



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE INFORMÁTICA (INF)
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
(PPGCC)

YURI KUIVJOGI FERNANDES

**Um sistema WebGIS para classificação supervisionada
de cobertura do solo utilizando inteligência artificial**

GOIÂNIA

2022



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

☒ [x] Dissertação ☐ [] Tese ☐ [] Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Yuri Kuivjogi Fernandes

3. Título do trabalho

Um sistema WebGIS para classificação supervisionada de cobertura do solo utilizando inteligência artificial

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento ☒ [X] SIM ☐ [] NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;

- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Ronaldo Martins Da Costa, Professor do Magistério Superior**, em 08/12/2022, às 08:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Yuri Kuivjogi Fernandes, Discente**, em 08/12/2022, às 21:22, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3387532** e o código CRC **C510A41F**.

Referência: Processo nº 23070.045922/2022-31

SEI nº 3387532

YURI KUIVJOGI FERNANDES

Um sistema WebGIS para classificação supervisionada de cobertura do solo utilizando inteligência artificial

Dissertação apresentada ao Programa de Pós-Graduação em 2022, do Instituto de Informática, da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Linha de pesquisa: Sistemas Inteligentes e Aplicações

Orientador: Professor Doutor Ronaldo Martins da Costa

GOIÂNIA

2022

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Fernandes, Yuri Kuivjogi

Um sistema WebGIS para classificação supervisionada de
cobertura do solo utilizando inteligência artificial [manuscrito] / Yuri
Kuivjogi Fernandes. - 2022.

LXVII, 67 f.: il.

Orientador: Prof. Dr. Ronaldo Martins da Costa.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto
de Informática (INF), Programa de Pós-Graduação em Ciência da
Computação em rede (UFG/UFMS), Goiânia, 2022.

Bibliografia.

Inclui gráfico, tabelas, lista de figuras, lista de tabelas.

1. sensoriamento remoto. 2. identificação espectral. 3. inteligência
artificial. 4. processamento em nuvem. 5. culturas agrícolas. I. Martins
da Costa, Ronaldo, orient. II. Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **19/2022** da sessão de Defesa de Dissertação de **Yuri Kuivjogi Fernandes**, que confere o título de Mestre em Ciência da Computação, na área de concentração em Ciência da Computação.

Aos vinte e um dias do mês de outubro de dois mil e vinte e dois, a partir das dez horas, na sala 150 do Instituto de Informática, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Um sistema WebGIS para classificação espectral de culturas utilizando inteligência artificial**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor Ronaldo Martins da Costa (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professora Doutora Bruna Mendes de Oliveira (EA/UFG), membra titular externa; Professor Doutor Édipo Henrique Cremon (IFG), membro titular externo, cuja participação ocorreu por videoconferência. Durante a arguição os membros da banca fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Ronaldo Martins da Costa, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos vinte e um dias do mês de outubro de dois mil e vinte e dois.

TÍTULO SUGERIDO PELA BANCA

Um sistema WebGIS para classificação supervisionada de cobertura do solo utilizando inteligência artificial



Documento assinado eletronicamente por **Ronaldo Martins Da Costa, Professor do Magistério Superior**, em 21/10/2022, às 12:57, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Édipo Henrique Cremon, Usuário Externo**, em 21/10/2022, às 12:58, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **YURI KUIVJOGI FERNANDES, Discente**, em 21/10/2022, às 15:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Bruna Mendes De Oliveira, Professor do Magistério Superior**, em 21/10/2022, às 17:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3158956** e o código CRC **3D53F075**.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Yuri Kuivjogi Fernandes

Graduou-se em Geologia na USP - Universidade de São Paulo. Durante a graduação foi bolsista por dois anos do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), tendo realizado atividades de iniciação científica voltada para datação de rochas alcalinas. Tem experiência na área de Sistemas de Informações Geográficas (GIS). Atualmente trabalha com projetos em Sensoriamento Remoto e Inteligência Artificial.

Resumo

Fernandes, Yuri K.. **Um sistema WebGIS para classificação supervisionada de cobertura do solo utilizando inteligência artificial**. Goiânia, 2022. 63p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

Com o avanço na geração de dados para observação da Terra e sua disponibilização de maneira gratuita, a área de Sensoriamento Remoto (SR) avançou de maneira expressiva. Ao longo dos anos, tem-se observado a migração de aplicações de SR para o ambiente da internet, facilitando pesquisas utilizações diversas. Este trabalho propõe uma nova abordagem para coleta e manipulação de dados espaciais para classificação espectral baseada em pixels. Uma aplicação web foi construída integrando serviços do Google Earth Engine, Google Maps e Auto Machine Learning para análise de desempenho. Experimentos usando regiões amostrais de cobertura do solo da região de Goiás, Brasil, demonstraram o ganho em tempo, processamento e armazenamento de dados. Tais contribuições estão relacionadas à grande quantidade de informações de imagens de satélite coletadas de maneira convencional, as quais posteriormente não são utilizadas. Como resultado final, tem-se uma imagem classificada por meio do processo de classificação representando as diferentes classes de cobertura do solo. O treinamento do modelo obteve uma acurácia de 99,85% usando o modelo Light Gradient Boosting Machine (LightGBM). Além destes benefícios, a otimização de processos permite a inclusão de pesquisadores de outras grandes áreas, contribuindo assim para a maior disseminação de conhecimento da área de SR e aplicações de reconhecimento de padrões.

Palavras-chave

Sensoriamento remoto, identificação espectral, inteligência artificial, processamento em nuvem, culturas agrícolas

Abstract

Fernandes, Yuri K.. **A WebGIS system for spectral classification of cultures using artificial intelligence**. Goiânia, 2022. 63p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

With the advancement in data generation for Earth observation and its availability free of charge, the Remote Sensing (SR) area advanced significantly. Over the years, it has been observed the migration of RS applications to the internet environment, facilitating searches of different uses. This work proposes a new approach for collecting and manipulating spatial data for spectral classification based on pixels. A web application was built integrating Google Earth Engine, Google Maps and Auto Machine Learning services for performance analysis. Experiments using samples from land cover regions in Goiás, Brazil, justifying the gain in time, processing and data storage. Such contributions are related to the large amount of information from satellite images collected in a conventional way, which are later not used. As a final result, there is an image classified through the classification process representing the different land cover classes. Model training achieved an accuracy of 99.85% using the Light Gradient Boosting Machine (LightGBM) model. In addition to these benefits, the optimization of processes allows the inclusion of research from other major areas, thus for the greater dissemination of knowledge in the area of SR and pattern recognition applications.

Keywords

Remote sensing, classification, artificial intelligence, cloud processing

Sumário

Lista de Figuras	8
Lista de Tabelas	10
1 Introdução	11
1.1 Contexto	11
1.2 Motivação	12
1.3 Objetivo	12
1.4 Organização do trabalho	13
2 Revisão da Literatura	14
2.1 Planejamento	14
2.1.1 Finalidade da Revisão Sistemática	14
2.1.2 Questões para condução da revisão	14
2.1.3 Base de dados e critérios para busca de trabalhos	15
2.1.4 Chaves de buscas utilizadas	15
2.1.5 Critérios de inclusão e exclusão de artigos	16
2.2 Condução	16
2.3 Análise de dados	17
2.3.1 Considerações finais da revisão	24
3 Fundamentação teórica	25
3.1 Sistema de Informações Geográficas - SIG	25
3.1.1 Aplicações SIG na internet	27
3.2 Sensoriamento Remoto	28
3.2.1 Energia eletromagnética	28
3.2.2 Imagens de Sensoriamento Remoto	29
3.2.3 Imagens multiespectrais	30
3.2.4 Índices espectrais	33
3.2.5 Classificação de imagens de SR	33
3.2.6 Classificação supervisionada baseada em pixel	34
3.3 Inteligência Artificial em Sensoriamento Remoto	34
3.3.1 Random Forest	34
3.3.2 Light Gradient Boosting Machine (LightGBM)	35
3.3.3 Automated Machine Learning (AutoML)	37

4	Material e métodos	38
4.1	Descrição da ferramenta	38
4.1.1	Escolhas de desenvolvimento	38
4.1.2	Arquitetura do sistema	39
4.1.3	Detalhamento do algoritmo	40
4.1.4	Coleta de dados	41
4.2	Experimentos	41
4.2.1	Criação dos dados de treinamento	41
4.2.2	Preparação de dados	42
4.2.3	Treinamento de modelos de classificação	42
5	Resultados e discussões	43
5.1	Regiões amostrais classificadas	43
5.2	Otimização do tempo	47
5.3	Otimização do armazenamento	47
5.4	Trabalhos correlatos	49
6	Conclusão	50
6.1	Contribuições	50
6.2	Trabalhos futuros	51
	Referências Bibliográficas	53

Lista de Figuras

2.1	Gráfico sobre aquisição de imagens.	23
3.1	Exemplo de dado vetorial. Extraído e adaptado de: Conceitos básicos em ciência da geoinformação. [15]	26
3.2	Exemplo de dado matricial. Extraído e adaptado de: Conceitos básicos em ciência da geoinformação. [15]	26
3.3	Propriedades de onda da energia eletromagnética. Sendo a) e c) ondas com uma dada frequência ν e b) 2ν . Extraído de: Aplicações de sensoriamento remoto [49]	29
3.4	Espectro eletromagnético. Extraído de: https://www.todamateria.com.br/espectro-eletromagnetico/	30
3.5	Traduzido de: http://landsat.gsfc.nasa.gov/?p=10643 . Ilustrando diferentes resoluções espectrais.	31
3.6	Extraído de: Remote Sensing and image interpretation [47]. As 4 imagens têm mesma resolução espacial porém diferentes resoluções radiométricas, com valores de 1-bit, 2-bits, 4-bits e 8-bits, respectivamente nos quadros a), b), c) e d).	31
3.7	Extraído de: Remote Sensing and image interpretation [47]. Ilustrando pixels em imagens e seus valores digitais. Em a) temos uma malha de 500 por 400 <i>pixels</i> , em b) temos 100 por 80 e em c) 10 por 8. A última figura em d) ilustra os valores dos <i>pixels</i> em c).	32
3.8	Traduzido de: Remote Sensing and image interpretation [47]. Ilustrando sobreposição de bandas espectrais.	32
3.9	Extraído de: Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities, [47]. Diferentes tipos de classificação. a) Nível de pixel, classificação foca em atrelar cada pixel à uma classe. b) Nível de objeto, classificação foca em reconhecer e segmentar objetos na imagem. c) Nível de cena, classificação foca em reconhecer o contexto da imagem.	34
3.10	Ilustração sobre fluxo do algoritmo de Árvore de Decisão	36
3.11	Extraída e adaptada da documentação do algoritmo LightGBM (disponível em: https://lightgbm.readthedocs.io/en/v3.3.2/). Ilustração apresentando a diferença na construção de árvores de decisão.	37
4.1	Arquitetura da ferramenta.	39
4.2	Modelo de classes implementado.	39
4.3	Ilustração contendo principais páginas da plataforma. Na sequência temos: a) a página inicial, b) menu de catálogos de variedades, c) detalhamento de região amostral, d) adição de região amostral.	40

4.4	Fluxo de execuções do algoritmo	40
5.1	Região amostral do estádio Serra Dourada, Goiânia.	44
5.2	Região amostral do autódromo de Goiânia.	44
5.3	Região amostral na região do aeroporto e bairro de Santa Genoveva, Goiânia.	45
5.4	Região amostral da cidade de Goianira.	45
5.5	Região amostral da cidade de Caldas Novas.	46
5.6	Comparativo da região classificada em Caldas Novas e uma imagem completa do Satélite Sentinel 2-MSI.	48

Lista de Tabelas

2.1	Quantidade de artigos selecionados por base de pesquisa	17
2.2	Categorias Monitoramento	18
2.3	Outras categorias	18
2.4	Resumo de satélite e sensores utilizados nos artigos	19
2.5	Resumo de características de fontes das imagens	19
2.6	Bandas e índices utilizados	20
2.7	Linguagens utilizadas	21
2.8	Métodos de Inteligência Artificial	22
2.9	Serviços em nuvem	23
3.1	Tabela com dados fictícios	36
4.1	Tabela com principais bibliotecas utilizadas	38
4.2	Relação de amostras coletadas de coberturas de solo	42
4.3	Métricas dos resultados de treinamentos	42
5.1	Tabela de resultados com ênfase em tempo de processamento	47
5.2	Tabela de resultados com ênfase em armazenamento	48

Introdução

1.1 Contexto

A área de Sensoriamento Remoto (SR), desde seu surgimento, sempre lidou com grande volume de dados. Além disso, há uma tendência de que esse volume continue em constante crescimento devido ao avanço nas tecnologias relacionadas ao desenvolvimento de sensores e processamento de imagens aliado ao aumento na disponibilidade desses dados. Os requisitos de software para gerenciamento e análise se tornam cada vez mais exigentes e menos viáveis para pacotes de software comuns e recursos de computação local.

De acordo com Agrawal e Gupta [2], ocorre uma migração dessas aplicações para o ambiente da internet desde 1993. Em 1999, foi lançada a *Web 2.0*, seguida de aplicações como *Google Maps*, *Google Earth* e *Microsoft Bing Map*. Em 2010, a *Google* lançou uma plataforma baseada em nuvem chamada *Google Earth Engine* (GEE), possibilitando análises científicas e visualização de dados geoespaciais em larga escala [32].

A quantidade de benefícios, principalmente observando aspectos financeiro e de escalabilidade, tem feito com que a computação em nuvem continue sendo amplamente utilizada no desenvolvimento de software, e se torne um atributo permanente [75]. Nos últimos anos temos visto o uso de Aprendizado de Máquina em diversas áreas de conhecimento. Na área de SR as aplicações de técnicas de reconhecimento de padrões são encontradas nos mais diversos assuntos, como afirmam Zhang et al. [102] em sua revisão sobre aplicações de Inteligência Artificial em Sensoriamento Remoto.

Mesmo com muitos avanços na área, a coleta de dados ainda é um ponto crítico no aprendizado de máquina. Para a correta utilização destas técnicas há a necessidade de uma grande quantidade de dados rotulados. E, se avançarmos para métodos mais robustos, como de aprendizado profundo, a necessidade se torna ainda maior [73]. Em contrapartida, a maioria dos estudos focam apenas no processo de análise e classificação. E de acordo com Press [65], a maior parte do tempo dispendido em análises de dados ocorre durante a aquisição e preparação dos dados.

As aplicações em classificação de cobertura do solo são inúmeras. Há iniciativas globais como pelo Globeland30 [16] fazendo monitoramento a classificação de grandes áreas. Uma nova tecnologia denominada *Auto Machine Learning* tem sido usada por muito classificadores para otimizar processos e, pretende-se usá-la neste trabalho para classificação de cobertura do solo.

1.2 Motivação

De maneira geral, apesar de softwares de Sensoriamento Remoto apresentarem interface, usuários não especializados têm bastantes dificuldades de uso. Isso ocorre devido às dezenas de menus e opções para manipulação e processamento de imagens com o fim de trazer maior versatilidade à ferramenta para as mais diversas aplicações.

Além disso, como dito anteriormente, em trabalhos que envolvam algoritmos de classificação, o maior tempo dispendido é nas tarefas de criação e preparação de banco de dados para análises, gerando maior nível de complexidade para usuários leigos.

O desenvolvimento de uma ferramenta cujo objetivo esteja bem definido e específico pode melhorar a usabilidade da ferramenta e trazer otimização de processos. Uma menor quantidade de informações demandadas para o usuário pode também diminuir a complexidade do uso da ferramenta.

Inicialmente o principal destinatário da ferramenta desenvolvida foi o projeto criado em conjunto com a Escola de Agronomia da Universidade Federal de Goiás, o qual visa a classificação de variedades de cana de açúcar através de SR. Mas ao longo do desenvolvimento, percebeu-se que há muitas outras aplicações para o uso da ferramenta, as quais podem culminar em futuros projetos e parcerias envolvendo a grande área de Cobertura do Solo.

1.3 Objetivo

Este trabalho propõe o desenvolvimento de uma ferramenta online, para aplicação em sensoriamento remoto para classificação supervisionada baseada em *pixel*. Tal *software* visa otimizar processos de: tempo de obtenção das imagens, necessidade de armazenamento, e preparação de dados para posterior processamento.

Para atingir tal objetivo foram identificadas as principais tarefas que necessitavam ser automatizadas:

- Obtenção de imagens multiespectrais;
- Manipulação de imagens e dados com referência espacial;
- Rotulação de *pixels* baseada na classe definida pelo usuário;

- Treinamento de modelos de classificação de imagens multiespectrais;
- Classificação de *pixels* em um problema multiclasse;
- Geração de imagem classificada;

1.4 Organização do trabalho

Este trabalho foi estruturado em cinco capítulos. No primeiro, são apresentados introdução e objetivo da pesquisa. No segundo, há um detalhamento sobre a revisão de trabalhos correlatos, por meio de uma metodologia sistematizada. Nele são apresentadas principais tendências, oportunidades e considerações sobre o panorama de softwares de classificação. Já no capítulo três, são descritos conceitos e definições importantes relacionados ao tema da pesquisa e aplicações. No capítulo quatro, detalha-se o material e métodos envolvidos no trabalho, com a descrição da ferramenta, incluindo arquitetura, processos e resultados. Por fim, no quinto capítulo são descritas as principais conclusões do trabalho, bem como as principais contribuições de pesquisa e sugestões para trabalhos futuros.

Revisão da Literatura

Com o intuito de se obter o entendimento do tema deste trabalho e seus principais avanços recentes, a revisão da literatura foi realizada seguindo um método de pesquisa específico denominado Revisão Sistemática, consolidado na área de Engenharia de *Software* por Kitchenham [46]. Tal método tem se mostrado eficiente quanto a construção de uma base de referências concentrada em um tema de pesquisa. A seguir serão detalhadas as etapas executadas durante a revisão, as quais em síntese são: planejamento, condução e análise de dados.

2.1 Planejamento

Durante a fase de planejamento alguns artigos classificados como relevantes para área foram utilizados para realização de uma pesquisa exploratória. Tal etapa possibilitou a definição de vários itens importantes da pesquisa como: objetivo geral, questões de pesquisa, palavras-chave, critérios de seleção de artigos e formato dos dados de extração.

2.1.1 Finalidade da Revisão Sistemática

Essa revisão visa encontrar e avaliar artigos referentes a aplicações online para lidar com classificação de imagens de satélites na área de agricultura e SR. Além disso, procurou-se entender quais são as fontes de dados mais utilizadas, quais são as principais aplicações, como tais ferramentas foram implementadas, e quais técnicas de classificação aplicadas.

2.1.2 Questões para condução da revisão

Com o intuito de atingir as metas acima descritas foram definidas as seguintes questões a serem respondidas:

- Q1: Quais ferramentas com aplicações em Sistemas de Informações Geográficas (SIG) estão sendo desenvolvidas em ambiente web na área de sensoriamento remoto e agricultura?
- Q2: Como são implementadas tais ferramentas?
- Q3: Quais os objetivos dessas ferramentas?
- Q4: Quais fontes de dados utilizados?
- Q5: Quais ferramentas fazem classificação? Com quais métodos?
- Q6: Quais ferramentas utilizam recursos de processamento em nuvem?
- Q7: Quais métricas de avaliação são utilizadas?

2.1.3 Base de dados e critérios para busca de trabalhos

A revisão da literatura foi feita nas seguintes bases de dados *ACM Digital Library*¹, *IEEE Digital Library*², *Science Direct*³, *Scopus*⁴ e *Springer*⁵. Apenas o período de 2015 a 2022, na língua inglesa ou portuguesa, foi abordado. Foram excluídas publicações de capítulos de livros e publicações incompletas.

2.1.4 Chaves de buscas utilizadas

Cada base de dados requer uma sintaxe específica para realização de buscas, podendo ser utilizados operadores lógicos e aplicar diversos tipos de filtros. Essa é uma etapa importante pois busca-se utilizar palavras-chave e sinônimos a fim de obter o maior número de artigos que tratam do assunto em questão. A seguir são listadas as bases de dados e respectivas chaves de buscas utilizadas:

- Scopus: (("webgis"OR "geoportal"OR "map service"OR "web based"OR "web application") AND ("remote sensing"OR "earth observation"OR "spatial analysis"))
- IEEE: (("webgis"OR "map service"OR "geoportal"OR "web based"OR "web application") AND ("remote sensing"OR "earth observation"OR "spatial analysis"))
- Springer: "webgis"or "map service"or "geoportal"or "web based"or "web application"and ("remote sensing"or "earth observation"or "spatial analysis")
- ACM: [All: (([] OR [[Keywords: "remote sensing"] AND [[Keywords: "machine learning"] OR [Keywords: "deep learning"]]] AND [[Keywords: "software"] OR [Keywords: "web interface"] OR [Keywords: "gui"] OR [Keywords: "open source"]

¹Disponível em: <https://dl.acm.org/>

²Disponível em: <https://ieeexplore.ieee.org/>

³Disponível em: <https://www.sciencedirect.com/>

⁴Disponível em: <https://www.scopus.com/>

⁵Disponível em: <https://link.springer.com/>

OR [Keywords: "ide"]]] OR [[Publication Title: "remote sensing"] AND [[Publication Title: "machine learning"] OR [Publication Title: "deep learning"]] AND [[Publication Title: "software"] OR [Publication Title: "web interface"] OR [Publication Title: "gui"] OR [Publication Title: "open source"] OR [Publication Title: "ide"]]] OR [All: (] OR [All: (] OR [[Abstract: "remote sensing"] AND [[Abstract: "machine learning"] OR [Abstract: "deep learning"]] AND [[Abstract: "software"] OR [Abstract: "web interface"] OR [Abstract: "gui"] OR [Abstract: "open source"] OR [Abstract: "ide"]]]]

- ScienceDirect: ("webgis"OR "map service"OR "geoportal"OR "web based"OR "web application") AND ("remote sensing"OR "earth observation"OR "spatial analysis")

Faz-se uma ressalva de que a base de dados Springer apresentou limitação de quantidade de palavras-chave e chaves-lógicas e por isso é ligeiramente diferente das outras bases.

2.1.5 Critérios de inclusão e exclusão de artigos

A fim de garantir uma maior relação de conteúdo da proposta deste trabalho com a lista de artigos revisados, foram adotados critérios de inclusão ou exclusão de artigos do rol daqueles retornados nas buscas anteriormente citadas.

Os critérios de seleção de artigos empregados foram:

- Critérios para inclusão:
 - **I1:** Apresentação de ferramenta web para manipulação de dados georreferenciados.
- Critérios para exclusão:
 - **E1:** Artigos fora da área de sensoriamento remoto;
 - **E2:** Artigos sem aplicação em agricultura;
 - **E3:** Artigos que não pudessem ser acessados na íntegra;

Adotando-se estes critérios conseguiu-se filtrar e revisar apenas artigos relacionados com a proposta deste trabalho.

2.2 Condução

Nesta etapa os artigos foram selecionados e elencados por ordem de relevância. Baseado nos critérios acima apresentados, fez-se a leitura dos campos título e resumo de cada artigo com o fim de selecioná-los e classificá-los. No total, 1567 artigos foram obtidos, dentre os quais 274 foram selecionados. A partir destes, foi feita a análise de todo

corpo do artigo aplicando novamente critérios de inclusão e exclusão, resultando em 64 trabalhos selecionados. Na Tabela 2.1 é apresentada a distribuição de artigos selecionados em cada base de pesquisa. Durante a seleção, foram encontrados 100 artigos duplicados, os quais foram excluídos.

Tabela 2.1: *Quantidade de artigos selecionados por base de pesquisa*

Base de pesquisa	Quantidade
Scopus	41
IEEE	15
Springer	4
Science Direct	4
ACM	0
Total	64

Após a etapa de seleção, houve uma revisita ao conteúdo dos artigos com o fim de elencá-los por relevância segundo os seguintes critérios de qualidade:

- 1) Apresenta classificação de culturas ou cobertura do solo?
- 2) A classificação é feita baseada em *pixel*?
- 3) Descreve os métodos de classificação?
- 4) Descreve a arquitetura da ferramenta?
- 5) Apresenta métodos de processamento em nuvem?
- 6) Apresenta alguma aplicação?
- 7) Apresenta métricas de avaliação da classificação?

A etapa de ranqueamento dos artigos é importante pois aumenta a eficiência no processo de extração de informações.

2.3 Análise de dados

Para estabelecer uma relação entre os artigos selecionados, a seguir serão respondidas algumas questões de interesse deste trabalho tomando por base as informações extraídas.

Quais principais temas de aplicação de ferramentas WebGIS?

Para facilitar o entendimento, os temas dos artigos foram divididos em dois grandes grupos: 1) ferramentas com aplicação em monitoramento, ou seja, análises temporais e 2) desenvolvimento de ferramentas para outros fins.

Do total, 42 artigos focam em monitoramento, os quais também foram categorizados conforme a tabela 2.2. Dentre os 22 artigos restantes, 15 apresentam o desenvolvimento de uma ferramenta sem aplicação específica. E, os outros sete artigos apresentam

o desenvolvimento com aplicação, que aqui foi agrupada como "classificação e análise de culturas", conforme mostrado na tabela 2.3.

Tabela 2.2: *Categorias Monitoramento*

Categoria	Quantidade	Artigos
Cobertura do solo	22	[3], [5], [6], [7], [17], [23], [25], [36], [37], [41], [42], [51], [55], [57], [58], [84], [87], [97], [98], [99], [101], [104]
Produção da colheita	11	[1], [22], [33], [52], [63], [67], [69], [70], [78], [82], [95]
Seca	3	[60], [76], [89]
Desmatamento	2	[59], [92]
Umidade	2	[9], [94]
Regeneração de vegetação	1	[85]
Incêndios	1	[83]
Total	42	

Tabela 2.3: *Outras categorias*

Categoria	Quantidade	Artigos
Desenvolvimento de ferramenta sem aplicação definida	15	[8], [11], [20], [24], [34], [28], [50], [51], [53], [68], [72], [79], [80], [105], [103]
Ferramenta para classificação e análise de culturas	7	[21], [74], [43], [56], [88], [93], [96]
Total	22	

Neste contexto percebemos a grande quantidade de artigos voltados para estudos de análises temporais, principalmente com aplicação em análise de cobertura do solo, ao passo que há poucas iniciativas fora desse tema, e ainda menos com usos específicos voltados para análise e classificação de culturas.

Quais as principais plataformas utilizadas?

Dos trabalhos que apresentaram as fontes de imagens para análise, alguns usaram imagens comerciais, imagens providas de organizações de monitoramento de clima e processos atmosféricos, imagens de satélite do tipo radar, imagens de veículo aéreo não tripulado (VANT), entre outros. Uma vez que o intuito do trabalho é lidar com dados multiespectrais gratuitos, tais fontes não foram listadas na tabela 2.4.

A grande maioria dos artigos utilizam imagens disponibilizadas gratuitamente, principalmente da série de satélites *Landsat*, *Sentinel*, e Terra e Acqua. Na tabela 2.5 são listadas suas principais características:

Tabela 2.4: *Resumo de satélite e sensores utilizados nos artigos*

Fonte de imagem	Aplicações / Artigo	Ano de publicação
Landsat 4 TM	Cobertura do solo [42]	2019
Landsat 5 TM	Cobertura do solo [5], [6], [23], [41], [42], [98], [99]; Desmatamento [92]; Ferramenta [8];	2015-2022
Landsat 7 ETM+	Cobertura do solo [5], [23], [36], [41], [98], [99]; Desmatamento [92]; Produção [95]; Umidade [9]	2015-2022
Landsat 8 OLI	Cobertura do solo [5], [6], [23], [36], [98], [99]; Desmatamento [92]; Ferramenta [8], [28], [53]; Incêndio [83]; Produção [22], [70], [95]; Umidade [9]	2017-2022
Terra/Aqua MO-DIS	Classificação de cultura [43]; Cobertura do solo [7], [17], [25], [37], [58], [84], [104]; Desmatamento [59]; Ferramenta [11], [24], [34], [50], [53], [105]; Incêndio [83]; Produção [22], [33]; Seca [76], [89]	2015-2022
Sentinel 2 MSI	Classificação de cultura [93]; Cobertura do solo [3], [23], [86], [99]; Ferramenta [68], [28]; Incêndio [83]; Produção [63]	2017-2022

Tabela 2.5: *Resumo de características de fontes das imagens*

Fonte	Resolução Espacial (m)	Resolução Temporal (dias)	Imagens disponíveis	Situação
Landsat 4 TM	30-120	16	07/1982 até 01/2013	Desativado
Landsat 5 TM	30-120	16	07/1982 até 01/2013	Desativado
Landsat 7 ETM+	30-120	16	04/1999 até 05/2003	Ativo (porém com problemas)
Landsat 8 OLI	15-100	16	Desde 02/2013	Ativo
Terra Aqua MO-DIS	250-1000	1-2	Desde 05/2002	Ativo
Sentinel 2 MSI	10-60	5 com o 52A e 52B	Desde 05/2002	Ativo

Além disso, vale apresentar as principais combinações de bandas espectrais da luz utilizadas nos trabalhos e suas aplicações, as quais serão apresentadas como índices de vegetação, conforme tabela 2.6.

Tabela 2.6: *Bandas e índices utilizados*

Índices	Bandas espectrais utilizadas	Aplicações / Artigos
Enhanced vegetation index (EVI)	Infravermelho próximo (NIR), Azul e Vermelho	Desmatamento [92]; Cobertura do solo [6], [7], [25]; Ferramenta [11], [24], [50]
Green Normalized Difference Vegetation Index (GNDVI)	Infravermelho próximo (NIR) e verde	Produção [78]
Normalized Difference Red Edge (NDRE)	Red Edge (Vermelho de borda) e Vermelho	Classificação de cultura [56], [67]
NIR	Infravermelho próximo (NIR)	Cobertura de solo [23]
Normalized Difference Vegetation Index (NDVI)	Infravermelho próximo (NIR) e Vermelho	Classificação de cultura [21], [74], [43], [56], [93]; Cobertura do solo [3], [5], [6], [7], [25], [55], [37], [57], [58], [84], [99], [104]; Desmatamento [59], [92]; Ferramenta [8], [11], [24], [28], [34], [53], [105]; Produção [33], [52], [67], [70], [69]; Seca [60], [76], [90]
Normalized difference water index (NDWI)	Infravermelho próximo (NIR) e Infravermelho de ondas curtas (SWIR)	Classificação de cultura [56]; Cobertura do solo [6], [23], [99]
Normalized Vegetation Anomaly Index (NVAI)	Infravermelho próximo (NIR) e Vermelho	Seca [89]
Red Green Ratio Index (RGRI)	Vermelho e Verde	Ferramenta [11]
Soil-Adjusted Vegetation Index (SAVI)	Infravermelho próximo e Vermelho	Classificação de cultura [74], [56]
SWIR	Infravermelho de ondas curtas (SWIR)	Cobertura do solo [23]
Visible Atmospherically Resistant Index (VARI)	Azul, Verde e Vermelho	Classificação de cultura [56]; Cobertura do solo [58]
Vegetation Condition Index (VCI)	Infravermelho próximo (NIR) e Vermelho	Ferramenta [105]; Seca [60], [89]
Weighted Difference Vegetation Index (WDVI)	Infravermelho próximo (NIR) e Vermelho	Classificação de cultura [74]; Umidade [94]

Quais linguagens utilizadas?

Nesta seção serão listadas por meio da tabela 2.7 as linguagens de programação citadas nos artigos. Nem todos artigos definiram a linguagem para desenvolvimento ou detalharam o uso destas.

Tabela 2.7: *Linguagens utilizadas*

Linguagens	Artigos	Ano de Publicação
Javascript	[1], [23], [33], [36], [37], [51], [56], [58], [59], [72], [78], [79], [80], [90] [99], [101]	2015-2021
Python	[3], [8], [34], [51], [56], [57], [58], [68], [70], [72], [78], [86], [90]	2017-2021
Java	[7], [20], [25], [28], [57]	2016-2021
R	[20], [70], [85]	2017-2020
PHP	[1], [37], [80]	2015-2020
C#	[97], [98]	2017-2019
C++	[103]	2015

Uma vez que a área de pesquisa se trata de aplicações web, era esperado linguagens como *Javascript*, *Python* ou *PHP*. Porém como dito anteriormente, alguns artigos focaram em descrever a estrutura da ferramenta, outros na obtenção dos dados, outros focaram no desenvolvimento da página de internet, não sendo possível assim indicar a função da linguagem em cada artigo.

Quais principais tipos de arquiteturas implementadas?

Dos 64 artigos, apenas 36 apresentaram alguma descrição da arquitetura. Em resumo, 12 podem ser classificadas como arquiteturas simples com duas ou três camadas, incluindo: cliente, serviços *web* e servidor, com fluxo simples de informações. O restante apresenta arquiteturas complexas, isso se deve principalmente a integração de várias fontes de dados, resultando em arquivos de tipos diferentes e diversos tipos de manipulação de dados, com processamento local ou por meio de serviços externos.

Quais métodos de classificação utilizados?

Quanto aos métodos de classificação, os métodos mais citados são *Random Forest* (RF), *Classification And Regression Tree* (CART) e *Support Vector Machine* (SVM), conforme exposto na tabela 2.8.

Tabela 2.8: Métodos de Inteligência Artificial

Métodos citados	Aplicações / Artigos	Resultados
Random Forest	Classificação de cultura [93], [96]; Cobertura do solo [6], [98]; Ferramenta [105]; Produção [78]	Acurácia: 0.96[6]
Classification And Regression Tree (CART)	Classificação de cultura [96]; Cobertura do solo [6], [99];	Acurácia: 0.9783[99]
Support Vector Machine (SVM)	Classificação de cultura [96]; Cobertura do solo [6], [99]; Produção [78]	Acurácia: 0.813[96]
Minimum Distance	Cobertura do solo [6], [99]	Acurácia: 0.83[6]
GMO Maximum Entropy classifier	Cobertura do solo [6]	Acurácia: 0.94
Gradient Boosting Decision Tree	Classificação de cultura [96]	Acurácia: 0.781
Naïve Bayes	Cobertura do solo [6]	Acurácia: 0.89

Quais métricas de avaliações utilizadas?

Dentro do contexto de resultados é de suma importância apresentar métricas de avaliação. Por meio da revisão dos artigos percebeu-se que muitos deles não apresentam nenhuma métrica de avaliação. Dentre os que apresentam, podemos dividir em três categorias:

- 1) Checagem ponto a ponto por meio de usuários: bastante custoso e demorado, há a necessidade de voluntários, ou seja, não é escalável;
- 2) Experimentos: em geral verifica-se o tempo de execução do algoritmo ou o consumo de memória, porém é um processo de difícil padronização de experimentos para equiparação com trabalhos correlatos;
- 3) Métricas por meio de cálculos matemáticos: podemos citar várias delas, acurácia, coeficiente *Kappa*, *Area under the curve* (AUC) entre outros. Estas apresentam a vantagem de já estarem implementados nos classificadores e possibilitam a comparação com artigos relacionados.

Quais serviços de nuvem utilizados para coleta de imagens?

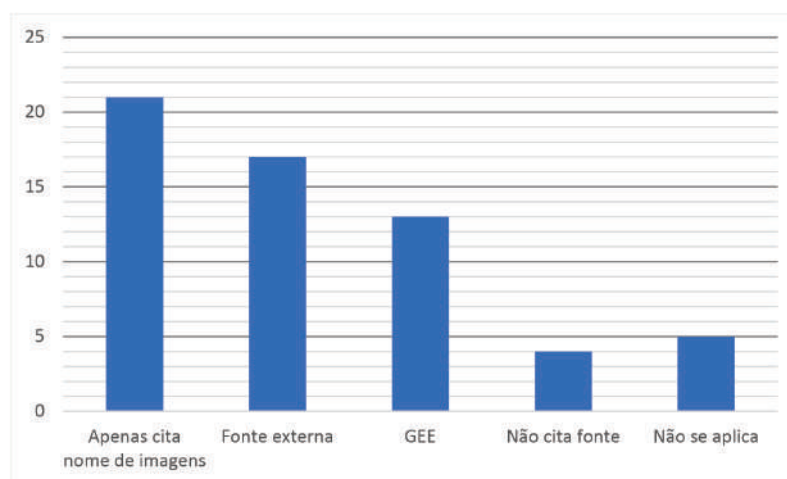
Como dito anteriormente, há uma tendência de migração de aplicações para a internet, incluindo também a coleta e processamento. Por isso foi levantado quantos artigos utilizam serviços em nuvem para essas etapas, e quais serviços são estes, conforme a tabela 2.9.

Tabela 2.9: *Serviços em nuvem*

Nomes dos serviços	Aplicações / Artigos
Google Earth Engine	Classificação de cultura [93]; Cobertura do solo [5], [6], [23], [36], [57], [84], [99], [101]; Desmatamento [92]; Ferramenta [24], [53]; Incêndio [83]; Produção [22], [70], [69]; Seca [76]
Amazon Web Services (AWS)	Classificação de cultura [56]; Ferramenta [72]; Produção [52]
ArcGis Server	Producao [33], [52]; Umidade [9]

Como é feita a aquisição de imagens?

Por se tratar de ferramentas para classificação de imagens espectrais, é esperado que haja a apresentação de como foram obtidos os dados para classificação. Porém, muitas vezes as fontes são apenas citadas, e o processo de obtenção não é abordado conforme apresentado na Figura 2.1.

**Figura 2.1:** *Gráfico sobre aquisição de imagens.*

O processo de aquisição por fonte externa muitas vezes se constitui da visita manual a portais com acesso às imagens gratuitas como *EarthExplorer*, ⁶, *Copernicus Open Access Hub* ⁷, entre outros, em que o usuário necessita de um cadastro, que pode ser realizado no mesmo dia, ou em algumas outras plataformas pode demorar alguns dias para validação. Dentro do site, o usuário precisa localizar a área de estudo por meio da interface ou adicionando coordenadas. As datas precisam ser definidas, bem como o sensor desejado. A depender da quantidade de imagens necessárias para o estudo esse processo pode ser bastante demorado.

⁶Disponível em: <https://earthexplorer.usgs.gov>

⁷Disponível em: <https://scihub.copernicus.eu/dhus/>

Dos 64 artigos analisados, 17 utilizam a plataforma GEE. Quatro utilizam as imagens dentro da própria plataforma e 13 fazem o *download* das imagens, para depois inserir em seus sistemas. Desse último grupo, a maior parte apenas cita o nome da plataforma, sem mencionar o processo de aquisição, se foi por meio de *script*⁸ ou foi feito interativamente por meio da interface gráfica oferecida.

2.3.1 Considerações finais da revisão

Em resumo, as ferramentas online para lidar com dados espaciais estão focadas em análises temporais, ou seja, monitoramento, com foco em cobertura do solo. As imagens para execução das análises em geral são de acesso gratuito e escolhidas de acordo com a aplicação. Quando o estudo é regional ou global, opta-se por imagens do sensor *MODIS*, com resolução espacial de 250m a 1000m. Por outro lado, em estudos mais localizados pode-se optar pelas imagens pertencentes a série de sensores embarcados nos satélites *Landsat* ou *Sentinel*, que apresentam melhor resolução espacial, temporal e espectral.

Dentre os algoritmos de classificação, *Random Forest*, CART e SVM continuam se destacando quanto à frequência de utilização e resultados. Porém, durante a análise dos artigos, devido à diversidade de aplicações, percebeu-se a dificuldade de padronização de métricas para fins de comparação de resultados. Possivelmente em uma seleção de artigos com filtro de aplicação mais restrita, haja maior facilidade em comparar trabalhos e resultados.

Quanto às arquiteturas, há uma tendência por sistemas mais complexos com uso de serviços externos, principalmente para aplicações em classificação de cobertura do solo, devido a quantidade de informações provindas de diversas fontes, e para reutilizar serviços já desenvolvidos e disponíveis.

Em relação à fonte de dados, fica claro que o processo de aquisição de imagens é um assunto pouco abordado, muitas vezes não citado, ou quando apresentado, feito de maneira breve e sem detalhamento. Mesmo em um cenário em que a maioria dos artigos lida com análises temporais, necessitando de várias imagens gerando o custo de armazenamento e tempo para aquisição dos dados, o processo é raramente mencionado.

⁸Linhas de códigos contendo instrução para o computador realizar.

Fundamentação teórica

3.1 Sistema de Informações Geográficas - SIG

Na literatura há muitas definições para SIG (do inglês *Geographic Information Systems* - GIS), algumas delas serão citadas e discutidas a seguir.

De acordo com Cowen [19], "SIG é um sistema de apoio à decisão envolvendo a integração de dados espacialmente referenciados em um ambiente de solução de problemas". Ainda segundo Rigaux et al. [71], SIG é definido como "uma ferramenta que armazena dados geográficos, recupera e combina esses dados para criar novas representações do espaço geográfico, fornece ferramentas para análise espacial e realiza simulações para ajudar usuários experientes a organizar seu trabalho em muitas áreas". Bolstad [10] se referiu a SIG como "um sistema baseado em computador para auxiliar na coleta, manutenção, armazenamento, análise, produção e distribuição de dados e informações espaciais".

Além destas, há muitas outras definições por outros autores. De acordo com Cowen [19], essa grande quantidade de definições se deve à multidisciplinaridade e à diversidade de aplicações em diferentes áreas da ciência.

Em seu livro, Gonzáles et al. [12], de modo muito didático, apresenta cinco componentes que constituem um SIG. São eles:

- **Hardware:** compõe os componentes físicos do sistema, como, por exemplo, discos de armazenamento, processadores, memórias para processamento, painel de visualização, etc.
- **Software:** é o sistema composto por diferentes funcionalidades, gerados por algoritmos para consultar e manipular os dados. Nesse quesito, há muitas companhias comerciais que oferecem serviços voltados a SIG, e Gonzáles et al. enfatiza, ainda em 2012 [12], o surgimento de muitos *softwares* livres, disponíveis gratuitamente.
- **Dados:** são as informações que serão consultadas, armazenadas e analisadas. Possuem as características específicas do objeto de estudo e carregam consigo a localização da informação.

- Organização de procedimentos: definições de objetivos, processos, controles de qualidade, etc. Muito semelhante ao conceito de regra de negócios que definirá a interação entre diferentes tipos de dados e quais procedimentos podem ser aplicados a eles.
- Usuários: podem ter formações específicas em SIG ou ser pesquisadores de outras áreas. Esse componente humano é tido como o mais instável do sistema, pois condicionará o funcionamento e a confiabilidade das análises.

Ainda sobre o componente dos dados, Câmara et al. [14] os define como uma representação do mundo real, e os segmenta em duas classes: representações vetoriais ou matriciais, como mostrado nas Figuras 3.1 e 3.2.

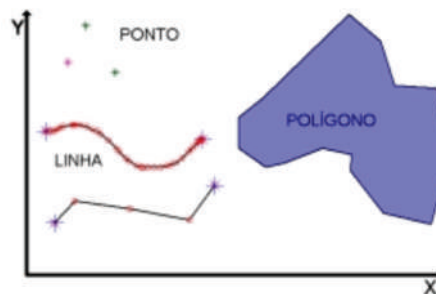


Figura 3.1: Exemplo de dado vetorial. Extraído e adaptado de: *Conceitos básicos em ciência da geoinformação*. [15]

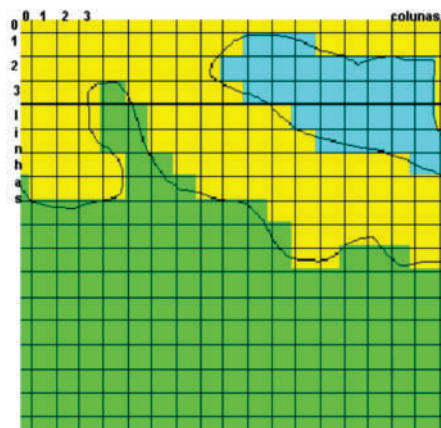


Figura 3.2: Exemplo de dado matricial. Extraído e adaptado de: *Conceitos básicos em ciência da geoinformação*. [15]

A representação vetorial é baseada em pares de coordenadas. Há três tipos de elementos: ponto, linha e polígono. O ponto é constituído por um par ordenado (x, y). As linhas ou elementos lineares são um conjunto de pares ordenados. E um polígono é a região delimitada por linhas conectadas, de forma que o primeiro ponto coincide com o último. Já a representação matricial é formada por uma matriz de linhas e colunas, em que cada pixel corresponde à determinada característica de estudo [14].

3.1.1 Aplicações SIG na internet

Em 1993, Putz foi um dos pioneiros a desenvolver sistemas SIG no ambiente da internet [66]. Putz inovou quanto à interatividade, apresentou uma ferramenta dentro de um site capaz de visualizar e também produzir informações espaciais a partir de arquivos estáticos.

Em 1994 fundou-se a *Open Geospatial Consortium* (OGC), uma comunidade mundial focada em definir padrões de desenvolvimento e implementação de conteúdos e serviços geoespaciais, processamento de dados SIG, serviços de localização, sensores e modelos de dados [91]. A partir daí muitas novas ferramentas foram disponibilizadas e passou-se a seguir padrões pré-estabelecidos de desenvolvimento para promover melhor interação entre serviços online.

Devido à possíveis implicações sociais, muitos pesquisadores começaram a indagar sobre um elitismo envolvendo tal tecnologia, principalmente, pelo fato de, países em desenvolvimento na época terem menor acesso à computadores [64], iniciativas visando a participação pública em aplicações SIG (*public participation GIS - PPGIS*) começaram a surgir. Os primeiros esforços consistiam em envolver um grande grupo de pessoas em processos de tomada de decisão. Com o advento da internet, a tomada de decisão em conjunto baseada em SIG se tornou muito mais fácil [61].

Em 1998, Hardie apresentou os fatores que possibilitaram o desenvolvimento desta área, desde a criação de navegadores de internet e o *HTML*¹ e suas versões, até soluções para interfaces de usuário como *Java*², *ActiveX*³ e *Scripts*[38]. Por meio dessas novas tecnologias, não se fazia mais necessário recarregar páginas, e os conteúdos passaram a ser dinâmicos.

Em 2013, Liu et al. [48] discorreu sobre a importância de levar as aplicações SIG para a internet. Uma das principais vantagens citadas é a facilidade de compartilhamento de dados espaciais com outras pessoas por meio da internet. Além disso, muito mais pessoas têm acesso a essa tecnologia, a qual antes era utilizada apenas por especialistas.

Outro avanço importante para os sistemas SIG online são as soluções de arquitetura, citadas também na revisão, anteriormente apresentada. Em seu artigo, Shi [81] abordou as arquiteturas baseadas em serviços, as quais aprimoraram em muito a questão de processamento. Rotinas anteriormente de alto custo de processamento e tempo, foram transferidas para serviços online. De acordo com Shi, esse é o conceito chave para a computação em nuvem.

¹Linguagem de Marcação de HiperTexto, em resumo se refere aos links que conectam entre si as páginas da internet.

²Java é uma linguagem de programação orientada a objetos

³Um framework que possibilita aplicações em páginas de internet.

3.2 Sensoriamento Remoto

Da mesma forma que SIG, há muitas definições para o SR, das quais apresentaremos algumas a seguir. De acordo com Fischer et al. [26], "sensoriamento remoto foi definido de várias maneiras, mas basicamente é a arte ou ciência de dizer algo sobre um objeto sem tocá-lo". Como o próprio autor afirma, essa é uma definição ainda bastante genérica.

Devido à grande quantidade de definições, sendo que muitas delas não são amplamente aceitas, Fussel et al. [29] comparou o histórico de definições apresentadas, concluindo que uma boa definição deve abordar as principais características: aquisição sem contato, uso da propriedade da radiação eletromagnética (REM), uso de instrumentos, plataformas fixas ou móveis, com tratamento de dados coletados e uso de técnicas de interpretação.

Podemos citar a definição de Schowengerdt [77], "sensoriamento remoto é definido, para nossos propósitos, como a medição das propriedades dos objetos na superfície da terra usando dados adquiridos de aeronaves e satélites".

Para facilitar o entendimento, Lillesand et al. [47] faz uma analogia de SR com a visão humana. Quando um texto é lido, os dados que os olhos obtém são correspondentes a quantidades de luz refletidas das porções pretas e brancas na página. O cérebro, assim como um sensor acoplado em um satélite, é capaz de obter essas informações, analisá-las e traduzí-las em informações interpretáveis. Nesse exemplo, essas informações correspondem às letras. O conjunto de informações torna possível identificar palavras, frases, parágrafos e textos completos, gerando as interpretações.

Por outro lado, há uma grande diferença entre os sensores de satélites e os olhos, a quantidade de informações obtidas em cada registro. Para o melhor entendimento dessa diferença é necessário um pouco mais de detalhe sobre a energia eletromagnética.

3.2.1 Energia eletromagnética

De acordo com Liu [49], há três propriedades importantes no que se refere a energia eletromagnética: comprimento de onda (λ), frequência (ν) e amplitude (A). O comportamento das três propriedades podem ser visualizadas na figura 3.3. O comprimento de onda se refere à distância entre picos. A frequência é a quantidade de ondas que por passam por um ponto fixo por segundo, medida em Hertz (Hz). Sendo um Hz referente à uma oscilação por segundo. E a amplitude se refere a altura de cada pico.

Em resumo, uma fonte de energia, como por exemplo o sol, emite energia eletromagnética que se propaga pela atmosfera. Ao interagir com algum objeto na superfície terrestre ou atmosfera, parte dessa energia pode ser absorvida ou refletida. Sensores são utilizados para registrarem essa energia refletida. Com base nas configurações de onda

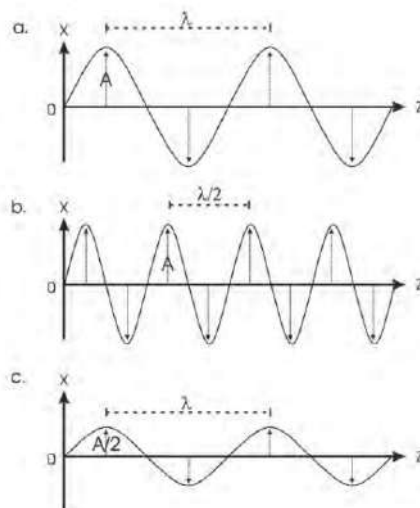


Figura 3.3: *Propriedades de onda da energia eletromagnética. Sendo a) e c) ondas com uma dada frequência ν e b) 2ν . Extraído de: Aplicações de sensoriamento remoto [49]*

provenientes dessa interação esses dados são tratados e transformados em imagens. Além disso, as quantidades de energia absorvidas também são importantes, pois definem informações sobre a composição dos alvos [54].

A luz visível, bem como suas cores, é só uma das muitas formas da energia eletromagnética. Ondas de rádio, raios ultravioletas, raio-x são outras formas amplamente conhecidas. Todas essas ondas fazem parte do espectro eletromagnético, mostrado na figura 3.4. Pela figura podemos observar que a frequência que corresponde a luz visível corresponde a um intervalo muito pequeno do espectro, em torno de 400 a 700nm.

Ao contrário dos olhos humanos, os sensores acoplados nos satélites são capazes de detectar uma maior variedade de frequência de ondas. Na figura 3.5, são listados três satélites (Landsat 7, Landsat 8 e Sentinel 2). O intervalo de comprimento de onda próximo de 400 e 750 é o correspondente ao visível ao olho humano. Cada um dos retângulos com números indica uma banda espectral referente ao satélite em questão. As bandas definem intervalos de comprimento de onda a partir dos quais o sensor é apto para obter dados espectrais.

3.2.2 Imagens de Sensoriamento Remoto

A depender da aplicação, deve-se escolher quais fontes de imagens serão utilizadas. A escolha pode considerar as características do objeto de estudo analisado. Lillesand et al. [47] citou algumas, como: geometria, tamanho, padrão de repetição, textura, sombras, localização. Pode-se também avaliar as características das imagens disponíveis, em geral comparam-se quatro aspectos principais:

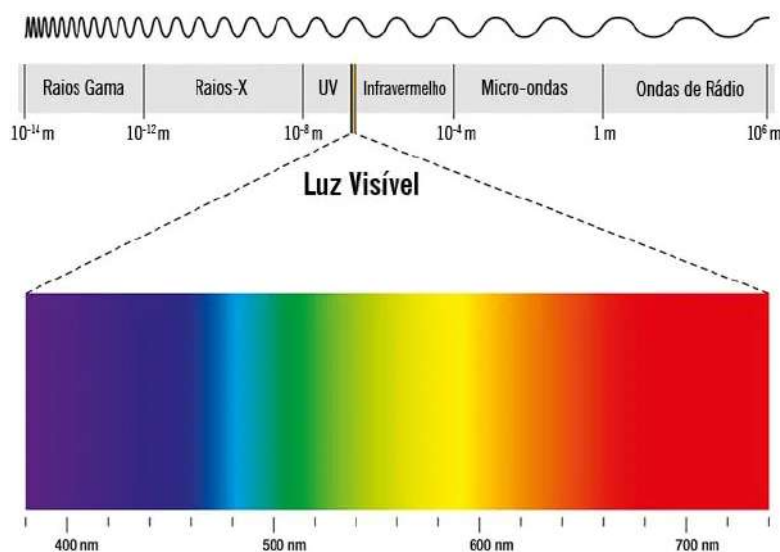


Figura 3.4: *Espectro eletromagnético. Extraído de:*
<https://www.todamateria.com.br/espectro-eletromagnetico/>

- Resolução espacial: define o quão pequeno um objeto pode ser a ponto de o sensor detectá-lo, diferenciando-o dos objetos ao redor. De modo prático, quanto maior a resolução espacial, mais detalhes a imagem terá, consequentemente aumentando espaço necessário para armazenar esta imagem. A Figura 3.7 ilustra diferentes resoluções espaciais.
- Resolução radiométrica: define a capacidade de o sensor diferenciar pequenas variações de brilho da imagem, também relacionada ao tamanho do intervalo da escala de cinza. A unidade de medida para essa resolução é o bit, e quanto menor seu valor, menos detalhe a imagem terá. O intervalo da escala de cinza pode ser calculado usando a potência 2 com base na quantidade de bits da imagem. Por exemplo, uma imagem com 16-bits terá 256 diferentes tons de cinza. Pode-se observar a diferença de detalhes comparando as imagens da Figura 3.6.
- Resolução espectral: refere-se à quantidade de bandas espectrais, largura e posicionamento no espectro eletromagnético, como foi mostrado na Figura 3.5.
- Resolução temporal: define a quantidade de tempo para que haja uma revisita do sensor ao mesmo ponto no globo terrestre. As diferentes resoluções temporais de alguns satélites foram apresentadas na Tabela 2.5.

3.2.3 Imagens multiespectrais

As imagens multiespectrais são dados de representação matricial, compostos por malhas sobrepostas geradas a partir de medições do processo físico da radiação

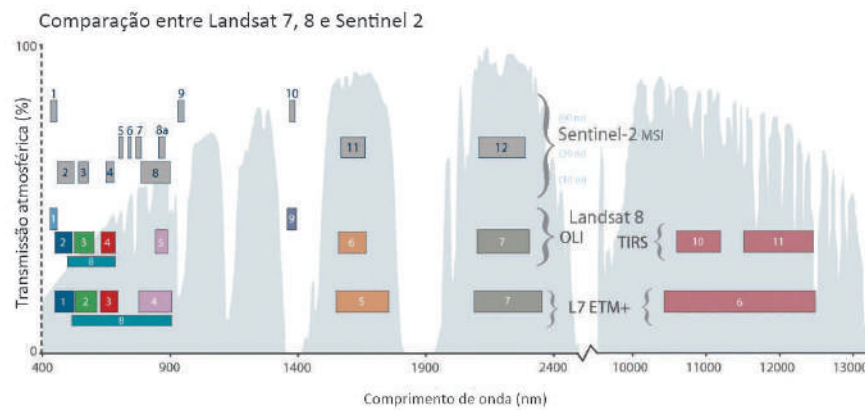


Figura 3.5: Traduzido de: <http://landsat.gsfc.nasa.gov/?p=10643>. Ilustrando diferentes resoluções espectrais.

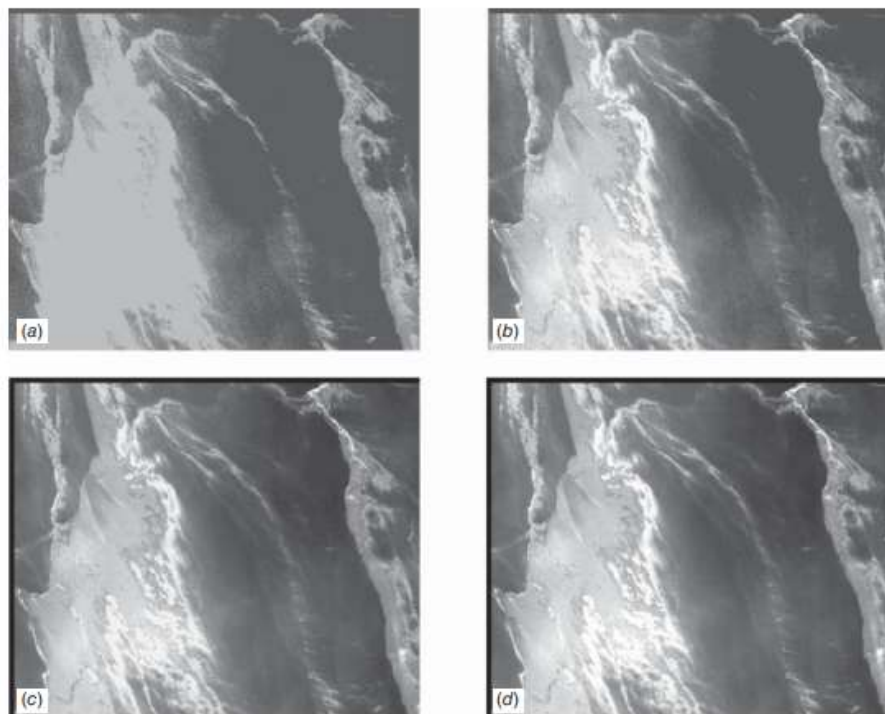


Figura 3.6: Extraído de: *Remote Sensing and image interpretation* [47]. As 4 imagens têm mesma resolução espacial porém diferentes resoluções radiométricas, com valores de 1-bit, 2-bits, 4-bits e 8-bits, respectivamente nos quadros a), b), c) e d).

eletromagnética. Os dados são captados pelo sensores os quais convertem medições analógicas em valores digitais. Cada valor é associado a um *pixel*, que define a área unitária dentro da imagem. Os *pixels* são organizados em linhas e colunas, e seus valores medidos correspondem a média dos valores dos diferentes alvos que podem estar contidos dentro de sua área [54].

A Figura 3.7, extraída do trabalho de Lillesand et al. [47], apresenta diferentes imagens com diferentes resoluções espaciais. Como pode-se observar, as imagens estão

em escala de cinza e representam apenas uma banda espectral. Uma imagem multiespectral é composta por várias camadas sobrepostas referentes a cada banda espectral, assim como mostrado na Figura 3.8 .

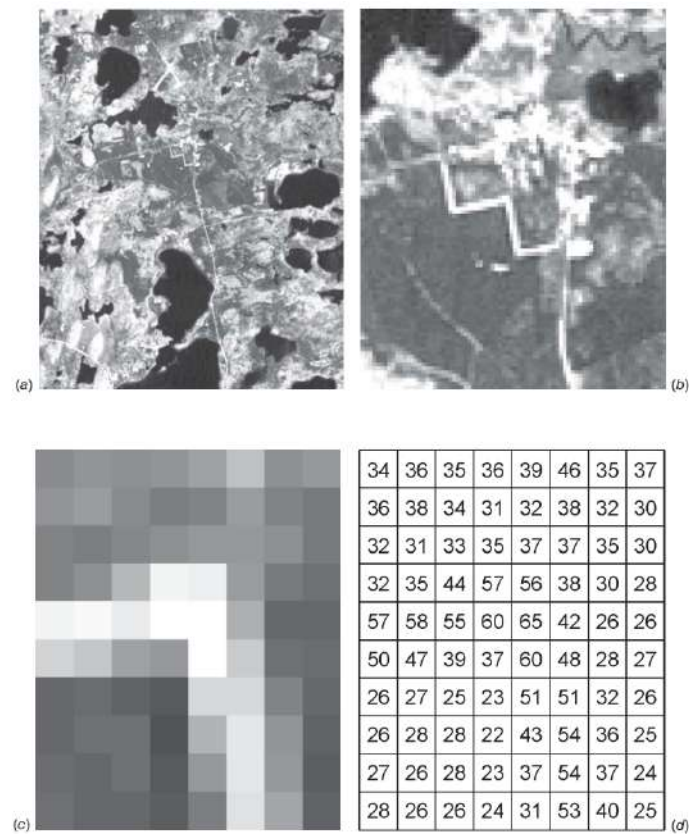


Figura 3.7: Extraído de: Remote Sensing and image interpretation [47]. Ilustrando pixels em imagens e seus valores digitais. Em a) temos uma malha de 500 por 400 pixels, em b) temos 100 por 80 e em c) 10 por 8. A última figura em d) ilustra os valores dos pixels em c).

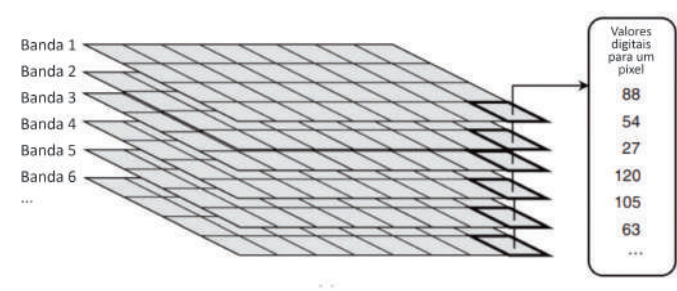


Figura 3.8: Traduzido de: Remote Sensing and image interpretation [47]. Ilustrando sobreposição de bandas espectrais.

3.2.4 Índices espectrais

Os índices espectrais referem-se às camadas sintéticas de números digitais, adicionadas às outras camadas da imagem multiespectral. Os valores dos *pixels* são definidos a partir de combinações e manipulações das bandas originais.

Desde 1983, Jackson [40] defendia a importância dos índices espectrais para a discriminação de vegetação e solo. Ainda em 1995, Xue [100] fez uma revisão sobre índices de vegetação, ou seja combinações de bandas específicas para análise espectral de vegetação. A seguir são apresentados alguns dos índices de vegetação convencionais e suas equações:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (3-1)$$

$$EVI = 2.5 \frac{NIR - RED}{(NIR + 6RED - 7.5BLUE) + 1} \quad (3-2)$$

$$SAVI = 2.5 \frac{800nm - 670nm}{800nm + 670nm + L} (1 + L) \quad (3-3)$$

Sendo $L = 0,5$.

$$GNDVI = \frac{NIR - [540 : 570]}{NIR + [540 : 570]} \quad (3-4)$$

$$GNDVI = \frac{860nm - 1240nm}{860nm + 1240nm} \quad (3-5)$$

Os valores desses índices em geral podem ser usados para análise de qualidade de vegetação, quantidade de umidade, desenvolvimento, etc. Podem derivar entre si, apresentando tratamentos de dados para situações específicas. Por exemplo, o EVI deriva do NDVI, porém com tratamentos para área com vegetação densa ou determinadas condições atmosféricas. Já o SAVI, também derivado do NDVI, se diferencia por realizar tratamento de dados para driblar instabilidades em análise com presença de solo exposto.

3.2.5 Classificação de imagens de SR

Cheng et al. [18] discutiram sobre três tipos de classificação espectral na área de SR com escalas diferentes: a nível de *pixel*, objeto e cenas. Tal diferenciação pode ser observada nas aplicações apresentadas na Figura 3.9. Através da imagem fica claro como cada tipo de classificação deve ser utilizada para objetivos diferentes.

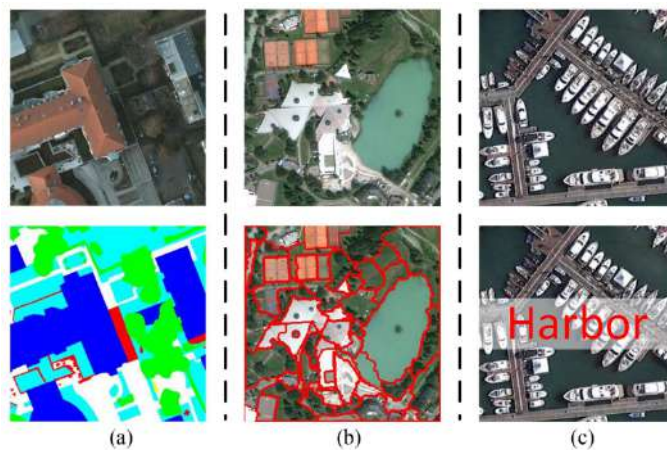


Figura 3.9: Extraído de: *Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities*, [47]. Diferentes tipos de classificação. a) *Nível de pixel*, classificação foca em atrelar cada pixel à uma classe. b) *Nível de objeto*, classificação foca em reconhecer e segmentar objetos na imagem. c) *Nível de cena*, classificação foca em reconhecer o contexto da imagem.

3.2.6 Classificação supervisionada baseada em pixel

Bhaskaran e Perumal [62] em 2010, descreveram esse tipo de classificação como o processo de atribuição de *pixels* ou unidades básicas de uma imagem à diferentes classes. A grosso modo, é como se fossem agrupados *pixels* com informações espectrais idênticas. Após isso, com base na semelhança de características espectrais, é determinado à qual classe cada grupo corresponde.

3.3 Inteligência Artificial em Sensoriamento Remoto

Uma análise dos algoritmos de aprendizado de máquina com aplicações em classificação de imagens de satélites foi feita por Ghamisi et al. [31], onde se observou que os métodos mais citados nos últimos anos foram: *support vector machine* (SVM), *artificial neural network* (ANN), *random forest* (RF), *extreme learning machine* (ELM), *multinomial logistic regression* (MLR), *Kernel ELM* (KELM). Até o ano de publicação de sua pesquisa o método *Convolutional Neural Network* (CNN) e *Light Gradient Boosting Machine* (LightGBM) eram recentes na área e por isso não entraram para esta lista.

3.3.1 Random Forest

O método de classificação *Random Forest* (RF) é um algoritmo de classificação supervisionado do tipo *ensemble*, ou seja, é constituído por múltiplos classificadores, o

qual foi apresentado por Breiman [13]. Resumidamente, de acordo com Khoshgoftaar et al. [45], as principais etapas para execução desse método são:

- 1) Criar um subconjunto B da amostra total chamada *bootstrap* com o mesmo tamanho do conjunto original.
- 2) Construir árvores de decisão baseadas no conjunto B porém selecionando n características aleatórias, sendo n menor ou igual ao número de características.
- 3) Repetir os passos anteriores quantas vezes forem definidas pelo usuário.
- 4) Classificar instâncias usando as árvores criadas. Cada árvore gerará uma classificação. A classe mais predita pelas árvores será a predição final do método RF.

Este método se destacou pois diferentemente de uma árvore de decisão comum, a entrada de dados não é realizada com o conjunto total de amostras mas sim um subconjunto aleatório deste. Tal processo cria uma grande quantidade de árvores geradas com amostragem aleatória a partir do conjunto original, reduzindo o sobreajuste (do inglês *overfitting*) do modelo [35].

3.3.2 Light Gradient Boosting Machine (LightGBM)

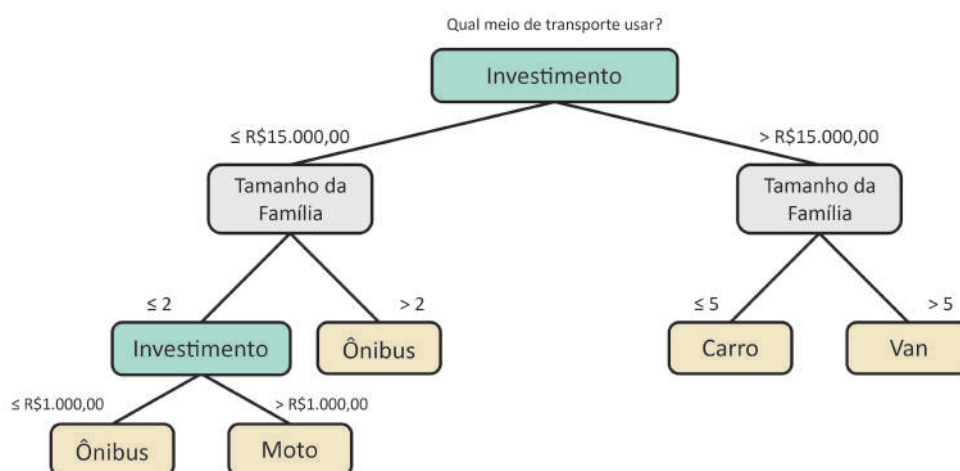
Um dos algoritmos de classificação mais usados atualmente é o LightGBM. Ele é resultado do aperfeiçoamento do algoritmo de Árvore de Decisão (do inglês *Decision Tree* - DT). A primeira melhoria do DT nesse processo de aperfeiçoamento foi com o algoritmo *Gradient Boosting Decision Tree* (GBDT). E posteriormente outras alterações foram feitas culminando no *LightGBM*.

Em resumo, o algoritmo de árvore de decisão consiste em uma árvore binária gerada a partir de condições baseadas em um conjunto de dados inicial. A raiz e os nós filhos representam condições em função das instâncias a serem estudadas. Os nós adjacentes se referem a resposta verdadeira ou falsa dada a condição do nó pai e os nós folhas indicam um rótulo da classe predita. As condições são selecionadas de modo a usar as instâncias mais relevantes para a classificação e para isso podem ser usadas técnicas de seleções como: *Information Gain*, *Gini Index*, entre outros [4]. Um exemplo pode ser visto na figura 3.10 referente a tabela 4.1. No entanto, o desempenho das árvores de decisão pode ser substancialmente melhorado usando métodos como *boosting* [27].

Com o GBDT, proposto por Gareth et al. [30], cada árvore é cultivada usando informações de árvores previamente construídas. Dado o modelo atual, ajustamos uma árvore de decisão aos resíduos do modelo, ou seja, usando a diferença entre o alvo real e os valores previstos como alvo em vez do alvo observado. A ideia é aprender com os dados lentamente, ao contrário de ajustar uma única árvore de decisão para todos os dados. O GBDT é um algoritmo de aprendizado de máquina popular e possui algumas implementações eficazes, como XGBoost e *Parallel Gradient Boosted Regression Trees*

Tabela 3.1: Tabela com dados fictícios

Investimento	Tamanho da família	Veículo
6000	1	Moto
10000	5	Ônibus
20000	2	Carro
18000	4	Carro
16000	7	Van
3000	2	Moto
500	1	Ônibus

**Figura 3.10:** Ilustração sobre fluxo do algoritmo de Árvore de Decisão

(pGBRT). O principal custo do GBDT está no aprendizado das árvores de decisão, e a parte mais demorada do aprendizado de uma árvore de decisão é encontrar os melhores pontos de divisão.

Uma ideia simples é reduzir o número de instâncias e o número de variáveis preditoras. No entanto, isso acaba por ser altamente não trivial. Para este objetivo, o modelo *LightGBM* propõe duas novas técnicas conhecidas como *Gradient-based One-Side Sampling* (GOSS) e *Exclusive Feature Bundling* (EFB). O primeiro trata de reduzir aleatoriamente os dados enquanto cuida do ganho de informação, enquanto o segundo procura reduzir o número de variáveis preditoras com base na esparsidade. O modelo *LightGBM* pode acelerar o processo de treinamento em até 20 vezes, alcançando quase a mesma precisão dos métodos GBDT anteriores [44]. De maneira bastante resumida, a construção das árvores de decisão não mais cria árvores completas, evitando nós não necessários para a análise, dessa forma são geradas menos instâncias e há uma diminuição da quantidade de variáveis preditoras, como pode ser visto na figura 3.11.

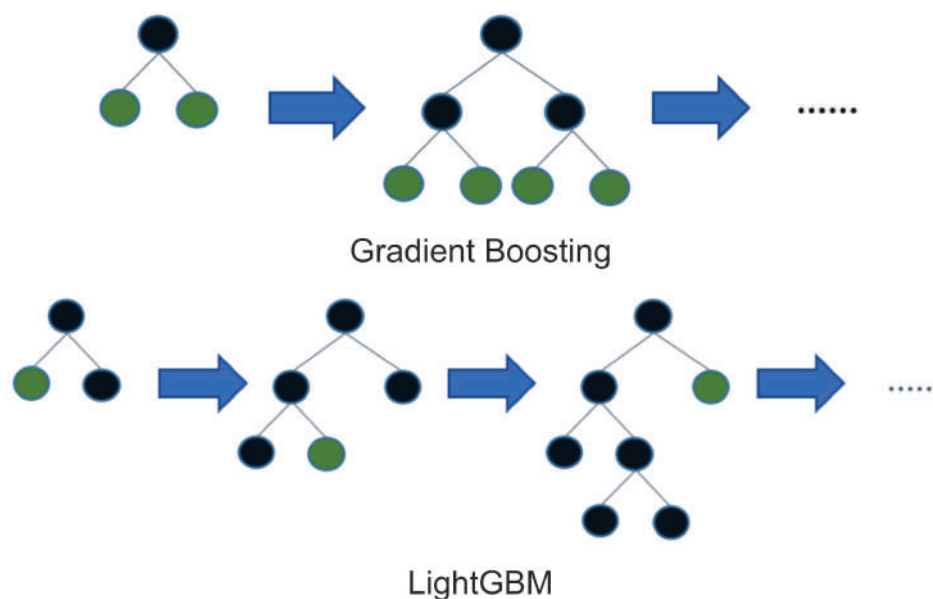


Figura 3.11: Extraída e adaptada da documentação do algoritmo LightGBM (disponível em: <https://lightgbm.readthedocs.io/en/v3.3.2/>). Ilustração apresentando a diferença na construção de árvores de decisão.

3.3.3 Automated Machine Learning (AutoML)

Uma vasta revisão foi realizada por [39] et al. sobre uma nova abordagem para condução de processos de classificação por meio de reconhecimento de padrão designada AutoML. Há uma grande complexidade envolvida nas várias etapas de um processo de classificação por meio de técnicas de Inteligência Artificial, principalmente ao lidar com modelos de aprendizado profundo. Ou seja, conhecimentos sobre estatística e algoritmos de classificação se tornam indispensáveis, ao lidar com cada uma dessas etapas, por exemplo, obtenção de dados, limpeza de dados, balanceamento, treinamento de modelo, validações, ajustes de parâmetros, métricas de avaliação de resultados, etc.

De acordo com Zöller [106], essa nova abordagem foca em otimizar esse fluxo por meio de automatização de cada um desses processos. Para especialistas da área de reconhecimento de padrões há um ganho de eficiência nas diferentes etapas, as quais muitas vezes são refinadas repetidamente por meio de tentativa e erro. Além disso, profissionais de outras áreas da ciência são incluídas e podem usufruir dos benefícios dessas técnicas, mesmo não sendo sua principal especialidade, e não mais limitadas a dependência de profissionais especializados em reconhecimento de padrões.

Material e métodos

4.1 Descrição da ferramenta

4.1.1 Escolhas de desenvolvimento

Para o desenvolvimento da ferramenta optou-se por utilizar *Python*, uma vez que é uma linguagem que está em constante desenvolvimento e há uma série de pacotes que a tornam muito versátil. O *framework* utilizado foi o *Django* já bastante conhecido e utilizado em diversos tipos de projetos.

A linguagem *Python* também possibilitou a integração do sistema com serviços de mapa do *Google Maps* e do *Google Earth Engine* para coleta de imagens. Uma tabela com as principais bibliotecas utilizadas, bem como com as principais funcionalidades é apresentada abaixo.

Tabela 4.1: Tabela com principais bibliotecas utilizadas

Nome	Funcionalidades
Rasterio	Ler, editar e criar imagens com referência geográfica.
Pillow	Ler, editar e criar imagens.
Matplotlib	Criar visualizações, estáticas ou dinâmicas. Ex.: gráficos.
Fiona	Ler, editar e criar dados com referência geográfica.
Pandas	Análise de dados, manipulação de vetores e matrizes.
Numpy	Manipulação de números, vetores, matrizes, etc.
Pycaret	Conjunto de algoritmos de Aprendizado de Máquina.

Como mostrado na tabela acima, as quatro primeiras bibliotecas apresentadas fazem referência à operações relacionadas a SIG. O restante foi utilizado em processos de tratamento e análise de dados e treinamento de modelos e classificação.

As imagens do produto Level-2A, do sensor *MSI* e satélite *Sentinel 2* foram escolhidas, uma vez que o acesso é gratuito, os dados são dados pré-tratados e há uma melhor resolução espectral, espacial e temporal em relação a outros sensores gratuitos. Futuramente pode-se adicionar imagens de outros sensores, porém seriam necessários processos para integração de dados e fusão de imagens com fontes diversas.

4.1.2 Arquitetura do sistema

A plataforma inclui quatro componentes principais entre cliente e servidor, ilustrados na Figura 4.1. O lado do cliente refere-se ao *front-end*, o qual disponibiliza uma interface gráfica de usuário (*Graphical User Interface* - GUI, em inglês) intuitiva, que pode ser vista na Figura 4.3. Nela o usuário pode cadastrar grupos de regiões amostrais, selecionar regiões através de um mapa interativo - definindo um nome e um intervalo de datas - e também fazer a classificação de áreas selecionadas. O lado do servidor, comumente tratado como *back-end*, é responsável por coletar as informações requisitadas pelo usuário usando o serviço do *Google Earth Engine* como fonte de dados, armazenar as informações no banco de dados, fazer o treinamento de modelos de classificação e também fazer a predição de classes das áreas selecionadas pelo usuário.



Figura 4.1: Arquitetura da ferramenta.

Para o banco de dados, foi escolhido o *SQLite* devido à simplicidade de implementação e a facilidade de integração com o *framework Django*. O modelo de classes para armazenamento de dados pode ser visualizado abaixo na Figura 4.2.

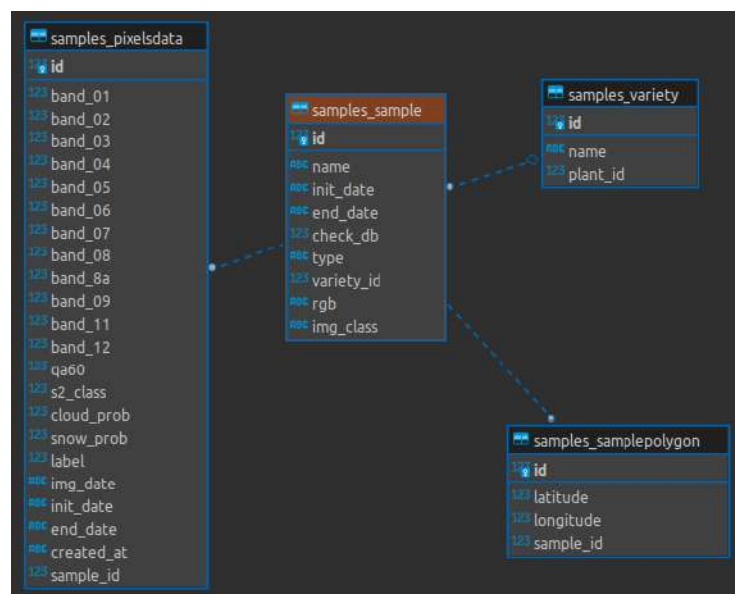


Figura 4.2: Modelo de classes implementado.

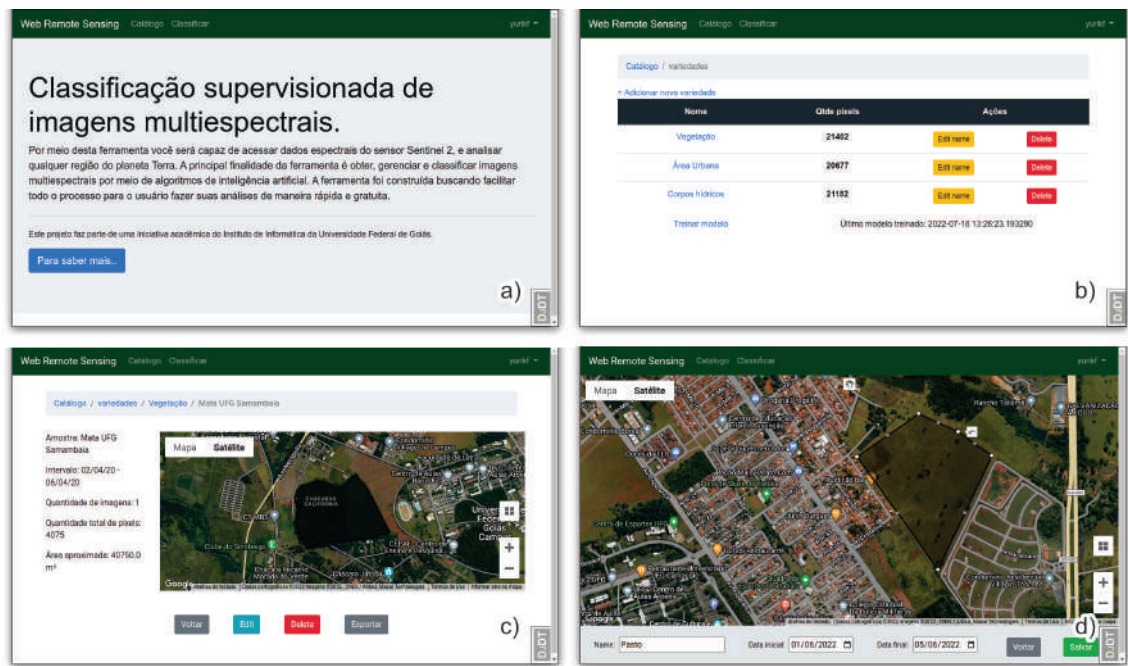


Figura 4.3: Ilustração contendo principais páginas da plataforma. Na sequência temos: a) a página inicial, b) menu de catálogos de variedades, c) detalhamento de região amostral, d) adição de região amostral.

4.1.3 Detalhamento do algoritmo

Para o melhor entendimento do funcionamento da plataforma, é apresentada a Figura 4.4, a qual mostra quais são os passos para o usuário realizar uma classificação supervisionada dentro do sistema, e quais são as rotinas realizadas pelo algoritmo durante esse fluxo.

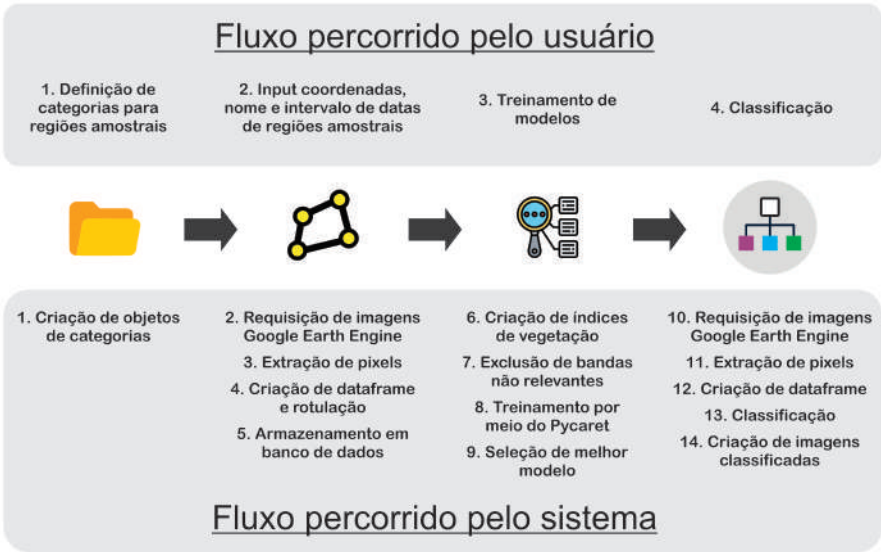


Figura 4.4: Fluxo de execuções do algoritmo

Detalhando um pouco mais sobre o fluxo do usuário, na etapa 1, inicialmente é necessário que o usuário defina grupos de estudo, como por exemplo: uma pasta para classificação da cobertura do solo, outra pasta para tipos de vegetação, e assim por diante. Dentro de cada um desses itens serão definidas as classes e amostras referentes àquele estudo criado anteriormente.

Durante a etapa 2, o usuário deve definir quais classes ele deseja usar no estudo. Para cada classe deverão ser inseridas regiões amostrais. Por exemplo, caso o estudo seja de classificação da cobertura do solo, pode-se inserir regiões amostrais de vegetação, áreas urbanas, corpos hídricos. As inserções são feitas por meio da definição de um polígono e o intervalo de datas definido pelo usuário. As etapas 3 e 4 de treinamento podem ser realizadas a qualquer momento pelo usuário, sem exigir informações de entrada além das amostras já cadastradas, ou região a ser classificada.

4.1.4 Coleta de dados

Diferentemente do processo de coleta de imagens convencional, aqui a coleta é feita por meio do serviço GEE. Enquanto, tradicionalmente, as imagens dos sensores são coletadas inteiras, com uma grande quantidade de *pixels*, a coleta por meio do serviço do *Google* é feita localmente. Ou seja, com a área de estudo definida, o serviço já extrai somente as informações da área solicitada, executando, internamente ao serviço, o recorte das imagens, evitando, dessa forma, a transmissão de grandes quantidades de dados que posteriormente seriam descartados por não estarem contidas na área de estudo.

4.2 Experimentos

Os testes a seguir foram realizados em um computador com sistema operacional *Linux*, equipado com 20 Ghz de memória ram, disco rígido com conexão Sata 3, processador *Amd Radeon Ryzen 5 3500u* de 8 núcleos. Durante o período de testes a taxa de *download* foi em média 266.64 Mbps e *upload* 207.37Mbps.

4.2.1 Criação dos dados de treinamento

Com o intuito de ilustrar o funcionamento da ferramenta foram classificadas regiões amostrais com variedades de cobertura de solo. Com esse objetivo, utilizando a própria ferramenta, foram coletados dados de três tipos de cobertura: vegetação, corpos hídricos e área urbana. Na Tabela 4.2 são apresentadas a quantidade de regiões amostrais e *pixels* referentes a cada classe.

Tabela 4.2: *Relação de amostras coletadas de coberturas de solo*

Tipo de cobertura	Quantidade de imagens	Quantidade de pixels
Vegetação	8	21402
Área urbana	8	20677
Corpos hídricos	21	21182

4.2.2 Preparação de dados

Após a coleta de imagens, é realizada uma preparação automática dos dados para o treinamento do modelo de classificação. Tal processo consiste em: converter as imagens em tabelas; limpar dados, excluindo os que não correspondem à região selecionada pelo usuário; vincular cada pixel à classe selecionada pelo usuário; integrar as informações de todas as imagens do estudo. Esse conjunto de dados fica armazenado no banco de dados obedecendo relações de chaves estrangeiras e rotulação dos *pixels* associados às regiões amostrais. Outra etapa importante é a manipulação dos valores dos *pixels* e produção de índices de vegetação. Para esse experimento foram incluídos os índices: NDVI, EVI, SAVI, GNDVI e NDWI.

4.2.3 Treinamento de modelos de classificação

A etapa de treinamento foi realizada usando a biblioteca do *Python*, chamada *Pycaret*, que consiste de um pacote desenvolvido para ser um AutoML, otimizando processos de: tratamento de dados, validações, escolha de parâmetros, treinamento de modelo, geração de métricas de avaliação, etc. Além disso, por meio dessa ferramenta é possível fazer treinamentos de vários tipos de modelos de classificação sequencialmente com a mesma base de dados, assim, pode-se ter uma visão mais ampla de quais modelos são propensos a prever melhor para o banco de dados em questão.

O modelo que apresentou melhores resultados foi o *Light Gradient Boosting Machine* (LightGBM), seguido do *Random Forest*. As métricas podem ser vistas na [tabela 4.3](#).

Tabela 4.3: *Métricas dos resultados de treinamentos*

Métrica	Light Gradient Boosting Machine	Random Forest Classifier
Acurácia	0.9985	0.9970
AUC	1.0000	1.0000
Recall	0.9981	0.9961
Precision	0.9985	0.9970
F1	0.9985	0.9970
Kappa	0.9976	0.9951

Resultados e discussões

5.1 Regiões amostrais classificadas

Todas regiões amostrais foram escolhidas no estado de Goiás. Buscou-se escolher polígonos de diferentes ordens de grandezas no quesito área analisada. A quantidade de *pixels* referente a cada área será apresentada a seguir.

As figuras 5.1, 5.2, 5.3, 5.4 e 5.5 apresentam:

- a) a área a ser classificada (a imagem se refere à base do *google maps* e não à imagem de satélite)
- b) A imagem do *google maps* como base e, por cima, a imagem provida pelo sensor, a qual será utilizada para extração de valores de *pixels*.
- c) a imagem colorida baseada na classificação.
- d) a área referente a cada classe.

A partir da visualização do mapa do serviço da *Google* juntamente com a imagem provida pelo sensor, pode-se perceber que o processo de rotulação, treinamento e classificação de imagens está funcional e em geral condizente com a classe real do *pixel*.

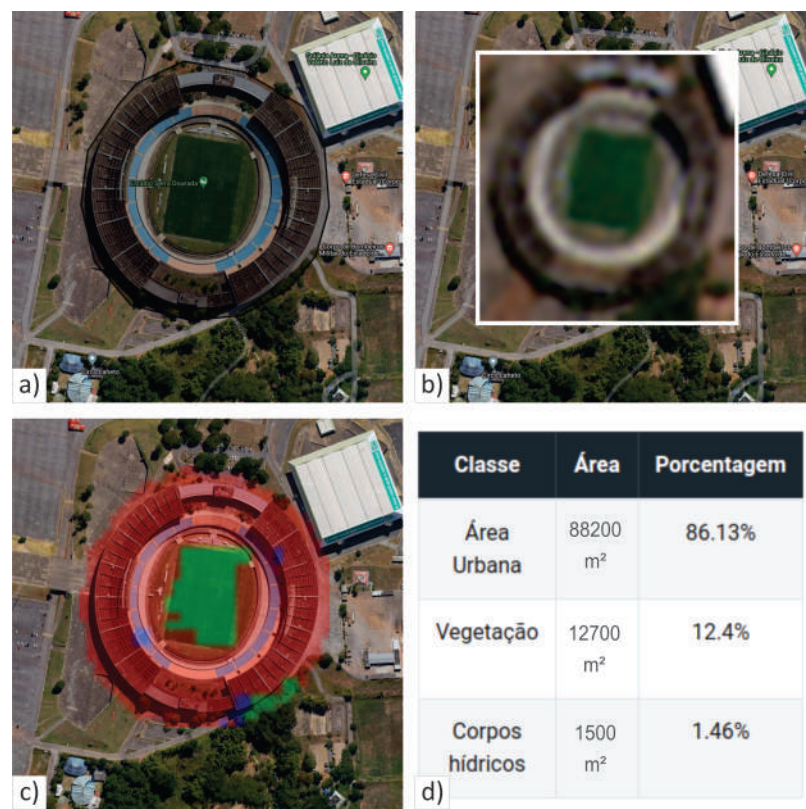


Figura 5.1: Região amostral do estádio Serra Dourada, Goiânia.

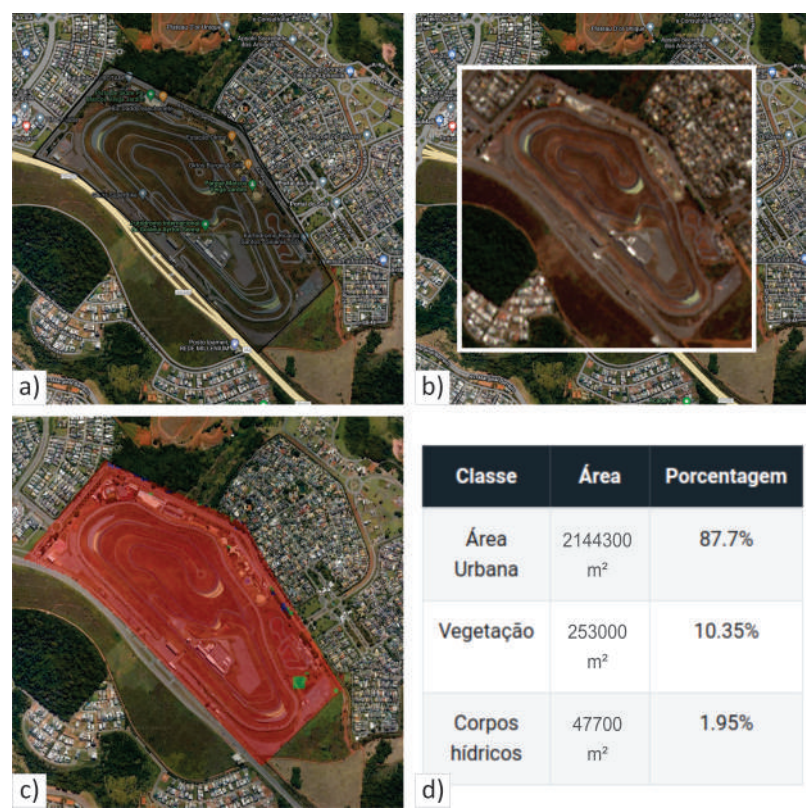


Figura 5.2: Região amostral do autódromo de Goiânia.

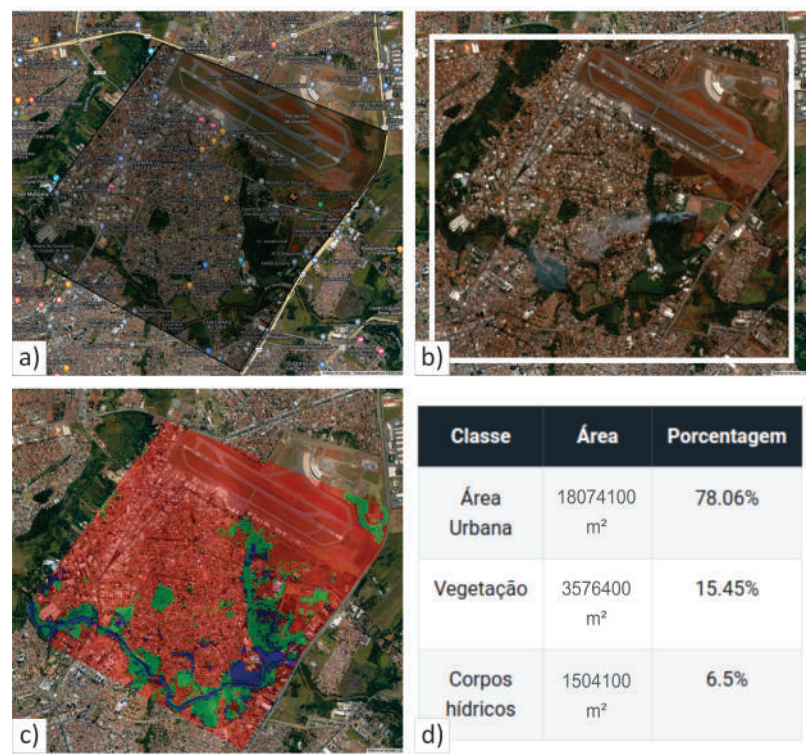


Figura 5.3: Região amostral na região do aeroporto e bairro de Santa Genoveva, Goiânia.



Figura 5.4: Região amostral da cidade de Goianira.

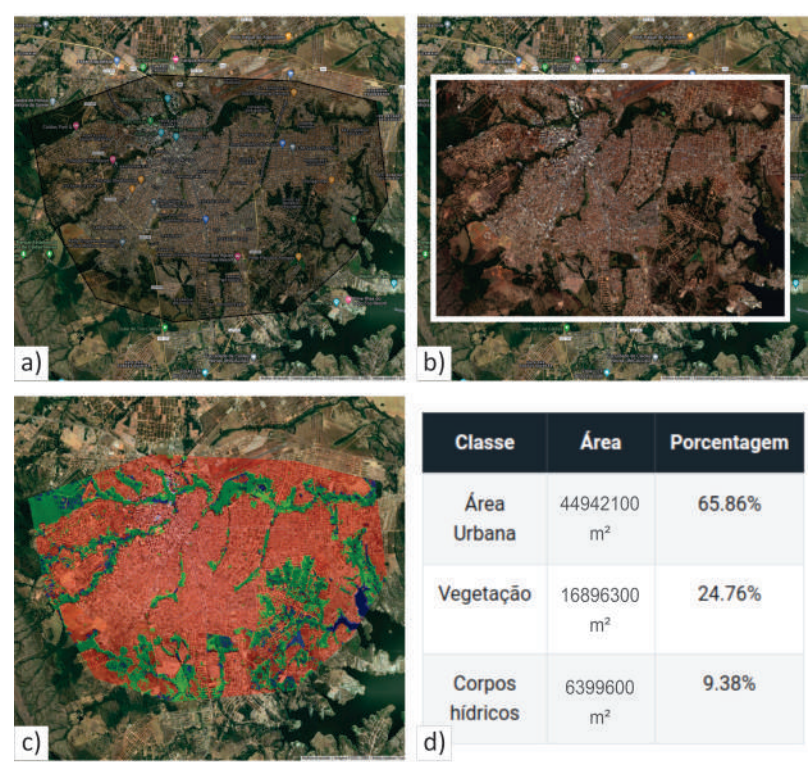


Figura 5.5: Região amostral da cidade de Caldas Novas.

5.2 Otimização do tempo

Para fins de comparação do uso da ferramenta com os softwares convencionais, foram registrados os tempos de coleta e manipulação de dados até a classificação. Na Tabela 5.2 é listado o tempo utilizado para coletar a quantidade referente de *pixels*, o tempo de rotulação, tempo de criação de imagem com cor natural e tempo agrupado classificação e criação de imagem classificada. Destaca-se que há um valor fixo mínimo de três a quatro segundos para fazer a conexão com a API do *Google Earth Engine*.

Tabela 5.1: Tabela de resultados com ênfase em tempo de processamento

Região Amostral	Quantidade de pixels	Download imagens (s)	Rotulação (s)	Imagem natural (s)	Imagem classificada (s)	Tempo total
Serra Dourada	802	2.5927	0.8535	2.4784	4.8937	10 seg
Autódromo	9897	8.1175	6.7054	2.5902	5.0505	22 seg
Região Aeroporto	129919	5.8553	61.1614	2.8217	16.065	1 min 25 seg
Goianira	236087	6.5483	91.9386	3.3887	11.2762	1 min 53 seg
Caldas Novas	538833	7.3767	192.6558	6.7896	25.6886	3 min 51 seg
Processo Convencional	10 ⁷	263.63	-	-	-	4 min 23 seg

Conforme a tabela, observamos que para áreas de até 5 km², equivalente a 5 milhões de *pixels* exemplificado pela região de Caldas Novas, a ferramenta construída leva pouco menos de 4 minutos para obtenção, rotulação, criação de imagem de cor natural, predição, e criação de imagem classificada. Esse mesmo tempo é dispendido apenas para obtenção da imagem do sensor por meio dos rotinas convencionais.

5.3 Otimização do armazenamento

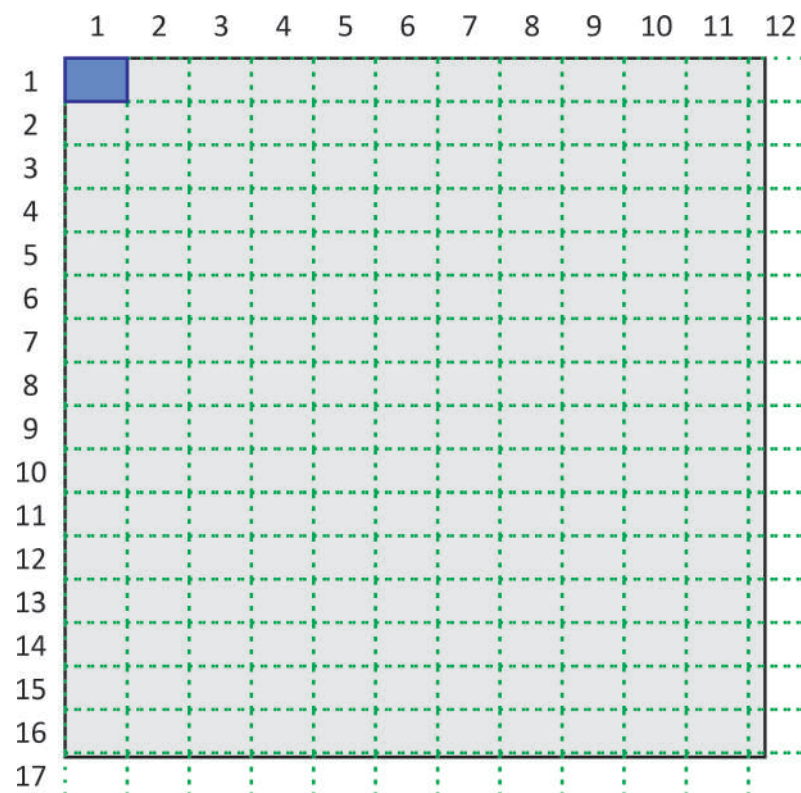
Outra variável importante para comparação é o tamanho dos arquivos gerados. Convencionalmente, após escolhida uma região de estudo, uma imagem inteira com 10 milhões de *pixels* é obtida usando sites como os citados anteriormente. Com base na geometria de área a ser estudada, essa imagem é recortada, e são extraídos os dados somente da região de estudo. Devido a grande quantidade de *pixels*, e as 12 bandas espectrais, esse processo se torna muito custoso. Principalmente, se o processo for repetido para várias imagens com datas diferentes.

Na tabela 5.3 podemos ver a necessidade de armazenamento necessária para cada uma das classificações mostradas acima. Pode-se observar que à medida que a quantidade de pixels aumenta, os valores tendem ao processo convencional. Porém, dificilmente uma imagem completa com 10 milhões de *pixels* é utilizada, fazendo-se necessário recortes de imagem.

Tabela 5.2: *Tabela de resultados com ênfase em armazenamento*

Região Amostral	Quantidade de pixels	Imagem original	Imagem natural	Imagem classificada	Total
Serra Dourada	802	24,9 kB	786,9 kB	3,0 kB	814,8 kB
Autódromo	9897	283,8 kB	1,3 MB	1,4 kB	1,58 MB
Região Aeroporto	129919	2,8 MB	2,3 MB	17,7 kB	5,11 MB
Goianira	236087	4,1 MB	2,3 MB	25,9kB	6,42 MB
Caldas Novas	538833	8,1 MB	2,2 MB	71,2 kB	10,62 MB
Processo Convencional	10^7	830,6 MB	-	-	830,6 MB

Para ficar ainda mais claro o ganho de otimização, abaixo é apresentada a Figura 5.6. Onde a área azul representa a área de Caldas Novas classificada acima, e a área cinza, representa uma imagem completa do satélite Sentinel-2 MSI. É nítida a diferença de grandeza entre as áreas. E uma vez que o estudo era localizado somente na região de Caldas Novas, não há necessidade de coletar informações de outras regiões. Pelo processo convencional, toda essa área cinza fora da área de estudo seria descartada para cada imagem analisada. Destaca-se novamente que essa quantidade de informações descartadas cresce muito mais quando são utilizadas múltiplas imagens.

**Figura 5.6:** *Comparativo da região classificada em Caldas Novas e uma imagem completa do Satélite Sentinel 2-MSI.*

5.4 Trabalhos correlatos

Durante a revisão sistemática também foram citados trabalhos que visam fazer a classificação da cobertura do solo. Dentre eles, podemos citar o trabalho de Ang et al. [6], o qual usou inteligência artificial, sensoriamento remoto e a plataforma do GEE para monitorar uma área de mineração por duas décadas, e analisar o impacto causado. Quanto a metodologia utilizada, todo o algoritmo é executado dentro da própria plataforma do GEE, inclusive o algoritmo de classificação *Random Forest*. Por um lado, utilizar uma ferramenta já pronta com algoritmos de classificação já inclusos traz agilidade quanto a execução de experimentos e geração de resultados, por outro, perde-se a praticidade do uso da ferramenta, e limita-se a aplicação do trabalho a apenas o objetivo específico dele.

Outro trabalho com aspectos muito semelhantes ao aqui apresentado é o de Xing et al. [99] o qual construiu a própria plataforma para classificação de cobertura do solo por meio de SR e IA com base na GEE. De acordo com o exposto no trabalho, o usuário pode definir a área de estudo, selecionar os dados de SR, rotular dados, selecionar algoritmos de classificação, avaliar resultados e apresentá-los por meio de uma interface. As principais diferenças observadas são: a) há uma limitação quanto as combinações de bandas disponíveis, b) a interface é composta por uma tela única e ainda está fortemente vinculada a plataforma original do GEE, e 3) o usuário fica responsável por escolher o o modelo de classificação. Isso nos leva a acreditar que os usuários dessa plataforma devem ter conhecimento técnico sobre SR e IA para que consigam usá-la corretamente.

Como citado anteriormente, este trabalho visa construir uma ferramenta que facilite o acesso de usuários leigos a classificações espectrais de cobertura do solo com o uso de SR e IA. Por isso foram construídas páginas de cadastro de estudos individualizados, cada um com suas amostras, e também há uma sessão somente para classificação e resultados, além disso todas as bandas disponíveis são coletadas, tudo isso com o fim de diversificar ao máximo as aplicações. As escolhas mais técnicas, como por exemplo, o algoritmo a ser utilizado para classificação, são feitas automaticamente por meio do AutoML, priorizando algoritmos com melhores precisões.

Essa abordagem exige o usuário de conhecer a fundo cada aspecto técnico da área de SR e IA, evitando que o usuário trave em alguma etapa por limitação de conhecimento e não conclua o estudo. Além disso, todo o processo de classificação se torna menos complexo e mais rápido e direto.

Nessa revisão não foram encontrados trabalhos que avaliam o processo e tempo de coleta de informações, evidenciando que é uma área que ainda pode ser bastante explorada. Os benefícios de se otimizar esse processos são claros e podem ser ainda considerados como gargalos, uma vez que esse processo ainda é pouco discutido.

Conclusão

Conforme levantado na revisão sistemática, há uma grande variedade de aplicações na área de Sensoriamento Remoto voltadas para a cobertura do solo e, em especial, para a agricultura. E há uma tendência de que com os avanços nas tecnologias, principalmente relacionada a serviços *web*, muitas novas aplicações se tornem ainda mais viáveis.

Grande parte dos estudos aqui apresentados propunham aplicações de monitoramento da superfície terrestre. Em geral, tais trabalhos lidam com áreas extensas. Devido a isso, optou-se por escolher imagens com tamanho de *pixels* maiores, abrangendo assim uma área maior utilizando menor quantidade de imagens. Mesmo assim, as análises temporais requerem imagens com datas distintas para uma mesma região amostral, aumentando assim a quantidade de imagens necessária a depender do estudo.

Como visto anteriormente, há ainda um número reduzido de aplicações que utilizam os serviços de processamento em nuvem disponíveis, principalmente quanto ao processo de coleta de dados. A depender da aplicação, são necessários grandes espaços de armazenamento para registrar tantas imagens e informações.

Por outro lado, para estudos com abrangência menor, por exemplo identificação de culturas, a quantidade de imagens em geral é reduzida. Devido a área de estudo ser menor, há uma transmissão excessiva de informações que provavelmente serão descartadas. Perde-se tempo durante a coleta, na manipulação das imagens e na extração de informações.

6.1 Contribuições

Por meio deste trabalho, pôde-se perceber que com uma arquitetura bem estruturada, porém simples, é possível coletar imagens multiespectrais, rotular amostras, treinar modelos e executar classificações em questão de alguns minutos ou até segundos, usando processos que integram serviços em nuvem gratuitos.

Os resultados das classificações evidenciaram a exatidão da classificação que pode ser executada em qualquer região do globo e em qualquer intervalo de data, respeitando os 5 dias de revisita do satélite a uma mesma região.

Com essa nova abordagem o custo com coletas e armazenamento de imagens é consideravelmente diminuído. De forma prática, pode-se chegar a custos próximos de zero para o armazenamento de informações, em apenas alguns segundos para a coleta, manipulação e classificação das informações. Além do mais, haveria um ganho de processamento, uma vez que o algoritmo não seria limitado as configurações locais quando da execução.

Outro ganho importante para o usuário é que por se tratar de uma aplicação online, o armazenamento das informações será feito no servidor que hospeda a aplicação, evitando, assim o uso excessivo do armazenamento por parte do usuário.

Devido à tecnicidade das áreas de Sensoriamento Remoto e Aprendizado de Máquina, muitos pesquisadores e profissionais de outras áreas não conseguem usufruir dessas tecnologias extremamente úteis quando o objetivo é diminuir custos - seja com transporte, equipe, materiais - e identificar, classificar e monitorar objetos na superfície da Terra à distancia.

Além disso, o ganho de tempo na preparação dos dados ficou evidente. Como afirmado anteriormente, essa etapa é uma das mais demoradas e trabalhosas. Para pessoas não especializadas isso se torna uma tarefa ainda mais complexa.

Por fim, por se tratar de uma ferramenta que integra o fluxo completo de uma classificação supervisionada, o usuário evita a troca de contextos para diversos ambientes com o fim de realizar as etapas de obtenção, preparação e rotulação de dados, treinamento de modelos, classificação e análise de resultados. A integração de todas etapas traz eficiência e retém o foco do usuário em sua tarefa de classificação, facilitando e agilizando o processo.

6.2 Trabalhos futuros

Há possibilidades de melhoria da pesquisa, a depender da aplicação. As classificações apresentadas neste trabalho, foram baseadas em apenas uma data. Porém, definindo um intervalo de datas maior, a ferramenta já está apta para coletar imagens da mesma área em diferentes datas, possibilitando a execução de análises temporais. Seria apenas necessário, a inclusão de modelos de reconhecimento de padrões aptos para tais análises.

Observando a Tabela 5.2, pode-se perceber também que a etapa de rotulação apresenta um crescimento alto relacionado a quantidade de *pixels*. Possivelmente um banco de dados mais robusto e performático reduziria o tempo com esse processo.

A inclusão de outras fontes de imagens, poderia melhorar a acurácia da classificação. Porém seriam necessárias adaptações durante todo o fluxo de trabalho, como, por

exemplo, processos de alinhamento de imagens, integração dos dados de imagens com tamanhos de *pixels* diferentes, possíveis correções de novas imagens, etc.

A depender da informação desejada, pode-se usar outro serviço de mapa que não seja do *Google Maps* para que o usuário identifique de maneira mais fácil a região a ser estudada, ou tenha outras imagens para comparação.

Por fim, essa pesquisa teve início a partir de uma demanda específica para classificação na área de agricultura, porém ao longo do desenvolvimento da ferramenta percebeu-se a grande variedade de aplicações que envolvam classificação da cobertura do solo. Destaca-se diminuição considerável em tempo, processamento, armazenamento e a reduzida quantidade de etapas para executar uma classificação supervisionada, abrindo uma grande possibilidade de inclusão de pesquisadores de outras grandes áreas.

Referências Bibliográficas

- [1] AGGARWAL, P.; SHIRSATH, P.; VYAS, S.; ARUMUGAM, P.; GOROSHI, S.; ARAVIND, S.; NAGPAL, M.; CHANANA, M. **Application note: Crop-loss assessment monitor—a multi-model and multi-stage decision support system.** *Computers and Electronics in Agriculture*, 175:105619, 2020.
- [2] AGRAWAL, S.; GUPTA, R. **Web gis and its architecture: a review.** *Arabian Journal of Geosciences*, 10(23):1–13, 2017.
- [3] AHMED, S.; KOTB, A.; SAMIR, E.; MOUSTAFA, M.; FARG, E.; ABD ELHAY, A.; ARAFAT, S. M. **A webgis decision support system for wadi el natrun rural land management.** In: *2020 21st International Arab Conference on Information Technology (ACIT)*, p. 1–7. IEEE, 2020.
- [4] ALI, J.; KHAN, R.; AHMAD, N.; MAQSOOD, I. **Random forests and decision trees.** *International Journal of Computer Science Issues (IJCSI)*, 9(5):272, 2012.
- [5] ALLRED, B. W.; BESTELMEYER, B. T.; BOYD, C. S.; BROWN, C.; DAVIES, K. W.; DUNIWAY, M. C.; ELLSWORTH, L. M.; ERICKSON, T. A.; FUHLENDORF, S. D.; GRIFITHS, T. V.; OTHERS. **Improving landsat predictions of rangeland fractional cover with multitask learning and uncertainty.** *Methods in Ecology and Evolution*, 12(5):841–849, 2021.
- [6] ANG, M. L. E.; ARTS, D.; CRAWFORD, D.; LABATOS JR, B. V.; NGO, K. D.; OWEN, J. R.; GIBBINS, C.; LECHNER, A. M. **Socio-environmental land cover time-series analysis of mining landscapes using google earth engine and web-based mapping.** *Remote Sensing Applications: Society and Environment*, 21:100458, 2021.
- [7] ARUNDEL, J.; WINTER, S.; GUI, G.; KEATLEY, M. **A web-based application for beekeepers to visualise patterns of growth in floral resources using modis data.** *Environmental Modelling & Software*, 83:116–125, 2016.
- [8] ASTSATRYAN, H.; HAYRAPETYAN, A.; NARSISIAN, W.; SARIBEKYAN, A.; ASMARYAN, S.; SAGHATELYAN, A.; MURADYAN, V.; GUIGOZ, Y.; GIULIANI, G.; RAY, N. **An inte-**

- roperable web portal for parallel geoprocessing of satellite image vegetation indices.** *Earth Science Informatics*, 8(2):453–460, 2015.
- [9] BAKIRMAN, T.; GÜMÜŞAY, M. Ü.; MUSAOĞLU, N.; TANIK, A. G. **Development of sustainable wetland management strategies by using the analytical hierarchy process and web-based gis: A case study from turkey.** *Transactions in GIS*, 26(3):1589–1608, 2022.
- [10] BOLSTAD, P. **GIS Fundamentals: A First Text on Geographic Information Systems.** 04 2012.
- [11] BORDOGNA, G.; KLIMENT, T.; FRIGERIO, L.; BRIVIO, P. A.; CREMA, A.; STROPPIANA, D.; BOSCHETTI, M.; STERLACCHINI, S. **A spatial data infrastructure integrating multisource heterogeneous geospatial data and time series: A study case in agriculture.** *ISPRS International Journal of Geo-Information*, 5(5):73, 2016.
- [12] BOSQUE GONZÁLEZ, I. D.; FERNÁNDEZ FREIRE, C.; MARTÍN-FORERO MORENTE, L.; PÉREZ ASENSIO, E. **Los sistemas de información geográfica y la investigación en ciencias humanas y sociales.** Confederación Española de Centros de Estudios Locales, 2012.
- [13] BREIMAN, L. **Random forests.** *Machine learning*, 45(1):5–32, 2001.
- [14] CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. V. **Introdução à ciência da geoinformação.** 2001.
- [15] CÂMARA, G.; MONTEIRO, A. M. V. **Conceitos básicos em ciência da geoinformação.** *Introdução a Ciência da Geoinformação.* São Jose dos Campos, 2001.
- [16] CHEN, J.; CAO, X.; PENG, S.; REN, H. **Analysis and applications of globe-land30: a review.** *ISPRS International Journal of Geo-Information*, 6(8):230, 2017.
- [17] CHEN, N.; ZHANG, X.; CHEN, Z.; YAN, S. **Integrated geospatial sensor web for agricultural soil moisture monitoring.** In: *2015 Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics)*, p. 28–32. IEEE, 2015.
- [18] CHENG, G.; XIE, X.; HAN, J.; GUO, L.; XIA, G.-S. **Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities.** *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3735–3756, 2020.
- [19] COWEN, D. J. **Gis versus cad versus dbms: What are the differences? photogrammetric engineering and remote sensing.** 1988.

- [20] DALL'AGNOL, R. W.; MICHELON, G. K.; BAZZI, C. L.; MAGALHÃES, P. S. G.; DE SOUZA, E. G.; BETZEK, N. M.; SOBJAK, R. **Web applications for spatial analyses and thematic map generation.** *Computers and Electronics in Agriculture*, 172:105374, 2020.
- [21] DAS, K.; PRAMANIK, D.; SANTRA, S.; SENGUPTA, S. **Parcel wise crop discrimination and web based information generation using remote sensing and open source software.** *The Egyptian Journal of Remote Sensing and Space Science*, 22(1):117–125, 2019.
- [22] DHEDIA, R.; PALIAKKARA, N.; LOBO, V. B.; GUPTA, D.; SHARMA, V. **Smart agri-farming on satellite imageries using machine learning.** In: *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, p. 1342–1347. IEEE, 2021.
- [23] DÍAZ-HORMAZÁBAL, I.; VALENCIA, D.; VALENZUELA, J. **Implementation of technologies in the public service: Geomatics in the cloud for monitoring wetlands in protected areas.** In: *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, p. 155–159. IEEE, 2020.
- [24] EBERLE, J.; TAYLOR, T.; SCHMULLIUS, C. **Easy to use time-series data access and analysis tools using standardbased geoprocessing services.** In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 3614–3617. IEEE, 2016.
- [25] ESQUERDO, J. C. D. M.; ANTUNES, J. F. G.; COUTINHO, A. C.; SPERANZA, E. A.; KONDO, A. A.; DOS SANTOS, J. L. **Satveg: A web-based tool for visualization of modis vegetation indices in south america.** *Computers and Electronics in Agriculture*, 175:105516, 2020.
- [26] FISCHER, W.; HEMPHILL, W.; KOVER, A. **Progress in remote sensing (1972–1976).** *Photogrammetria*, 32(2):33 – 72, 1976.
- [27] FRIEDMAN, J. H. **Greedy function approximation: a gradient boosting machine.** *Annals of statistics*, p. 1189–1232, 2001.
- [28] FU, Y.; LUAN, G.; CAI, J.; LI, Y.; ZHAO, F. **Construction and application research of ecological evaluation platform based on cesium.** *Water Supply*, 21(3):983–991, 2021.
- [29] FUSSELL, J.; RUNDQUIST, D.; HARRINGTON, J. **On defining remote sensing.** *Photogrammetric Engineering and Remote Sensing*, 52(9):1507–1511, 1986.

- [30] GARETH, J.; DANIELA, W.; TREVOR, H.; ROBERT, T. **An introduction to statistical learning: with applications in R**. Springer, 2013.
- [31] GHAMISI, P.; PLAZA, J.; CHEN, Y.; LI, J.; PLAZA, A. J. **Advanced spectral classifiers for hyperspectral images: A review**. *IEEE Geoscience and Remote Sensing Magazine*, 5(1):8–32, 2017.
- [32] GOMES, V. C.; QUEIROZ, G. R.; FERREIRA, K. R. **An overview of platforms for big earth observation data management and analysis**. *Remote Sensing*, 12(8):1253, 2020.
- [33] GRANELL, C.; MIRALLES, I.; RODRÍGUEZ-PUPO, L. E.; GONZÁLEZ-PÉREZ, A.; CASTELEYN, S.; Busetto, L.; PEPE, M.; BOSCHETTI, M.; HUERTA, J. **Conceptual architecture and service-oriented implementation of a regional geoportal for rice monitoring**. *ISPRS International Journal of Geo-Information*, 6(7):191, 2017.
- [34] GUO, J.; HUANG, C.; HOU, J. **A scalable computing resources system for remote sensing big data processing using geopyspark based on spark on k8s**. *Remote Sensing*, 14(3):521, 2022.
- [35] GUO, L.; MA, Y.; CUKIC, B.; SINGH, H. **Robust prediction of fault-proneness by random forests**. In: *15th international symposium on software reliability engineering*, p. 417–428. IEEE, 2004.
- [36] HAMUNYELA, E.; ROSCA, S.; MIRT, A.; ENGLE, E.; HEROLD, M.; GIESEKE, F.; VERBESSELT, J. **Implementation of bfastmonitor algorithm on google earth engine to support large-area and sub-annual change monitoring using earth observation data**. *Remote Sensing*, 12(18):2953, 2020.
- [37] HAN, G.; CHEN, J.; HE, C.; LI, S.; WU, H.; LIAO, A.; PENG, S. **A web-based system for supporting global land cover data production**. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:66–80, 2015.
- [38] HARDIE, A. **The development and present state of web-gis**. *Cartography*, 27(2):11–26, 1998.
- [39] HE, X.; ZHAO, K.; CHU, X. **Automl: A survey of the state-of-the-art**. *Knowledge-Based Systems*, 212:106622, 2021.
- [40] JACKSON, R. D. **Spectral indices in n-space**. *Remote sensing of environment*, 13(5):409–421, 1983.

- [41] JAYAKUMAR, K.; MALARVANNAN, S. **A webgis based decision support system for land use and land cover changes: a case study of tiruvallur block, tamil nadu.** *Int J Earth Sci Eng*, 8(4):1892–1898, 2015.
- [42] JAYAKUMAR, K.; SELVAM, V.; PRABAVATHY, V. **A webgis-based study for managing mangroves of godavari wetland, andhra pradesh, india.** *Journal of the Indian Society of Remote Sensing*, 47(12):2027–2045, 2019.
- [43] KALPOMA, K. A.; RAHMAN, A. **Web-based monitoring of boro rice production using improvised ndvi threshold of modis mod13q1 and myd13q1 images.** In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, p. 6877–6880, 2021.
- [44] KE, G.; MENG, Q.; FINLEY, T.; WANG, T.; CHEN, W.; MA, W.; YE, Q.; LIU, T.-Y. **Lightgbm: A highly efficient gradient boosting decision tree.** *Advances in neural information processing systems*, 30, 2017.
- [45] KHOSHGOFTAAR, T. M.; GOLAWALA, M.; VAN HULSE, J. **An empirical study of learning from imbalanced data using random forest.** In: *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, volume 2, p. 310–317. IEEE, 2007.
- [46] KITCHENHAM, B. **Procedures for performing systematic reviews.** *Keele, UK, Keele University*, 33(2004):1–26, 2004.
- [47] LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote sensing and image interpretation.** John Wiley & Sons, 2015.
- [48] LIU, J. G.; MASON, P. J. **Essential image processing and GIS for remote sensing.** John Wiley & Sons, 2013.
- [49] LIU, W. T. H. **Aplicações de sensoriamento remoto.** Oficina de Textos, 2015.
- [50] LONG, P. R.; BENZ, D.; MARTIN, A. C.; HOLLAND, P. W.; MACIAS-FAURIA, M.; SEDDON, A. W.; HAGEMANN, R.; FROST, T. K.; SIMPSON, A.; POWER, D. J.; OTHERS. **Left—a web-based tool for the remote measurement and estimation of ecological value across global landscapes.** *Methods in Ecology and Evolution*, 9(3):571–579, 2018.
- [51] MACHWITZ, M.; HASS, E.; JUNK, J.; UDELHOVEN, T.; SCHLERF, M. **Cropgis—a web application for the spatial and temporal visualization of past, present and future crop biomass development.** *Computers and Electronics in Agriculture*, 161:185–193, 2019.

- [52] MCCORD, P.; TONINI, F.; LIU, J. **The telecoupling geoapp: A web-gis application to systematically analyze telecouplings and sustainable development.** *Applied Geography*, 96:16–28, 2018.
- [53] MELTON, F. S.; HUNTINGTON, J.; GRIMM, R.; HERRING, J.; HALL, M.; ROLLISON, D.; ERICKSON, T.; ALLEN, R.; ANDERSON, M.; FISHER, J. B.; OTHERS. **Openet: Filling a critical data gap in water management for the western united states.** *JAWRA Journal of the American Water Resources Association*, 2021.
- [54] MENESES, P. R.; ALMEIDA, T. D. **Introdução ao processamento de imagens de sensoriamento remoto.** *Universidade de Brasília, Brasília*, 2012.
- [55] MILLAN, V. E. G.; FAUDE, U.; BICSAN, A.; KLINK, A.; TEUWSEN, S.; PAKZAD, K.; MÜTERTHIES, A. **Monitoring flooding damages in vegetation caused by mining activities using optical remote sensing.** *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86(1):1–13, 2018.
- [56] MILLER, I. J.; SCHIEBER, B.; DE BEY, Z.; BENNER, E.; ORTIZ, J. D.; GIRDNER, J.; PATEL, P.; CORADAZZI, D. G.; HENRIQUES, J.; FORSYTH, J. **Analyzing crop health in vineyards through a multispectral imaging and drone system.** In: *2020 Systems and Information Engineering Design Symposium (SIEDS)*, p. 1–5. IEEE, 2020.
- [57] MURO, J.; STRAUCH, A.; FITOKA, E.; TOMPOULIDOU, M.; THONFELD, F. **Mapping wetland dynamics with sar-based change detection in the cloud.** *IEEE Geoscience and Remote Sensing Letters*, 16(10):1536–1539, 2019.
- [58] NDUNGU, L.; OWARE, M.; OMONDI, S.; WAHOME, A.; MUGO, R.; ADAMS, E. **Application of modis ndvi for monitoring kenyan rangelands through a web based decision support tool.** *Frontiers in Environmental Science*, 7:187, 2019.
- [59] PASCOAL, L. M. L.; PARENTE, L. L.; NOGUEIRA, H. S.; JÚNIOR, L. G. F. **Deforestation polygon assessment tool: Providing comprehensive information on deforestation in the brazilian cerrado biome.** In: *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, p. 428–433. IEEE, 2020.
- [60] PENG, C.; DENG, M.; DI, L.; HAN, W. **Delivery of agricultural drought information via web services.** *Earth Science Informatics*, 8(3):527–538, 2015.
- [61] PENG, Z.-R. **Internet gis for public participation.** *Environment and Planning B: Planning and Design*, 28(6):889–905, 2001.

- [62] PERUMAL, K.; BHASKARAN, R. **Supervised classification performance of multispectral images**. *arXiv preprint arXiv:1002.4046*, 2010.
- [63] PICCARD, I.; GOBIN, A.; WELLENS, J.; TYCHON, B.; GOFFART, J.-P.; CURNEL, Y.; PLANCHON, V.; LECLEF, A.; COOLS, R.; CATTOOR, N. **Potato monitoring in belgium with “watchitgrow”**. In: *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, p. 1–4. IEEE, 2017.
- [64] PICKLES, J. **Representations in an electronic age: Geography, gis, and democracy**. Praxis (e) Press, 2008.
- [65] PRESS, G. **Cleaning big data: Most time-consuming, least enjoyable data science task, survey says**. *Forbes*, March, 23:15, 2016.
- [66] PUTZ, S. **Interactive information services using world-wide web hypertext**. *Computer Networks and ISDN Systems*, 27(2):273 – 280, 1994. Selected Papers of the First World-Wide Web Conference.
- [67] QUACH, C. H.; PHAM, M. T.; NGUYEN, T. S.; PHUNG, M. D. **Real-time agriculture field monitoring using iot-based sensors and unmanned aerial vehicles**. In: *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, p. 492–497. IEEE, 2021.
- [68] RAPIŃSKI, J.; BEDNARCZYK, M.; ZINKIEWICZ, D. **Jupyter ide as an online tool for earth observation data processing**. *Remote Sensing*, 11(17):1973, 2019.
- [69] REMBOLD, F.; MERONI, M.; URBANO, F.; CSAK, G.; KERDILES, H.; PEREZ-HOYOS, A.; LEMOINE, G.; LEO, O.; NEGRE, T. **Asap: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis**. *Agricultural systems*, 168:247–257, 2019.
- [70] REMBOLD, F.; MERONI, M.; URBANO, F.; LEMOINE, G.; KERDILES, H.; PEREZ-HOYOS, A.; CSAK, G. **Asap-anomaly hot spots of agricultural production, a new early warning decision support system developed by the joint research centre**. In: *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, p. 1–5. IEEE, 2017.
- [71] RIGAUX, P.; SCHOLL, M.; VOISARD, A. **Spatial databases: with application to GIS**. Elsevier, 2001.
- [72] RIZVI, S. R.; KILLOUGH, B.; CHERRY, A.; GOWDA, S. **The ceos data cube portal: A user-friendly, open source software solution for the distribution, exploration, analysis, and visualization of analysis ready data**. In: *IGARSS 2018-2018 IEEE*

- International Geoscience and Remote Sensing Symposium*, p. 8639–8642. IEEE, 2018.
- [73] ROH, Y.; HEO, G.; WHANG, S. E. **A survey on data collection for machine learning: a big data-ai integration perspective.** *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1328–1347, 2019.
- [74] RUIZ GUZMAN, H. A.; GONZÁLEZ-NAVARRO, F. F.; SELVARAJ-GOMEZ, M.; VALENCIA, M.; DELGADO, A. **Field phenomics: A web based image analysis platform using open source tools.** In: *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, p. 851–852, 2015.
- [75] SATYANARAYANAN, M. **The emergence of edge computing.** *Computer*, 50(1):30–39, 2017.
- [76] SAZIB, N.; MLADENOVA, I.; BOLTEN, J. **Leveraging the google earth engine for drought assessment using global soil moisture data.** *Remote sensing*, 10(8):1265, 2018.
- [77] SCHOWENGERDT, R. A. **Remote sensing: models and methods for image processing.** Elsevier, 2006.
- [78] SELVARAJ, M. G.; VALDERRAMA, M.; GUZMAN, D.; VALENCIA, M.; RUIZ, H.; ACHARJEE, A. **Machine learning for high-throughput field phenotyping and image processing provides insight into the association of above and below-ground traits in cassava (*manihot esculenta crantz*).** *Plant methods*, 16(1):1–19, 2020.
- [79] SHAH, P.; HIREMATH, D.; CHAUDHARY, S. **Towards development of spark based agricultural information system including geo-spatial data.** In: *2017 IEEE International Conference on Big Data (Big Data)*, p. 3476–3481. IEEE, 2017.
- [80] SHARMA, V. K.; BANU, V.; CHANDRASEKAR, K.; BHATTACHARYA, B. K.; SAI, M. S.; BHANUMURTHY, V. **Web processing service integrated with mobile application to identify suitable grain storage facility location.** *Spatial Information Research*, 25(1):131–140, 2017.
- [81] SHI, X. **High performance computing: fundamental research challenges in service oriented gis.** In: *Proceedings of the ACM SIGSPATIAL international workshop on high performance and distributed geographic information systems*, p. 31–34, 2010.

- [82] SHUKLA, S.; LANDSFELD, M.; ANTHONY, M.; BUDDE, M.; HUSAK, G. J.; ROWLAND, J.; FUNK, C. **Enhancing the application of earth observations for improved environmental decision-making using the early warning explorer (ewx).** *Frontiers in Climate*, 2:583509, 2021.
- [83] SINGH, S.; SINGH, H.; SHARMA, V.; SHRIVASTAVA, V.; KUMAR, P.; KANGA, S.; SAHU, N.; MERAJ, G.; FAROOQ, M.; SINGH, S. K. **Impact of forest fires on air quality in wolgan valley, new south wales, australia—a mapping and monitoring study using google earth engine.** *Forests*, 13(1):4, 2021.
- [84] SLINGSBY, J. A.; MONCRIEFF, G. R.; WILSON, A. M. **Near-real time forecasting and change detection for an open ecosystem with complex natural dynamics.** *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:15–25, 2020.
- [85] STEVENS, J. T.; COLLINS, B. M.; MILLER, J. D.; NORTH, M. P.; STEPHENS, S. L. **Changing spatial patterns of stand-replacing fire in california conifer forests.** *Forest Ecology and Management*, 406:28–36, 2017.
- [86] SUDMANN, M.; AUGUSTIN, H.; VAN DER MEER, L.; BARALDI, A.; TIEDE, D. **The austrian semantic eo data cube infrastructure.** *Remote Sensing*, 13(23):4807, 2021.
- [87] SZANTOI, Z.; BRINK, A.; BUCHANAN, G.; BASTIN, L.; LUPI, A.; SIMONETTI, D.; MAYAUX, P.; PEEDELL, S.; DAVY, J. **A simple remote sensing based information system for monitoring sites of conservation importance.** *Remote Sensing in Ecology and Conservation*, 2(1):16–24, 2016.
- [88] TAYYEBI, A.; TAYYEBI, A.; VAZ, E.; ARSANJANI, J. J.; HELBICH, M. **Analyzing crop change scenario with the smartscape™ spatial decision support system.** *Land use policy*, 51:41–53, 2016.
- [89] VAN HOEK, M.; ZHOU, J.; JIA, L.; LU, J.; ZHENG, C.; HU, G.; MENENTI, M. **A prototype web-based analysis platform for drought monitoring and early warning.** *International Journal of Digital Earth*, 2019.
- [90] VAN HOEK, M.; ZHOU, J.; JIA, L.; LU, J.; ZHENG, C.; HU, G.; MENENTI, M. **A prototype web-based analysis platform for drought monitoring and early warning.** *International Journal of Digital Earth*, 13(7):817–831, 2020.
- [91] VAN REES, E. **Open geospatial consortium (ogc).** *GeoInformatics*, 16(8):28, 2013.

- [92] VENTER, Z. S.; SCOTT, S. L.; DESMET, P. G.; HOFFMAN, M. T. **Application of landsat-derived vegetation trends over south africa: Potential for monitoring land degradation and restoration.** *Ecological Indicators*, 113:106206, 2020.
- [93] VOLPI, I.; MARCHI, S.; PETACCHI, R.; HOXHA, K.; GUIDOTTI, D. **Detecting olive grove abandonment with sentinel-2 and machine learning: The development of a web-based tool for land management.** *Smart Agricultural Technology*, 3:100068, 2023.
- [94] VUOLO, F.; D'URSO, G.; DE MICHELE, C.; BIANCHI, B.; CUTTING, M. **Satellite-based irrigation advisory services: A common tool for different experiences from europe to australia.** *Agricultural water management*, 147:82–95, 2015.
- [95] WACHOWIAK, M. P.; WALTERS, D. F.; KOVACS, J. M.; WACHOWIAK-SMOLÍKOVÁ, R.; JAMES, A. L. **Visual analytics and remote sensing imagery to support community-based research for precision agriculture in emerging areas.** *Computers and Electronics in Agriculture*, 143:149–164, 2017.
- [96] WANG, H.; HAN, D.; MU, Y.; JIANG, L.; YAO, X.; BAI, Y.; LU, Q.; WANG, F. **Landscape-level vegetation classification and fractional woody and herbaceous vegetation cover estimation over the dryland ecosystems by unmanned aerial vehicle platform.** *Agricultural and Forest Meteorology*, 278:107665, 2019.
- [97] XING, H.; CHEN, J.; WU, H.; HOU, D. **A web service-oriented geoprocessing system for supporting intelligent land cover change detection.** *ISPRS International Journal of Geo-Information*, 8(1):50, 2019.
- [98] XING, H.; CHEN, J.; WU, H.; ZHANG, J.; LI, S.; LIU, B. **A service relation model for web-based land cover change detection.** *ISPRS Journal of Photogrammetry and Remote Sensing*, 132:20–32, 2017.
- [99] XING, H.; HOU, D.; WANG, S.; YU, M.; MENG, F. **O-lcmapping: a google earth engine-based web toolkit for supporting online land cover classification.** *Earth Science Informatics*, 14(1):529–541, 2021.
- [100] XUE, J.; SU, B. **Significant remote sensing vegetation indices: A review of developments and applications.** *Journal of sensors*, 2017, 2017.
- [101] YALEW, S.; VAN GRIENSVEN, A.; VAN DER ZAAG, P. **Agrisuit: A web-based gis-mcda framework for agricultural land suitability assessment.** *Computers and Electronics in Agriculture*, 128:1–8, 2016.

- [102] ZHANG, L.; ZHANG, L.; DU, B. **Deep learning for remote sensing data: A technical tutorial on the state of the art.** *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, 2016.
- [103] ZHANG, T.; HUA, G.; LIGMANN-ZIELINSKA, A. **Visually-driven parallel solving of multi-objective land-use allocation problems: a case study in chelan, washington.** *Earth Science Informatics*, 8(4):809–825, 2015.
- [104] ZHANG, W.; CHEN, J.; LIAO, A.; HAN, G.; CHEN, X.; CHEN, L.; PENG, S.; WU, H.; ZHANG, J. **Geospatial knowledge-based verification and improvement of globeland30.** *Science China Earth Sciences*, 59(9):1709–1719, 2016.
- [105] ZHAO, H.; DI, L.; SUN, Z. **Watersmart-gis: A web application of a data assimilation model to support irrigation research and decision making.** *ISPRS International Journal of Geo-Information*, 11(5):271, 2022.
- [106] ZÖLLER, M.-A.; HUBER, M. F. **Benchmark and survey of automated machine learning frameworks.** *Journal of artificial intelligence research*, 70:409–472, 2021.