Disponível em: <<https://posit.co/download/rstudio-desktop/>>. Acesso em: 26/nov./24.

RAJA SAKTI ARIEF, D.; SYAHRIL, E.; SUHERMAN. **Review of Literature on Improving the KNN Algorithm.** *Transactions on Machine Learning and Artificial Intelligence*, [s.l.], v. 11, n° 3, 2023. ISSN: 20547390, DOI: 10.14738/tecs.113.14768.

RAMACHANDRA, T. V.; MONDAL, T.; SETTURU, B. **Relative performance evaluation of machine learning algorithms for land use classification using multispectral moderate resolution data.** *SN Applied Sciences*, [s.l.], v. 5, n° 10, p. 274, 2023. ISSN: 2523-3963, DOI: 10.1007/s42452-023-05496-4.

RAMEZAN, C. A. **Transferability of Recursive Feature Elimination (RFE)-Derived Feature Sets for Support Vector Machine Land Cover Classification.** *Remote Sensing*, [s.l.], v. 14, n° 24, p. 6218, 2022. ISSN: 2072-4292, DOI: 10.3390/rs14246218.

RAO, N. S. V. **Study of Overfitting by Machine Learning Methods Using Generalization Equations**. Em: *2023 26th International Conference on Information Fusion (FUSION)*. [s.l.]: IEEE, 2023. ISBN: 979-8-89034-485-4, DOI: 10.23919/FUSION52260.2023.10224198.

RICHARDSON, E. *et al.* **The receiver operating characteristic curve accurately assesses imbalanced datasets**. *Patterns*, [s.l.], v. 5, n° 6, p. 100994, 2024. ISSN: 26663899, DOI: 10.1016/j.patter.2024.100994.

RITU; BHAMBRI, P. **Software Effort Estimation with Machine Learning – A Systematic Literature Review**. *Agile Software Development*. [s.l.]: Wiley, 2023. p. 291–308. DOI: 10.1002/9781119896838.ch15.

SALAMA, M. **Optimization of Regression Models Using Machine Learning: A Comprehensive Study with Scikit-learn**. *International Uni-Scientific Research Journal*, [s.l.], v. 5, p. 119–129, 2024. ISSN: 27354660, DOI: 10.59271/s45500.024.0624.16.

SAMI, Jairson Alberto; FRANCISCO, Cristiane Nunes; FERNANDES, Pedro José Farias; HOLANDA, Tiago Fernando de. **Mapeamento do índice de sensibilidade no litoral da Guiné-Bissau por GEOBIA e Machine Learning**. *Revista de Gestão Costeira Integrada*, v. 24, n. 2, p., dez. 2024. DOI: <https://doi.org/10.5894/rgci-n568>, Disponível em: <https://www.aprh.pt/rgci>. Acesso em: 01 de maio. 2025

SÁNCHEZ-AZOFEIFA, G. A. *et al.* **Experiences in Field Data Collection**. *Remote Sensing of Forest Environments*. Boston, MA: Springer US, 2003. p. 433–446. DOI: 10.1007/978-1-4615-0306-4\_16.

SCHONLAU, M. **Random Forests**. [s.l.]: [s.n.], 2023. p. 183–204. DOI: 10.1007/978-3-031-33390-3\_10.

SCIKIT-LEARN. **Eliminação Recursiva de Características (RFE)**. 2024. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_selection.RFE.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html)>. Acesso em: 27/nov./24.

SECA, A. I.; PEREIRA, H. D. S.; SILVA, S. C. P. Da. **Expansion of Cashew Cultivation and its Environmental and Economic Impacts on Guinea-Bissau.** *Revista Brasileira de Ciências Ambientais*, [s.l.], v. 56, n° 3, p. 385–397, 2021. ISSN: 21769478, DOI: 10.5327/Z21769478950.

SHAIK, A. S.; SHAIK, N.; PRIYA, Dr. C. K. **Predictive Modeling in Remote Sensing Using Machine Learning Algorithms.** *International Journal of Current Science Research and Review*, [s.l.], v. 07, n° 06, 2024. ISSN: 25818341, DOI: 10.47191/ijcsrr/V7-i6-62.

SHALU; GURJEET SINGH. **ENVIRONMENTAL MONITORING WITH MACHINE LEARNING.** *EPRA International Journal of Multidisciplinary Research (IJMR)*, [s.l.], p. 208–212, 2023. ISSN: 24553662, DOI: 10.36713/epra13330.

SINGH, P. S. *et al.* **Enhanced classification of hyperspectral images using improvised oversampling and undersampling techniques.** *International Journal of Information Technology*, [s.l.], v. 14, n° 1, p. 389–396, 2022. ISSN: 2511-2104, DOI: 10.1007/s41870-021-00676-0.

SONKO, S. M.; DEAC, A. L. **FROM SIMPLE TOURISM TO SMART TOURISM: THE BET FOR THE DEVELOPMENT OF A RESPONSIBLE AND SUSTAINABLE TOURISM. CASES IN THE ZIGUINCHOR REGION (SENEGAL).** *Analele Universității din Oradea, Seria Geografie*, [s.l.], v. 30, n° 1, p. 10–19, 2020. ISSN: 12211273, DOI: 10.30892/auog.301102-838.

SOUSA, J. *et al.* **The (dis)engagement of mangrove forests and mangrove rice in academic and non-academic literature on Guinea-Bissau—a systematic review protocol.** *PLOS ONE*, [s.l.], v. 18, n° 4, p. e0284266, 2023. ISSN: 1932-6203, DOI: 10.1371/journal.pone.0284266.

SUYKENS, J. A. K.; VANDEWALLE, J. **Least Squares Support Vector Machine Classifiers.** *Neural Processing Letters*, [s.l.], v. 9, n° 3, p. 293–300, 1999. ISSN: 13704621, DOI: 10.1023/A:1018628609742.

TEOH, T. T.; RONG, Z. **Python for Data Analysis**. [s.l.]: [s.n.], 2022. p. 107–122. DOI: 10.1007/978-981-16-8615-3\_7.

THANH NOI, P.; KAPPAS, M. **Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery**. *Sensors*, [s.l.], v. 18, n° 1, p. 18, 2017. ISSN: 1424-8220, DOI: 10.3390/s18010018.

UPRETI, A. **Machine learning application in GIS and remote sensing: An overview**. *International Journal of Multidisciplinary Research and Growth Evaluation*, [s.l.], p. 546–553, 2022. ISSN: 25827138, DOI: 10.54660/anfo.2022.3.4.20.

WEATHER, S. **Clima e condições meteorológicas médias em Bissau no ano todo**. 2024. Disponível em: <<https://pt.weatherspark.com/y/31647/Clima-caracter%C3%ADstico-em-Bissau-Guin%C3%A9-Bissau-durante-o-ano>>. Acesso em: 26/nov./24.

WU, L. *et al.* **Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure**. *International Journal of Remote Sensing*, [s.l.], v. 40, n° 15, p. 5973–5991, 2019. ISSN: 0143-1161, DOI: 10.1080/01431161.2019.1584929.

YAN, X. *et al.* **Rapid Land Cover Classification Using a 36-Year Time Series of Multi-Source Remote Sensing Data**. *Land*, [s.l.], v. 12, n° 12, p. 2149, 2023. ISSN: 2073-445X, DOI: 10.3390/land12122149.

YANG, X. *et al.* **Artificial neural network modelling in GIS spatial analysis**. *Academic Journal of Computing & Information Science*, [s.l.], v. 7, n° 6, 2024. ISSN: 26165775, DOI: 10.25236/AJCIS.2024.070605.

ZHANG, C.; GUO, J.; LU, J. **Research on Classification Method of High-Dimensional Class-Imbalanced Data Sets Based on SVM**. Em: *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)*. [s.l.]: IEEE, 2017. ISBN: 978-1-5386-1600-0, DOI: 10.1109/DSC.2017.63.

ZHANG, M.; SHI, W.; XU, Z. **Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data.** *Hydrology and Earth System Sciences*, [s.l.], v. 24, n° 5, p. 2505–2526, 2020. ISSN: 1607-7938, DOI: <https://doi.org/10.22044/jme.2019.7956.1664>.

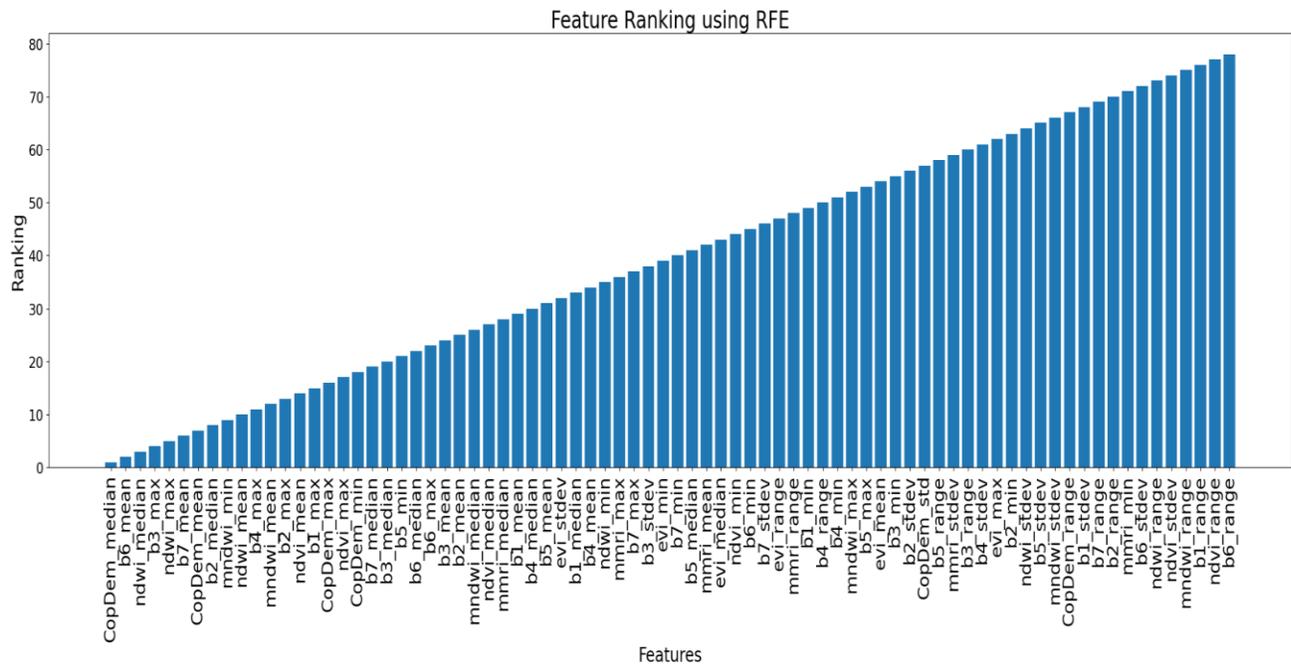
ZHAO, Q.; QU, Y. **The Retrieval of Ground NDVI (Normalized Difference Vegetation Index) Data Consistent with Remote-Sensing Observations.** *Remote Sensing*, [s.l.], v. 16, n° 7, p. 1212, 2024. ISSN: 2072-4292, DOI: 10.3390/rs16071212.

ZHU, B.; LIAO, J.; SHEN, G. **Combining time series and land cover data for analyzing spatio-temporal changes in mangrove forests: A case study of Qinglangang Nature Reserve, Hainan, China.** *Ecological Indicators*, [s.l.], v. 131, p. 108135, 2021. ISSN: 1470160X, DOI: 10.1016/j.ecolind.2021.108135.

ZHU, Z. **Systematic Optimization of Overfitting Problem in Machine Learning.** *Highlights in Science, Engineering and Technology*, [s.l.], v. 111, p. 353–359, 2024. ISSN: 2791-0210, DOI: 10.54097/3tkzrj84.

# 7. APÊNDICE

## 7.1 Classificação das variáveis segundo o RFE



## 7.2 Registro Fotográfico Do Trabalho De Campo







