

## Desenvolvimento de ferramenta de Machine Learning para classificação automatizada de uso e cobertura da terra em linhas de distribuição da CELESC

Pedro Henrique Machado Porath<sup>1\*</sup>, Fernanda Oliveira da Silva<sup>2</sup>, Francisco Caruso Jr<sup>3</sup>, Jessica Finco<sup>4</sup>, Bruno Tiago Paulo<sup>5</sup>, Dionatan Oster Scherer<sup>6</sup>

<sup>1</sup>Mestrado em Engenharia de Transportes e Gestão Territorial, CARUSO Soluções Ambientais & Tecnológicas, Brasil. (\*Autor correspondente: pedroporath@giscarusu.com)

<sup>2</sup>Mestrado em Ecossistemas Agrícolas e Naturais, CARUSO Soluções Ambientais & Tecnológicas, Brasil.

<sup>3</sup>Doutorado em Geociência, CARUSO Soluções Ambientais & Tecnológicas, Brasil.

<sup>4</sup>Mestrado em Oceanografia, CARUSO Soluções Ambientais & Tecnológicas, Brasil.

<sup>5</sup>MBA em Gestão de Projetos, CELESC Distribuição S. A., Brasil.

<sup>6</sup>Engenheiro Elétrico, CELESC Distribuição S. A., Brasil.

Histórico do Artigo: Submetido em: 26/01/2024 – Revisado em: 13/02/2024 – Aceito em: 27/03/2024

### RESUMO

A CELESC está entre as maiores empresas do setor elétrico brasileiro, tendo como área de concessão de distribuição de energia elétrica quase todo o território catarinense e o município de Rio Negro (PR). Possui mais de 5.000 km de linhas de distribuição, representando assim um complexo cenário quando o assunto é manutenção das faixas de servidão especialmente quanto aos serviços de podas e roçadas. Possuir o mapeamento do uso e ocupação da terra atualizado é um importante elemento para auxiliar o processo de identificação do grau de dificuldade e frequência necessária para a realização dessas manutenções. Dessa forma, a partir do projeto de P&D ANEEL “PD-05697-0122/2022” junto à CELESC, dentre os produtos gerados, por meio de linguagem de programação Python, foi proposto o desenvolvimento de uma *toolbox* para o QGIS utilizando a técnica de *Machine Learning Random Forest* e imagens do Satélite Sentinel-2 para gerar automaticamente mapeamento do Uso e Cobertura da Terra para o estado de Santa Catarina. O mapeamento obteve o coeficiente Kappa e Acurácia Geral, o valor de 0,96. Por fim, na sequência foi realizada uma análise espacial para identificar o perfil do uso da terra nas linhas de distribuição da CELESC. A ferramenta proposta possui um potencial para otimizar a extração e o processamento de dados em vastas áreas, facilitando a classificação do uso e cobertura da terra.

**Palavras-Chaves:** Aprendizado de máquina, Random Forest, Linha de distribuição.

### Development Of Machine Learning Tool For Automated Classification Of Land Use And Coverage In Celesc Distribution Lines

CELESC is among the largest companies in the Brazilian electrical sector, with almost the entire territory of Santa Catarina and the municipality of Rio Negro (PR) as its electricity distribution concession area. It has more than 5.000 km of distribution lines, representing a complex scenario when it comes to maintaining rights of way, especially regarding pruning and mowing services. Having an up-to-date mapping of land use and occupation is an important criterion to assist in the process of identifying the degree of difficulty and frequency required to carry out this maintenance. Thus, based on the ANEEL P&D project “PD-05697-0122/2022” with CELESC, among the products generated, using the Python programming language, was proposed the development of a toolbox for QGIS using Random Forest (technique of Machine Learning) and images from the Sentinel-2 Satellite to automatically generate Land Use and Cover mapping for the state of Santa Catarina. The mapping got for Kappa index and Accuracy, a value of 0.96. Finally, a spatial analysis was then carried out to identify the land use and occupation profile in CELESC's distribution lines. The proposed tool has the potential to optimize data extraction and processing across vast areas, facilitating the classification of land use and coverage.

**Keywords:** Machine Learning, Random Forest, Distribution Line.

Porath, P. H. M., Silva, F. O., Caruso Jr, F., Finco, J., Paulo, B. T., Scherer, D. O. (2024). Desenvolvimento de ferramenta de Machine Learning para classificação automatizada de uso e cobertura da terra em linhas de distribuição da CELESC. *Revista Brasileira de Sensoriamento Remoto*, v.5, n.2, p.31-42.



## 1. Introdução

A empresa Centrais Elétricas de Santa Catarina S.A. (CELESC) é responsável pela distribuição de energia elétrica em quase todo território catarinense, possuindo 177 subestações e mais de 5.000 km de linhas de distribuição. A extensão considerável dessas linhas implica que seja onerosa a manutenção de ativos, especialmente para serviços como podas e roçadas em faixas de servidão (CELESC, 2022).

A faixa de servidão, enquanto corredor de segurança essencial, protege as linhas e redes de distribuição contra interferências externas, como vegetação não controladas, minimizando os riscos de curtos-circuitos e interrupções no fornecimento de energia, conforme a ABNT NBR 5422:2024. Um dos critérios utilizados pelos gestores da CELESC para cálculo do custo e frequência de contratação de empresa terceirizada para realizar atividades de limpeza de faixa de servidão, é o mapeamento atualizado do uso e cobertura da terra. A detenção deste dado atualizado e preciso e forma automatizada representa um desafio significativo.

Ao mapear grandes regiões (em escala nacional ou regional), surgem desafios significativos na seleção e processamento de imagens de satélites. A aquisição de imagens para abranger uma extensão territorial considerável resulta em uma grande quantidade de dados, demandando uma infraestrutura robusta e eficiente. Lidar com esse desafio implica não apenas em capacidade computacional considerável, mas também estratégias sofisticadas para organizar, armazenar e processar essas informações de maneira a extrair dados cruciais para a tomada de decisão (Mohanrajan; Loganathan; Manoharan, 2020; Zhang; Li, 2022; Zhu et al., 2023). A extração dessas informações pode ser realizada utilizando diferentes linguagem de programação, como Python e R, possuem bibliotecas estáveis para o processamento de imagens utilizando algoritmos *Machine Learning* (Lemenkova; Debeir, 2022).

As potencialidades do *Machine Learning*, uma subdivisão da inteligência artificial, sua aplicação em imagens providas por Sensoriamento Remoto tem se destacado. Os seus diferentes algoritmos são capazes de modelar as assinaturas espectrais de classes complexas, também conseguem receber múltiplos dados preditores de entrada e não fazem suposições sobre a distribuição de dados (ou seja, não são paramétricos) (Talukdar et al., 2020). As técnicas de *Machine Learning* foram categorizadas em dois subtipos: não supervisionadas e supervisionadas. As não supervisionadas incluem por exemplo os algoritmos *de Affinity Propagation (AP)*, *Cluster Algorithm*, *Fuzzy C-Means Algorithms*, *K-Means Algorithm*. Enquanto supervisionadas incluem *Support Vector Machine (SVM)*, *Spectral Angle Mapper (SAM)*, *Decision Tree (DT)*, *Naive Bayes (NB)*, *Neural Networks (RNA)* (Halder, Ghosh & Ghosh, 2011) e *Random Forest (RF)* (Breiman, 2001).

A precisão de um mapa de classificação supervisionada está diretamente relacionada a qualidade das amostras de treinamento usada pelo método de *Machine Learning* (Vellido et al., 2019). Entre os algoritmos de classificação supervisionada, o *Random Forest* se destaca por ser aplicado para fins de classificação e regressão, logo, pode ser usado como variáveis categóricas e contínuas (Woznicki et al., 2019). *Random Forest* aplicado ao sensoriamento remoto, destaca-se pela sua precisão na classificação, bem como a capacidade de lidar com a alta dimensionalidade e multicolinearidade de dados. Além disso, é notável por sua velocidade de processamento e a robustez contra a overfitting (Belgiu; Drăgu, 2016) e outlier (Corcoran, Knight & Gallant, 2013).

Essa técnica tem sido utilizada para o mapeamento de uso e cobertura da terra em larga escala, como por exemplo, o MapBiomias que reconstruiu informações anuais de uso e cobertura da terra no período de 1985 a 2017 para o Brasil, com a coleção imagens dos sensores Landsat (Souza et al., 2020). Carrasco et al. (2019) propôs mapeamento do uso e cobertura do solo na Inglaterra aplicando o Random Forest. O mapeamento de uso e cobertura da terra envolve uma série de etapas, desde a seleção de dados, processamento de dado e classificação do uso e cobertura da terra.

A gestão eficiente do processamento desses grande conjunto de dados é crucial para garantir que o produto gerado seja preciso, relevante e escalável. Para o presente projeto de pesquisa foi desenvolvido uma

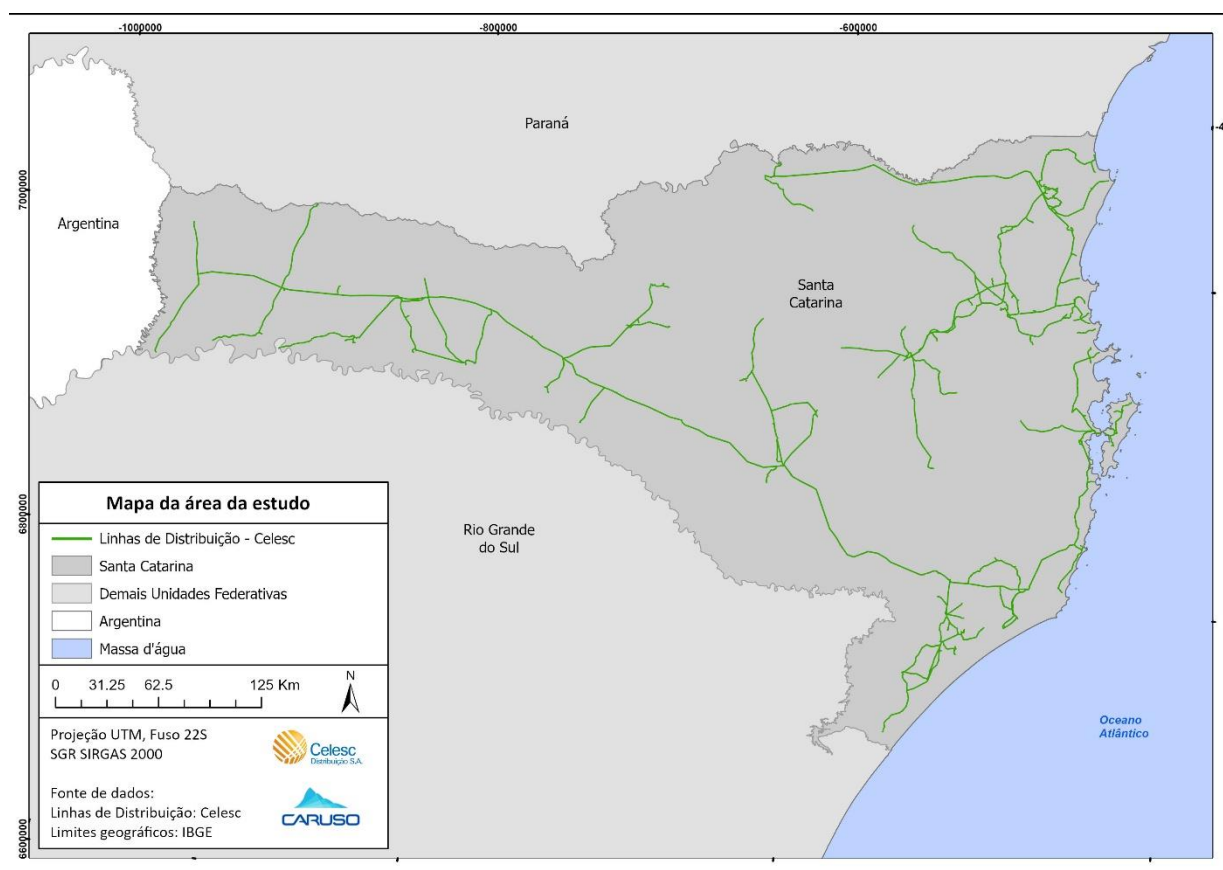
*toolbox* para o software QGIS contendo três ferramentas, onde o seu fim é realizar a classificação do uso e cobertura da terra para o estado de Santa Catarina por meio do algoritmo de *Random Forest* com imagens de satélites do Sentinel-2 do ano de 2023.

## 2. Material e Métodos

### 2.1 Área de Estudo

A delimitação da área para mapeamento do uso e ocupação da terra foi de 5 km além dos limites geográficos do estado de Santa Catarina, devido à alguns municípios dos estados vizinhos serem parcialmente atendidos pela CELESC, além da inclusão do município de Rio Negro (PR) por ter seu fornecimento de energia integralmente suprido pela CELESC, totalizando assim 108,24 mil km<sup>2</sup>. Mas para este artigo, a área de estudo é delimitada pelas linhas de distribuição da CELESC, conforme ilustrado pela Figura 1, possuindo cerca de 92,98 km<sup>2</sup> quando considerado as faixas de servidão.

**Figura 1** - Fluxograma de processos da ferramenta desenvolvida.



Fonte: Autores (2024)

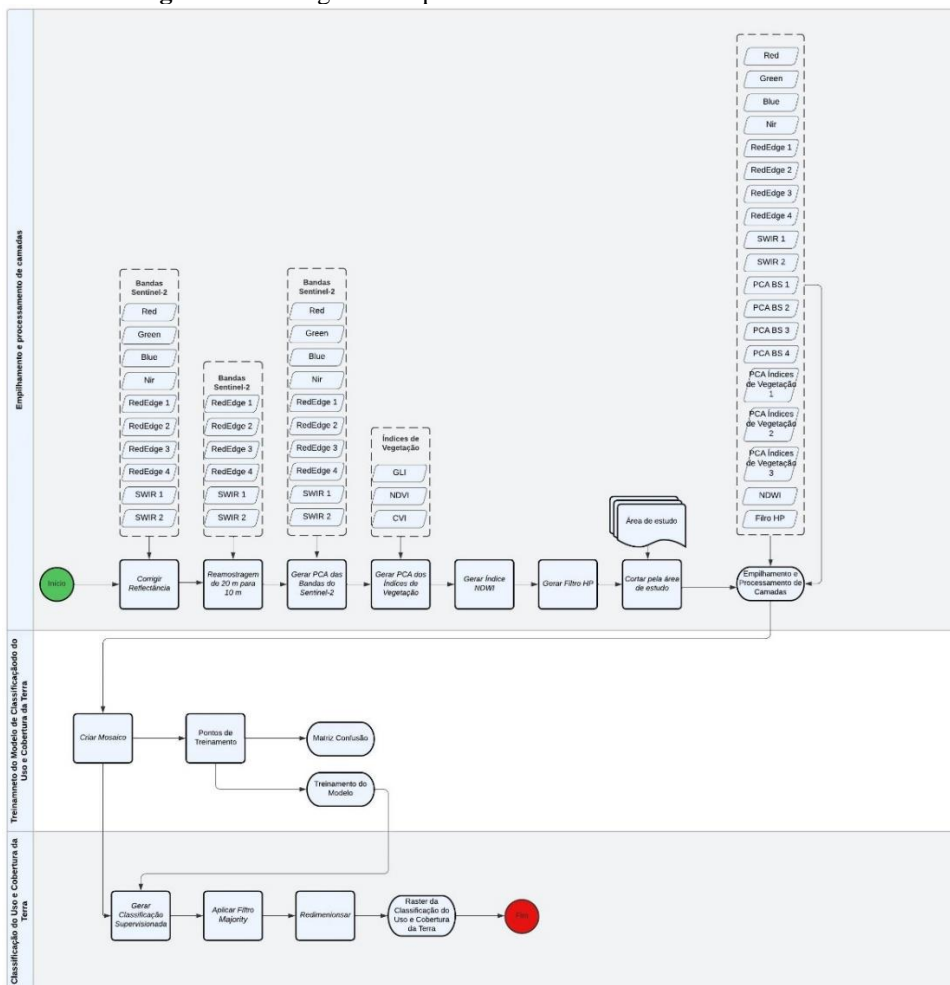
Assim, foram utilizadas imagens do satélite Sentinel-2 capturadas pelo sensor multiespectral MSI que conta com treze bandas espectrais. A missão Sentinel é um projeto da Agência Espacial Europeia (ESA) tendo

seu uso voltado a ações como uso e cobertura da terra, monitoramento de desastres ambientais, aptidão agrícola etc. As imagens desta missão foram adquiridas gratuitamente por meio do site do Sentinel Hub.

Para a compor a área de estudo foi necessário a utilização de vinte e duas cenas de imagens do Sentinel-2. Para elas, foi ainda adicionado o filtro de imagens do verão do ano de 2023 e com no máximo 5% de nuvem foi trabalhado com a menor porcentagem de nuvens possível. As nuvens podem mascara a reflectância da superfície da Terra.

Em atividades de geoprocessamento, a automatização de fluxos de processos é cada vez mais requisitada, devido aos benefícios de agilidade e eliminação de riscos de erros provocados pela atividade humana. Para isso, a partir da linguagem de programação Python e bibliotecas geoespaciais como GDAL, GRASS, SAGA e PyQGIS, foi desenvolvido uma toolbox para uso no software QGIS voltada para processamento de imagens do Sentinel-2, composta por três ferramentas: Empilhamento e Processamento de Camadas; Treinamento do Modelo de Classificação do Uso e Cobertura Terra e Classificação do Uso e Cobertura Terra.

**Figura 2 - Fluxograma de processos da ferramenta desenvolvida.**



Fonte: Autores (2024)

## 2.2 Processamento da primeira ferramenta

A primeira ferramenta da toolbox, Empilhamento e Processamento de Camadas, tem por função preparar um empilhamento de camadas de arquivos matriciais gerados a partir das bandas multiespectrais do Sentinel-2, com o objetivo de gerar um conjunto de dados de entrada para o treinamento e classificação do uso e cobertura da terra. Esta ferramenta foi projetada para processar e empilhar as bandas do Sentinel-2, incluiu Red; Green; Blue; RedEdge 1; RedEdge 2; RedEdge 3; RedEdge 4; Nir; SWIR1 e SWIR2 como entrada de dados. O processo resulta na geração de imagem aprimorada, que passou pelas etapas que incluem correção atmosférica, normalização radiometria e aplicação de técnicas avançadas, culminando um conjunto final de dados otimizados e para a análise. As seguintes etapas detalham o fluxo do processo: Conversão dos números digitais (DN) em refletância no topo da atmosfera (TOA) para as bandas: Red; Green; Blue; RedEdge 1; RedEdge 2; RedEdge 3; RedEdge 4; Nir; SWIR1 e SWIR2.

Com auxílio da biblioteca do GRASS foi implementada uma função que reamostra as imagens para uma resolução maior de 10 metros, uma vez que as bandas RedEdge e SWIR possuem uma resolução de 20 metros e nesta ferramenta optou-se por padronizar a resolução espacial para 10 metros.

Geração da análise de componentes principal (PCA) das bandas: Red; Green; Blue; RedEdge 1; RedEdge 2; RedEdge 3; RedEdge 4; Nir; SWIR1 e SWIR2. Nesta função, são retornadas as componentes principais primeira, segunda, quarta e nona das bandas multiespectrais mencionadas.

Além disso, foram desenvolvidas funções para calcular os seguintes índices de vegetação:

- O Green Lead Index (GLI) enfatiza a presença de folhas verdes saudáveis na vegetação (Louhaichi, Borman & Johnson, 2001), determinado pela equação 1:

$$GLI = \frac{((2 \times GREEN) - RED - BLUE)}{((2 \times GREEN) + RED + BLUE)} \quad (1)$$

Onde: GREEN é referente ao valor de refletância na banda do verde; RED é referente ao valor de refletância na banda do vermelho; BLUE é referente ao valor de refletância na banda do azul.

- O NDVI é amplamente utilizado para avaliar a vegetação e é sensível a mudanças na quantidade e saúde da vegetação (Huang et al., 2021), determinado pela equação 2.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

Onde: NIR é referente ao valor de refletância na banda do infravermelho próximo.

- O CVI (Chlorophyll Vegetation Index) está relacionado à concentração de clorofila nas folhas das plantas (Vincini, Frazzi e D'Alessio, P, 2008), sendo sensível à quantidade de clorofila presente

$$CVI = \frac{NIR \times RED}{GREEN^2} \quad (3)$$

Obtendo os três índices de vegetação, em seguida, a função aplica uma análise de componente principal (PCA) para reduzir a dimensionalidade dos dados, eliminar redundâncias e realçar características importantes

relacionadas à saúde das plantas. O resultado inclui as três primeiras componentes, que têm finalidades semelhantes trazer informações sobre a sanidade da vegetação.

Foi implementada a função para determinar o NDWI (Normalized Difference Water Index), conforme proposto por McFeeters, 1996. Essa função utiliza dados de refletância de infravermelho próximo e infravermelho de onda curta.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (4)$$

Onde: SWIR é referente ao valor de refletância na banda do infravermelho de onda curta.

Por fim, foi empregado o filtro Laplaciano da ferramenta Saga, que utiliza a banda NIR (Infravermelho Próximo) para processamento.

### 2.3 Processamento da segunda e terceiras ferramentas

Já a segunda ferramenta, Treinamento do Modelo da Classificação do Uso e Cobertura da Terra, recebe como entrada o empilhamento e processamento de uma ou mais imagens do Sentinel-2, que foram combinadas em um único mosaico para fins de classificação, juntamente com o arquivo contendo amostras para o treinamento do modelo Random Forest. No contexto deste estudo, foram escolhidas as seguintes classes para treinamento do modelo: solo exposto, floresta, silvicultura, área urbana, pastagem/agricultura e corpo d'água.

A coleta de amostras de categorias da classificação de uso e cobertura da terra é um passo importante para o treinamento de modelos de classificação de uso e cobertura da terra a partir de imagens de sensoriamento remoto. Neste estudo, foram coletadas 9.468 amostras ou polígonos de treinamento que abrangem diferentes classes presentes na sua área de estudo. Cada polígono foi rotulado com a classe, o que permite ao algoritmo de aprendizado de máquina Random Forest aprender a associar características das imagens com classes específicas de uso e cobertura da terra.

Como saída, a ferramenta fornece o empilhamento das diferentes camadas em um mosaico abrangendo a área de estudo, o modelo treinado e a matriz de confusão. A matriz de confusão é um instrumento que ajuda a avaliar o desempenho do modelo, permitindo comparar as previsões do modelo com os resultados reais e determinar a precisão da classificação das diferentes classes de uso e cobertura da terra. Esse procedimento desempenha papel importante na interpretação e utilização das imagens de satélite no contexto do estudo.

A terceira e última ferramenta de classificação do uso e cobertura da terra tem como entrada o mosaico gerado e o arquivo de treinamento do modelo. Esse, tem como objetivo gerar um arquivo contendo o mapeamento do uso e ocupação da terra a partir das classes de solo exposto, floresta, silvicultura, área urbana, pastagem/agricultura e corpo d'água. Por fim, ainda é aplicado ao raster recém gerado o filtro Majority, a classificação por pixel geralmente tem o efeito “sal e pimenta”, assim removendo os pixels isolados ou mal classificados, aprimorando assim a qualidade da classificação.

## 3. Resultados e Discussão

A primeira ferramenta, chamada de “Empilhamento e Processamento de Camadas Matriciais”, resulta em um arquivo composto por 18 camadas, listadas na seguinte sequência: Red (Banda 4 do Sentinel-2); Green (Banda 3 do Sentinel-2); Blue (Banda 2 do Sentinel-2); REDEDGE 1 (Banda 5 do Sentinel-2); REDEDGE 2 (Banda 6 do Sentinel-2); REDEDGE 3 (Banda 7 do Sentinel-2); NIR (Banda 8 do Sentinel-2); SWIR 1 (Banda 11 do Sentinel-2); SWIR 2 (Banda 12 do Sentinel-2); Primeira componente das bandas multiespectrais (PCA BS 1); Segunda componente das bandas multiespectrais (PCA BS 2); Quarta componente das bandas

multiespectrais (PCA BS 3); Nona componente das bandas multiespectrais(PCA BS 4); Primeira componente dos índices de vegetação (PCA Índice de Vegetação 1); Segunda componente dos índices de vegetação (PCA Índice de Vegetação 2); Terceira componente dos índices de vegetação(PCA Índice de Vegetação 3); NDWI; Filtro Laplacian. A Figura 3 exibe o resultado obtido por meio da aplicação da ferramenta “Empilhamento e Processamento de Camadas Matriciais”.

Essa ferramenta foi desenvolvida para conter diversos processo, desde o tratamento das imagens do Sentinel-2, gerar índices de vegetações e aplicar PCAs, retorna as camadas já empilha em um único arquivo. Entre as vantagens apresentadas, redução do erro humano e otimização do tempo do usuário. Já que todos esses processos que compõem a primeira ferramenta executados sequencialmente, consomem muito tempo. Assim, mesmo pessoas com menos conhecimento em sensoriamento remoto podem utilizar o plugin com facilidade, bastando inserir as bandas do Sentinel-2.

A segunda ferramenta retorna o treinamento do modelo Random Forest e o resultado da matriz de confusão. A partir dos resultados obtidos, foi realizado uma validação cruzada com o total de 204.063 pixels das diferentes classes visando a obtenção da matriz de confusão (Tabela 1).

**Tabela 1** - Matriz de confusão e Coeficiente Kappa.

Classe pred/verd	Área						Corpo d'água	Total	Comissão	Acurácia Usuário
	Solo	Floresta	Silvicultura	urbana	Past/Ag					
Solo exposto	23.745	176	206	434	1.199	41	25.801	0,08	0,92	
Floresta	146	41.342	656	6	340	22	42.512	0,03	0,97	
Silvicultura	62	717	32.685	3	36	10	33.513	0,02	0,98	
Área urbana	357	7	3	29.520	92	128	30.107	0,02	0,98	
Pastagem/ Agricultura	2.105	331	110	124	36.458	18	39.146	0,07	0,93	
Corpo d'água	6	1	19	29	34	32.895	32.984	0,00	1,00	
<b>Total</b>	<b>26.421</b>	<b>42.574</b>	<b>33.679</b>	<b>30.116</b>	<b>38.159</b>	<b>33.114</b>	<b>204.063</b>	<b>0,22</b>		
Omissão	0,10	0,03	0,03	0,02	0,04	0,01	0,23			
Acurácia Produtor	0,90	0,97	0,97	0,98	0,96	0,99				
Acurácia Geral	0,96									
Kappa	0,96									
p®	0,17									

Fonte: Autores (2024)

Essa matriz traz as informações relevantes sobre a precisão do mapa e a confusão espectral entre as classes. Ao avaliar a acurácia geral e coeficiente Kappa do modelo, foi identificado que ambos os indicadores atingiram o valor de 0,96 (Tabela 1), classificado como excelente segundo Richads (2019).

A validação da precisão é de suma importância no processamento de dados de sensoriamento remoto. Uma forma de avaliar é pela matriz de confusão, pois ela traz dados de acurácia do produto e acurácia do usuário. Acurácia do produtor determina os erros de omissão e quão bem os dados coletados refletem as características reais da superfície da Terra. Já a acurácia do usuário, mede os erros de comissão que

representam a probabilidade de um pixel classificado corresponder ao tipo de cobertura do solo de sua localização correspondente no mundo real (Rwanga; Ndambuki, 2017).

A partir da matriz de confusão, foi gerada a Tabela 2, que resume as estimativas de precisão do usuário e do produtor. Como se pode observar, a acurácia do usuário variou de 0,92% (solo exposto) a 1,00% (corpo d'água). A alta acurácia pode estar relacionada ao NDWI, proposto por (Mcfeeters, 1996), que é amplamente utilizado na detecção e mapeamento de corpos d'água. Já a menor precisão do produtor e do usuário aconteceu na classe de solo exposto onde os valores foram de 0,90% e 0,92% respectivamente, que pode ser explicado em função da confusão espectral com a classe de pastagem/agricultura.

**Tabela 2** - Acurácia e erro padrão do usuário e produtor para as classes do uso e cobertura da terra.

Classes	Usuário		Produtor	
	Acurácia	Erro	Acurácia	Erro
Solo exposto	0,92	0,08	0,90	0,10
Floresta	0,97	0,03	0,97	0,03
Silvicultura	0,98	0,02	0,97	0,03
Área urbana	0,98	0,02	0,98	0,02
Pastagem/Agricultura	0,93	0,07	0,96	0,04
Corpo d'água	1,00	0,00	0,99	0,01
MÉDIA	0,98	0,03	0,99	0,01

Fonte: Autores (2024)

Logo como resultado da última ferramenta, a Figura 3 representa o mapeamento do uso e cobertura da terra para a área de estudo, dando destaque para os municípios de Joaçaba (SC) e Herval D'Oeste (SC) com o objetivo de evidenciar as classes corretamente identificadas.

Com o mapeamento do uso e cobertura da terra para o estado de Santa Catarina e verificação da acurácia do modelo, pôde ser realizado o recorte espacial pelas áreas que interseccionam as linhas de distribuição da CELESC. Essa análise permitiu analisar o perfil do uso e ocupação nas áreas estudadas, revelando que a classe de pastagem/agricultura é a mais predominante, representando 40,43% do total. Seguido das classes de floresta (24,66%), solo exposto (19,87%), área urbana (13,04%) e com menor expressão as classes de silvicultura (1,06%) e corpo d'água (0,94%), como ilustrado na Tabela 3.

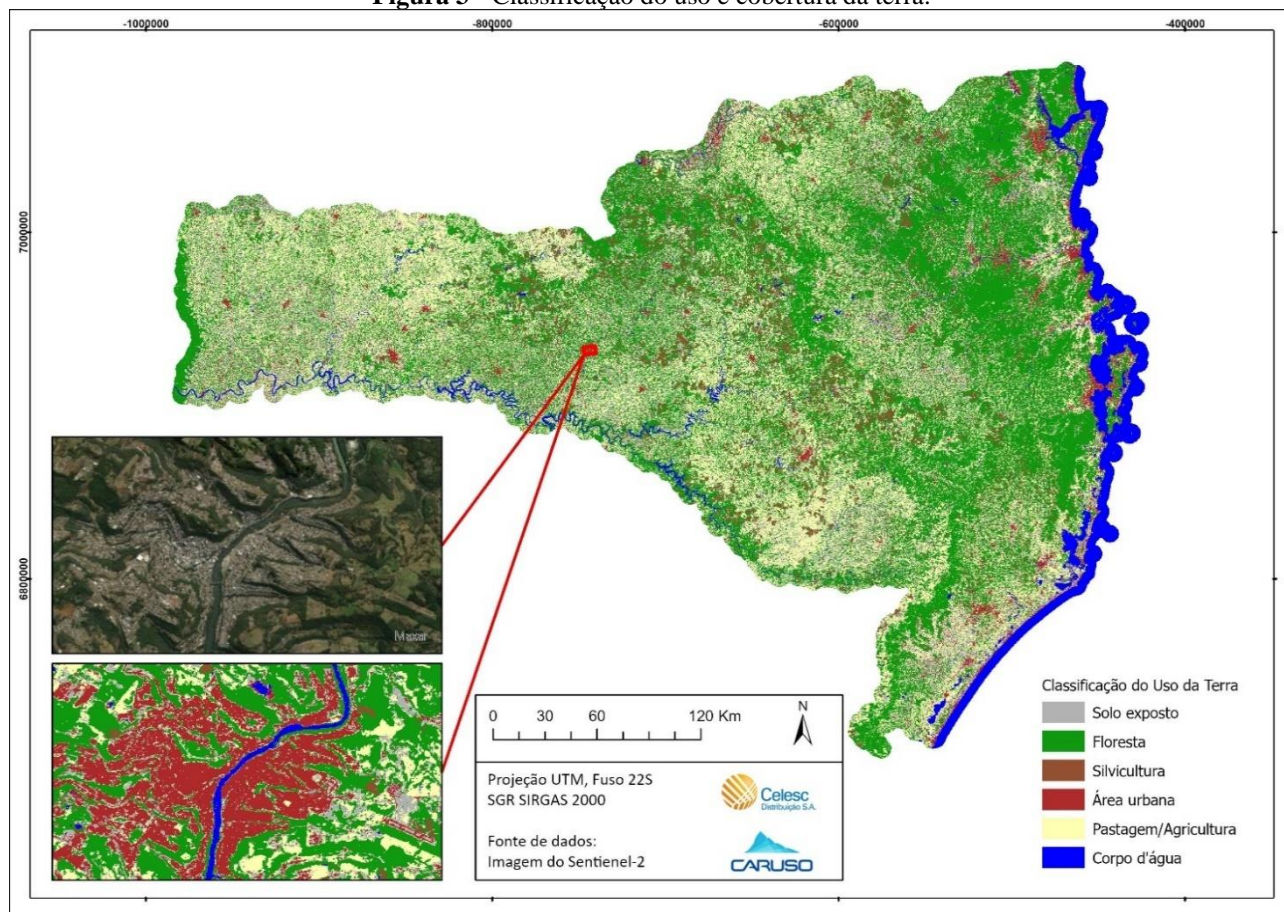
**Tabela 3** - Classes do uso e cobertura da terra, área em km<sup>2</sup> e porcentagem destas classes nas linhas de distribuição da empresa CELESC.

Classes	Área (Km <sup>2</sup> )	Área (%)
Pastagem/Agricultura	37,59	40,43%
Floresta	22,93	24,66%
Solo exposto	18,48	19,87%
Área urbana	12,13	13,04%
Silvicultura	0,99	1,06%
Corpo d'água	0,87	0,94%
<b>TOTAL</b>	<b>92,98</b>	<b>100%</b>

Fonte: Autores (2024)



**Figura 3 - Classificação do uso e cobertura da terra.**



Fonte: Autores (2024)

Por meio da metodologia e ferramentas propostas foi possível mapear e quantificar as classes de uso e cobertura da terra ao longo das linhas de distribuição da CELESC. O conhecimento do tipo de classe é crucial para a CELESC estimar os custos das atividades de poda e roçadas. Por exemplo, em classes como silvicultura e floresta nativa, são áreas com maior densidade de árvores que podem comprometer a operação da concessionária. O solo exposto é uma classe que demandam menor atenção para execução desses serviços, devido a característica de ausência de vegetação.

Com o aprimoramento das técnicas de *Machine Learning*, foi possível observar o aumento significativo na acurácia geral dos modelos. Como é o caso do projeto MapBiomias, onde houve uma melhora nos resultados das acurácias gerais da Coleção 3.1 nível 1 (proposto abril de 2019) para a Coleção 7.1 nível 1 (proposto em abril de 2023), saindo de 89,1% para 91,5% baseado no uso da coleção Landsat (MAPBIOMAS, 2023). Embora para o trabalho aqui proposto o mapeamento seja menos complexo, devido a sua geração para um único ano e bioma (Mata Atlântica), a acurácia geral obtida por meio de imagens do Sentinel-2 com valores de 0,96 é expressivo. O modelo apresenta limitações na classificação, principalmente quando se trata das classes de agricultura e solo exposto.

No entanto, demonstra um potencial em mapear diversas classes em uma área extensa, como é a área de estudo. O monitoramento das classes de uso e ocupação da terra nas linhas de servidão da CELESC é um processo que deve ser realizado anualmente. O toolbox proposto representa uma maneira eficiente de otimizar desde o processo mais onerosos, que é o processamento das imagens do sentinel-2 até classificação do uso e

cobertura da terra.

#### 4. Conclusão

Neste trabalho, todas as cenas do Sentinel-2 utilizadas são referentes aos meses de janeiro a abril de 2023. Todavia a utilização de series temporais unido ao aprendizado de máquina tem sido amplamente empregado no uso e cobertura da terra e detecção de mudanças com bons resultados. Franklin et al. (2015) demonstrou em seu estudo que a aplicação de séries temporais na classificação melhorou a precisão geral em 6,38% em relação aos resultados de dados de uma única data. É importante reforçar que a pesquisa de classificação de uso e cobertura da terra está em fase de desenvolvimento e a metodologia prevê a utilização de series temporais para otimizar a classificação do uso e cobertura da terra para o estado de Santa Catarina nas faixas de servidão da CELESC.

Ao avaliar a usabilidade do *toolbox* proposto para o *software* QGIS, pode-se concluir que a geração da classificação do uso e cobertura da terra a partir da técnica de *Machine Learning Random Forest* com dados do Sentinel-2 foi bastante eficiente, uma vez que o *toolbox* automatizou uma série de etapas e processamentos complexos, diminuindo a onerosidade e a exposição a erros provocados pelo ser humano. Demonstrando que a metodologia e conjunto de ferramentas apresentada foi de extrema importância para a CELESC mapear a classificação do uso e ocupação de terra em seus ativos facilitando assim a tomada de decisão de gestores em extensas e complexas áreas.

#### 5. Agradecimentos

Por fim, os autores são gratos à CELESC por todo o apoio financeiro recebido durante o desenvolvimento do projeto e ao Programa de Pesquisa, Desenvolvimento e Inovação - PDI da ANEEL. Agradecemos também a todos os membros da equipe da Divisão de Pesquisa, Desenvolvimento e Inovação (DVPI) da CELESC Distribuição S.A., nas pessoas da Sra. Maria de Lourdes Machado Viccari, Lucas Lapolli Briguenti e Thiago José de Chaves, pelo acompanhamento de execução administrativa e financeira do projeto. Em especial, agradecemos também ao Eng. Roberto Kinceler, Gerente da DVPI do Programa de P&D e Inovação da CELESC, pela atenção e presteza, que tornaram possível a realização deste projeto.

#### 6. Referências

ABNT - Associação Brasileira de Normas Técnicas (2024). **NBR 5422 - Projeto de linhas aéreas de energia elétrica – Critérios técnicos**. Rio de Janeiro, 2024.

Belgiu, M., Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, 144, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.

Breiman, L. (2001). **Random Forests**. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>.

Carrasco, L. et al. (2019). Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. **Remote Sensing**, 11 (3), 288. <https://doi.org/10.3390/rs11030288>.

CELESC - Centrais Elétricas de Santa Catarina S.A. (2022). Área de Concessão. Brasil. Disponível em:

<https://www.celesc.com.br/a-celesc-distribuicao>. Acesso em: 01/08/2023.

Corcoran, J. M., Knight, J. F., Gallant, A. L. (2013). Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. **Remote Sensing**, 5 (7), 3212–3238. <https://doi.org/10.3390/rs5073212>.

Franklin, S. E. et al (2015). Large Area Mapping of Annual Land Cover Dynamics Using Multitemporal Change Detection and Classification of Landsat Time Series Data. **Canadian Journal of Remote Sensing**, 41(4), 293–314. <https://doi.org/10.1080/07038992.2015.1089401>.

Halder, A., Ghosh, A., Ghosh, S. (2011). Supervised and unsupervised land use map generation from remotely sensed images using ant based systems. **Applied Soft Computing Journal**, 11 (8), 5770–5781. <https://doi.org/10.1016/j.asoc.2011.02.030>.

Huang, S., Tang, L., Hupy, J. P., Wang, Y., Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. **Journal of Forestry Research**, 32(5). <https://doi.org/10.1007/s11676-020-01155-1>.

Lemenkova, P.; Debeir, O. (2022). Satellite Image Processing by Python and R Using Landsat 9 OLI/TIRS and SRTM DEM Data on Côte d'Ivoire, West Africa. **Journal of Imaging**, 8, 12. <https://doi.org/10.3390/jimaging8120317>.

Louhaichi, M., Borman, M. M., Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, 16, 65–70. <https://doi.org/10.1080/10106040108542184>.

Mcfeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, 17, 1425–1432. <https://doi.org/10.1080/01431169608948714>.

Mohanrajan, S. N.; Loganathan, A.; Manoharan, P (2020). Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges. **Environmental Science and Pollution Research**, 27(24), 29900-29926. <https://doi.org/10.1007/s11356-020-09091-7>.

Rwanga, S. S., Ndambuki, J. M. (2017) Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. **International Journal of Geosciences**, 08, 611–622. <https://doi.org/10.4236/ijg.2017.84033>.

Souza, C. M. et al (2020). Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, 12, 17. <https://doi.org/10.3390/rs12172735>.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal., S., Liou, Y., Rahman, A. (2020). Land-use land-cover classification by machine learning classifiers for satellite observations-A review. **Remote Sensing**, 12(7), 1135. <https://doi.org/10.3390/rs12071135>.

Vellido, A., Gibert, K., Ângulo, C., Guerrero, J. D. M. (2019). *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*, Barcelona: Springer.

Vincini, M., Frazzi, E., D'Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. **Precision Agriculture**, 9 (5), 303 -319. <https://doi.org/10.1007/s11119-008-9075-z>.

Woznicki, S. A, Baynes, J., Panlasigui, S., Mehaffey, M., Neale, A. (2019). Development of a spatially complete floodplain map of the conterminous United States using random forest. **Science of the Total Environment**, 647, 942–953. <https://doi.org/10.1016/j.scitotenv.2018.07.353>.

Zhang, C.; Li, X (2022). Land Use and Land Cover Mapping in the Era of Big Data. **Land**, 11, 1692. <https://doi.org/10.3390/land11101692>.

Zhu, J.; Zhang, Z.; Zhao, F.; Su, H.; Gu, Z.; Wang, L (2023). Efficient Management and Scheduling of Massive Remote Sensing Image Datasets. **ISPRS International Journal of Geo-Information**, 12 (5), 199. <https://doi.org/10.3390/ijgi12050199>.