UNIVERSIDADE DO ESTADO DE MATO GROSSO
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E AGROECOSSISTEMAS AMAZÔNICOS

RAUL PIO DE AZEVEDO

IMAGENS REMOTAMENTE SENSORIADAS E MACHINE LEARNING NO MAPEAMENTO DA CULTURA DO GERGELIM

Dissertação de Mestrado

Alta Floresta - Mato Grosso - Brasil
2023

UNIVERSIDADE DO ESTADO DE MATO GROSSO
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E AGROECOSSISTEMAS AMAZÔNICOS

RAUL PIO DE AZEVEDO

IMAGENS REMOTAMENTE SENSORIADAS E MACHINE LEARNING NO MAPEAMENTO DA CULTURA DO GERGELIM

Dissertação apresentada à Universidade do Estado de Mato, como parte das exigências do Programa de Pós-graduação em Biodiversidade e Agroecossistemas Amazônicos, para obtenção do título de Mestre em Biodiversidade e Agroecossistemas Amazônicos.

Orientador: Prof. Dr. Carlos Antônio da Silva Junior.

Alta Floresta – Mato Grosso - Brasil
2023

AUTORIZO A DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO, CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE

Catalogação na publicação

Faculdade de Ciências Biológicas e Agrárias

AZEVEDO, Raul Pio de.

A994i Imagens Remotamente Sensoriadas e Machine Learning no Mapeamento da Cultura do Gergelim / Raul Pio de Azevedo – Alta Floresta, 2023.

79 f.; 30 cm. (ilustrações) Il. color. (sim)

Trabalho de Conclusão de Curso (Dissertação/Mestrado) – Curso de Pós-graduação Stricto Sensu (Mestrado Acadêmico) Biodiversidade e Agroecossistemas Amazônicos, Faculdade de Ciências Biológicas e Agrárias, Câmpus de Alta Floresta, Universidade do Estado de Mato Grosso, 2023.

Orientador: Carlos Antonio da Silva Junior

1. Gergelim. 2. Machine Learnig. 3. Mapeamento. 4. Classificação. 5. Random Forest. I. Raul Pio de Azevedo. II. Imagens Remotamente Sensoriadas e Machine Learning no Mapeamento da Cultura do Gergelim: .

CDU 633.85

IMAGENS REMOTAMENTE SENSORIADAS E MACHINE LEARNING NO MAPEAMENTO DA CULTURA DO GERGELIM

Raul Pio de Azevedo

Dissertação apresentada à Universidade do Estado de Mato Grosso, como parte das exigências do Programa de Pós-Graduação em Biodiversidade e Agroecossistemas Amazônicos, para a obtenção do título de Mestre em Biodiversidade e Agroecossistemas Amazônicos

Aprovada em: ___/___/_____

_____

Prof. Dr. Carlos Antônio da Silva Junior

Orientador – UNEMAT/ PPGBioAgro

_____

Prof. Dr. Rivanildo Dallacort

UNEMAT

_____

Prof. Dr. Cácio Luiz Boechat

Universidade Federal do Piauí-UFPI/ CPCE

DEDICATÓRIA

Dedico a minha querida avó Antônia Teodora (in *memoriam*) que deixou um legado de valores de amor, vivendo seus dias nessa Terra com expressões diárias de gratidão a Deus pelo dom da vida, desapega do materialismo, sempre sorrindo pela satisfação de criar os netos, passando pela velhice na companhia da família. Te amo!

AGRADECIMENTO

A Deus, pelo privilégio do dom da vida e no sustento diário.

Aos meus pais Rubens e Luzia, pelos importantes valores e ensinamentos transmitidos, muitas vezes por palavras, mas principalmente por atitudes, além de sempre exigirem dedicação nos estudos, bem como em tudo na vida.

A minha esposa Danubia, pela compreensão nos momentos desafiadores durante o mestrado, abdicando de vários momentos de lazer e comunhão.

Aos colegas Ismael Maciel e Rogério Souza, companheiros tanto no mestrado quanto no trabalho, pela troca de experiencias e conhecimento.

Aos meus chefes Gabriel Heise, Marly Silva e Francielle Tonietti, por viabilizarem a conciliação dos estudos com o trabalho, tenho certeza de que fizeram tudo que estava ao alcance.

Ao orientador Carlos Antônio da Silva Junior, por todo o trabalho de mapeamento e divulgação dos dados da cultura da soja no estado de Mato Grosso, pois foi o meio que tivemos contato inicialmente e, principalmente, pela orientação, dedicação de tempo e conhecimento.

Aos colegas do laboratório GAAF, José Wagner, João Della Silva, Tatiane Deoti e Fernado Saragosa, que em diversos momentos deram o suporte necessário.

A Universidade do Estado do Mato Grosso-UNEMAT e o Programa de Pós-graduação em Biodiversidade e Agroecossistemas Amazônicos-PPGBIOAgro, pela oportunidade de viver a rica experiência de aprender com excelentes professores que se dedicaram ao máximo em ensinar com qualidade mesmo em período de pandemia do Covid-19.

E a todos que de alguma forma contribuíram para este trabalho.

Obrigado!

SUMÁRIO

LISTA DE FIGURAS 15
LISTA DE TABELAS 16
RESUMO 17
ABSTRACT 18
1. INTRODUÇÃO 13
2. OBJETIVOS 16
2.1 Objetivo Geral 16
2.2 Objetivos Específicos 16
3. JUSTIFICATIVA 17
4. REVISÃO DE LITERATURA 18
4.1 A cultura do gergelim no Brasil 18
4.2 Fases fenológicas da cultura do gergelim 19
5. SENSORIAMENTO REMOTO AGRÍCOLA 20
5.1 Mapeamento de alvos agrícolas 20
5.1.1 Comportamento espectral da vegetação de ciclo anual 20
6. SENSORES ORBITAIS 24
6.1 Panorama geral dos sensores de órbita heliossíncrona 24
6.1.1 Landsat-8 (OLI) 26
7. PROCESSAMENTO DIGITAL DE IMAGENS 28
7.1 Correção atmosférica de imagens 28
8. CLASSIFICADORES DIGITAIS NA DETECÇÃO DE ALVOS AGRÍCOLAS 30
8.1 Análise via Machine Learning 31
8.1.1 Support Vector Machines 32
8.1.2 Random Forest 35
9. MATERIAL E MÉTODOS 37
9.1 Área de estudo 37
9.2 Aquisição das Imagens 39
9.3 Pré-processamento 40
9.4 Índices de Vegetação 42
9.5 Dados de Referência 44
9.6 Treinamento e Teste dos Algoritmos 46

9.7 Avaliação e Validação da Acurácia 47
10. RESULTADOS E DISCUSSÃO 51
11. CONCLUSÃO 61
12. REFERÊNCIAS BIBLIOGRÁFICAS 62
APÊNDICES 85

LISTA DE FIGURAS

Figura 1. Espectro eletromagnético.....23
Figura 2. Ilustração do trajeto típico de satélites em órbita polar e geoestacionário. 25
Figura 3. Capacidade do SVM de classificar conjuntos de dados que não são linearmente separáveis. 33
Figura 4. Arquitetura do modelo de Random Forest.....35
Figura 5. Localização geográfica/ distribuição dos biomas e tipologia dos solos da área de estudo.....38
Figura 6. Fluxograma para mapeamento e estimativa da cultura do gergelim. 39
Figura 7. Distribuição dos pontos de controle coletados na área de estudo. 45
Figura 8. Resultado da classificação para o ano-safra 2021/2022 no município de Canarana-MT.....51
Figura 9. Resultado da classificação para os classificadores para o ano-safra 2021/2022 no município de Canarana-MT. 52
Figura 10. Índices de vegetação para área de estudo. a) NDVI; b) NDBI; c) EVI; d) SAVI; e) PVI. 56
Figura 11. Cultura do gergelim com dosséis mal formados (a, b). 57
Figura 12. Classificação dos algoritmos para a classe gergelim: a) Random Forest; b) SVM; c) Composição falsa cor RGB. 60

LISTA DE TABELAS

Tabela 1. Fenologia da cultura do gergelim. 20
Tabela 2.Parâmetros de entrada do modelo para correção atmosférica da cena órbita/ponto 224-069. 41
Tabela 3. Parâmetros de entrada no modelo para correção atmosférica da cena órbita/ponto 225-069. 41
Tabela 4. Características do sensor Operational Land Imager satélite Landsat 8. 42
Tabela 5. Distribuição do conjunto amostral de pontos de controle, de acordo com as classes de uso e ocupação do solo, nas fases de treinamento/teste* e validação** da classificação. 46
Tabela 6. Definição das classes de uso e ocupação do solo e suas descrições. 48
Tabela 7. Matrizes de confusão (número de pixels) das tabulações cruzadas, corretas e incorretas do conjunto de validação, para os classificadores. 55
Tabela 8. Resultado da acurácia da classificação. 55
Tabela 9. Precisão do produtor, do usuário e índice Kappa. 57
Tabela 10. Teste de hipótese comparando os resultados de exatidão entre os índices - ano safra 2021/2022. 59
Tabela 11. Áreas estimadas para as classes de uso do solo no município de Canarana-MT. 60

RESUMO

AZEVEDO, Raul Pio. M.Sc. Universidade do Estado de Mato Grosso, fevereiro de 2023. **Imagens remotamente sensoriadas e machine learning no mapeamento da cultura do gergelim.** Orientador: Carlos Antônio da Silva Junior.

As mudanças nas paisagens têm sido acompanhadas mais intensamente nas últimas décadas graças aos avanços científicos, tanto no campo do melhoramento tecnológico dos satélites como nas técnicas de sensoriamento remoto. Técnicas avançadas e eficientes de machine learning têm ajudado os profissionais de sensoriamento remoto a determinar estas mudanças, desde às paisagens mais simples até as mais complexas, permitindo a identificação dos mais variados usos e ocupação do solo, bem como a estimativa das áreas que estes usos ocupam, permitindo uma gestão mais dinâmica dos recursos naturais, especialmente na exploração agrícola, fornecendo informação confiável aos tomadores de decisões. Assim, o objetivo deste trabalho é, através de técnicas de machine learning, estimar a área de cultivo do gergelim (*Sesamum indicum*) no ano agrícola 2021/2022, no município de Canarana, no estado de Mato Grosso, comparando o desempenho dos classificadores Random Forest e Support Vector Machine (SVM), utilizando imagens do satélite Landsat 8/OLI. Como fonte de informação para a classificação supervisionada, foram coletados pontos de controlo em coordenadas geográficas na área de estudo, para a identificação das áreas cultivadas com gergelim. Para a elaboração de mapas temáticos, foram utilizados os índices de vegetação NDVI, EVI, NDBI, PVI e SAVI, juntamente com as imagens do Landsat 8/OLI. Na avaliação dos mapas temáticos foram utilizados como regra a Exatidão Global e o índice Kappa, comparados pelo teste Z, com significância a α=0,05. O teste revelou que o classificador Random Forest apresentou melhor desempenho na identificação das áreas cultivadas de gergelim, com Exatidão Global de 0,95 e Kappa de 0,90, quando comparado ao SVM, que apresentou 0,91 e 0,81, respectivamente. A utilização de técnicas de machine learning em imagens Landsat 8/OLI revelou-se satisfatória na estimativa de áreas cultivadas com gergelim no município de Canarana-MT, demonstrando confiança no mapeamento. A maneira como o Random Forest estruturou seu modelo de treinamento, criando quantas árvores de decisão forem necessárias, acabou mitigando mais os erros de classificação e demonstrou-se mais promissora quando comparada ao SVM. Em regra, ambos os algoritmos apresentaram potencial para o mapeamento da cultura de gergelim.

Palavras-chave: Gergelim, Random Forest, SVM, classificação, monitoramento, machine learning.

AZEVEDO, Raul Pio. M.Sc. Universidade do Estado de Mato Grosso, fevereiro de 2023. **Imagens remotamente sensoriadas e machine learning no mapeamento da cultura do gergelim.** Orientador: Carlos Antônio da Silva Junior.

ABSTRACT

The changes in landscapes have been followed more intensely in recent decades thanks to scientific advances, both in the field of technological improvement of satellites and in remote sensing techniques. Advanced and efficient machine learning techniques have helped remote sensing professionals to determine these changes, from the simplest to the most complex landscapes, allowing the identification of the most varied land uses and occupation, as well as the estimation of the areas that these uses occupy, allowing a more dynamic management of natural resources, especially in agricultural exploitation, providing reliable information to decision makers. Thus, the objective of this work is, through machine learning techniques, to estimate the area of sesame (Sesamum indicum) cultivation in the agricultural year 2021/2022, in the municipality of Canarana, in the state of Mato Grosso, comparing the performance of the Random Forest and Support Vector Machine (SVM) classifiers, using images from the Landsat 8/OLI satellite. As a source of information for the supervised classification, control points in geographic coordinates were collected in the study area to identify the areas cultivated with sesame. The vegetation indices NDVI, EVI, NDBI, PVI and SAVI were used for the elaboration of thematic maps, along with the Landsat 8/OLI images. Global Accuracy and Kappa index were used as a rule of thumb in the evaluation of the thematic maps, compared by the Z test, with significance at $\alpha=0.05$. The test revealed that the Random Forest classifier showed better performance in identifying the sesame cultivated areas, with Global Accuracy of 0.95 and Kappa of 0.90, when compared to SVM, which showed 0.91 and 0.81, respectively. The use of machine learning techniques in Landsat 8/OLI images proved satisfactory in estimating areas cultivated with sesame in the municipality of Canarana-MT, demonstrating confidence in the mapping. The way Random Forest structures its training model, creating as many decision trees as necessary, ended up mitigating more classification errors and proved to be more promising when compared to SVM. As a rule, both algorithms showed potential for mapping the sesame crop.

Keywords: Sesame, Random Forest, SVM, classification, monitoring, machine learning.

1. INTRODUÇÃO

A produção de alimentos é essencial para suprir as demandas nutricionais diárias do ser humano. A maximização do uso da terra para o cultivo de diversas culturas em um mesmo ano/safra é crucial para o aumento da produção de alimentos, haja visto a contínua demanda global que, até o ano de 2050, a população poderá ser de 9,8 bilhões, 29% superior ao número atual e o crescimento maior será nos países em desenvolvimento, em torno de 70% da população será urbana e os níveis de renda serão maiores do que os atuais (FAO, 2017).

Ainda de acordo a FAO (2017), a produção de cereais terá que aumentar para 3 bilhões de toneladas/ano, em relação aos 2,5 bilhões produzidos atualmente, e a produção de proteína animal precisará aumentar mais de 200 milhões de toneladas.

O gergelim (*Sesamum indicum*) desempenha um papel importante na nutrição humana. Suas sementes são utilizadas essencialmente para a produção de óleo, mas também na produção da pasta (tehineh) e em formulações alimentares como o halaweh (tehineh adoçado), feijão java e saladas (ELLEUCH, et al., 2007).

Cultivado amplamente de regiões húmidas a moderadas do mundo, as sementes de gergelim contêm de 25 a 35% de proteínas, bem como 55% de óleo rico em gordura insaturada. Os ácidos graxos mais ricos em óleo de gergelim são ácidos linoleicos (37 a 42%) e ácido oleico (41 a 45%). Assim, o óleo de gergelim se torna adequado ao consumo humano por possuir elevada quantidade de ácidos graxos insaturados (ASGHAR; et al., 2014).

Os consumidores na atualidade preferem uma mercadoria com alto valor nutricional e baixo preço de mercado, e as sementes de gergelim atendem perfeitamente a esses requisitos, e mudar os padrões de consumo do consumidor e aumentar a conscientização sobre a saúde está impulsionando o mercado (INDUSTRY RESEARCH, 2019).

O potencial da cultura do gergelim tem atraído a atenção dos setores produtivo, pesquisa e melhoramento genético e industrial. No ano de 2020, as exportações de gergelim produzido no Brasil nos anos de 2018 e 2019 foram de $ 3.7 milhões e $ 25,4 milhões de dólares, respectivamente, um incremento de

586,5%. No ano de 2019, as exportações tiveram como destino o Vietnã (25%), Guatemala (15%), Turquia (12%) e Arábia Saudita (12%) (MINISTÉRIO DA AGRICULTURA, 2020).

Neste contexto, a cultura do gergelim foi introduzida no Brasil no século XVI pelos portugueses, tradicionalmente cultivada no nordeste brasileiro para consumo local (EMBRAPA, 2007a). No estado de Mato Grosso, o gergelim é considerado uma cultura que apresenta expansão exponencial devido aos seus baixos custos de produção, sua rentabilidade e excelente adaptação ao clima de Mato Grosso, na segunda safra, apresentando resistência à estiagem hídrica (CONAB, 2020a).

Ainda conforme Conab (2020a), em seu início, a cultura se restringiu apenas ao Vale do Araguaia, mais precisamente ao município de Canarana e, devido ao excelente desempenho e adaptação, a cultura se expandiu para outras regiões, e em âmbito estadual, sendo atualmente o estado de Mato Grosso o maior produtor nacional, cultivando 157 mil hectares e produzindo 125,4 mil toneladas.

Assim, o governo necessita de métodos eficazes para o acompanhamento da produção agrícola, na geração de estatísticas confiáveis que subsidiem as negociações locais e internacionais que mitiguem a ação especulativa. Essa iniciativa gera mais estabilidade e previsibilidade na rentabilidade do setor produtivo, fomentando a expansão da produção. No entanto, ainda para o autor, o sistema atual de previsão de safras utilizado no Brasil é dependente de informações subjetivas e marginais, obtidas por meio da avaliação de técnicos que percorrem os municípios produtores, coletando informações entre agentes da cadeia produtiva. Apesar da experiência dos agentes envolvidos, essa metodologia é sujeita a interesses particulares, gerando questionamentos sobre a qualidade dos números apresentados (FERNANDES, 2009).

Dessarte, Jamali (2019), depreende que o mapeamento e monitoramento da cobertura da terra são relevantes para monitorar as mudanças ambientais como um fator essencial para estudar as mudanças ad cobertura da terra como entrada primordial para várias associações científicas que trabalham em investigações de mudanças climáticas, desenvolvimento sustentável, geomorfologia, gestão de recursos naturais e monitoramento de terras agrícolas.

Nesta perspectiva, Fernandes (2009), destaca que o uso de técnicas de sensoriamento remoto pode contribuir significativamente no monitoramento de áreas agrícolas, permitindo obter metodologias eficazes e de baixo custo, especialmente no Brasil que possui grande extensão territorial e volume de produção.

Para auxiliar na distinção de alvos similares é recomendável a utilização de série temporal, pois as culturas agrícolas são caracterizadas por apresentar alta dinâmica temporal, sendo recomendável um acompanhamento dos alvos por meio de séries temporais de imagens de satélite (VERICA, 2018).

O monitoramento da superfície terrestre e da dinâmica de uso por meio de técnicas de sensoriamento remoto é de grande importância para a agricultura. Atualmente, um dos contextos que envolvem as culturas agrícolas se relaciona à estimativa de áreas cultivadas. A dinâmica temporal das culturas, tanto a soja como outras, requerem dados atualizados além de se fazer necessário o uso de tecnologias e métodos para análise da grandeza desses dados (SILVA JUNIOR et al., 2014).

Embora durante algumas fases do seu desenvolvimento as culturas possam assemelhar-se espectralmente a outras classes de cobertura vegetal da superfície terrestre, seu peculiar comportamento espectral, decorrente da dinâmica fenológica, pode ser usado para realçar diferenças em relação às classes fonologicamente mais estáveis ao longo do tempo, como a vegetação nativa (BERNARDES et al., 2011).

Para áreas agrícolas com maior extensão espacial, o problema de identificação é minimizado em relação a imagens de resolução espacial e diferentes estágios fenológicos da cultura, que pode ser espectralmente diferenciada por meio de imagens multi-temporais (SILVA JUNIOR et al., 2017).

É possível extrair informações desses alvos porque estes em uma cena refletem, absorvem e emitem radiação eletromagnética de uma maneira diferente, dependendo de sua composição e forma molecular (VALLS, 2009). Trabalhos recentes utilizando imagens do satélite Landsat-8 OLI (Operational Land Imager) por meio do parâmetro NDVI (Normalized Difference Vegetation Index), obtiveram resultados satisfatórios na classificação da cobertura vegetal e alterações de uso

do solo (HERNANDEZ et al., 2015; LEITE; SANTOS; SANTOS, 2017; WANG et al., 2018; MOAYEDI et al., 2020).

Entretanto, com a utilização de séries temporais a dificuldade atrelada ao tempo necessário para o processamento dos dados é aumentada. Uma solução para contornar isto, é por meio da automatização dos processos, de modo que este possa ser realizado várias vezes ao longo dos anos, tornando-o mais dinâmico (ZHONG et al., 2016). Métodos de aprendizado de máquina têm sido o foco de intensas pesquisa nos últimos anos, com aplicações em campos tão diversos como química, física, biologia, ciência dos materiais e engenharia (DELHOMMELLE, 2018).

O principal objetivo do aprendizado múltiplo de máquinas é mapear dados dimensionais em uma dimensão inferior, preservando as principais características dos dados originais para uma melhor análise. Em sensoriamento remoto, a aplicação de métodos ativos de aprendizagem que visa selecionar as amostras mais relevantes para o treinamento é bastante recente (VALLS, 2009).

2. OBJETIVOS

2.1 Objetivo Geral

Estimar a área de cultivo da cultura de gergelim no município de Canarana-MT na safra 2021/2022 através de algoritmos de aprendizado de máquinas (Machine Learning).

2.2 Objetivos Específicos

Avaliar o desempenho dos algoritmos Support Vector Machine e Random Forest na classificação da cultura do gergelim;

Estimar a área cultivada de gergelim na safra 2021/2022 por diferentes algoritmos;

Testar e comparar as precisões dos mapas temáticos resultantes das classificações, através dos métodos estatísticos Exatidão Global e Índice Kappa.

3. JUSTIFICATIVA

A disponibilidade de informações oportunas e confiáveis sobre o uso da terra está se tornando mais importante do que nunca nos processos de apoio à decisão em vários níveis, tanto dentro de um país como entre países (LU et al., 2012). Ainda para o autor, as estimativas anuais para a maioria dos tipos de commodities agrícolas geralmente estão disponíveis como dados estatísticos em nível de país (por exemplo, unidade). No entanto, o usuário de tais dados permanece desinformado sobre onde, ou seja, a localização geográfica onde essas commodities são cultivadas.

A localização de safras bem como sua extensão espacial, são informações úteis para melhorar planejamento da agricultura e gestão dos recursos hídricos para irrigação, bem como para o monitoramento da saúde das culturas, para o aumento da produção de alimentos e redução do uso de água (JAMAL-EDDINE et al., 2018).

Desta forma, os esforços em desenvolver cultivares mais produtivas e resistentes a pragas, doenças e clima adverso, tem levado a pesquisa a níveis mais profundos, e a Empresa Brasileira de Pesquisa Agropecuária-EMBRAPA, é pioneira no desenvolvimento de pesquisas em técnicas de cultivo, colheita, armazenagem e cultivares para as principais regiões produtoras (EMBRAPA, 2007b; EMBRAPA, 2008; EMBRAPA, 2015).

Para Jansen e Di Gregorio (2003), um exemplo de fatores influentes na economia de um país é o uso de incentivos econômicos, como os subsídios à cana-de-açúcar que influenciaram uso da terra no Brasil e a Política Agrícola da União Europeia que alterou os padrões de uso/cobertura da terra.

Quando há a presença de políticas públicas atuantes no meio rural, observa-se o desenvolvimento e consequente avanço nas produções propostas. Um exemplo de êxito a nível regional para Paiva e Aragão (2020) é o que aconteceu no estado de São Paulo, o maior produtor de cítricos do país. O agronegócio citrícola é responsável pela geração de renda e empregos de maneira direta e indireta, sendo uma atividade de grande importância para o estado e para o Brasil. Dessa forma, é importante que haja dados confiáveis para formulação das políticas públicas necessárias.

A efetividade das políticas públicas depende de uma série extensa de fatores, mas a informação estatística cumpre papel instrumental relevante em todas as fases de implementação de um programa público, da formulação à avaliação deste (JANNUZZI, 2018).

A precisão e o tempo da previsão são aspectos importantes aos tomadores de decisão a ajustar sua gestão adequadamente. As previsões são valiosas quando uma mudança na decisão (por exemplo, compra a termo) leva a uma mudança desejável no resultado, por exemplo, maior lucratividade, o que permite ao tomador de decisão passar de um ponto de aceitação passiva da previsão para um ponto de resposta ativa à previsão (POTGIETER et al., 2010).

Uma característica peculiar da região onde está localizado o município de Canarana no estado de Mato Grosso, é o comportamento do regime chuvoso. Com base na classificação climática de Köppen, Alvares et al., (2013) descreve que o município está localizado na porção leste do estado, apresenta verão chuvoso de novembro a abril, e inverno seco de maio a outubro. No entanto, dada as mudanças climáticas observadas no planeta Terra, a sistemática climática da região tem se apresentada alterada nas últimas safras e, conforme relato dos produtores rurais locais, o período chuvoso tem se iniciado tardiamente e encerrado prematuramente, comprometendo a produção das culturas de segunda safra, como o milho.

Diante disso, a cultura do gergelim tem sido uma opção viável para substituir culturas de segunda safra, uma vez que a oleaginosa é mais tolerante a escassez hídrica quando comparada a cultura do milho.

Assim, a proposta de identificar e mapear as áreas de cultivo de gergelim no município de Canarana-MT tem por finalidade gerar dados estatísticos robustos que poderão subsidiar políticas públicas, dando respaldo científico à estas, uma vez que tanto o setor produtivo quanto o poder público não possuem dados confiáveis da área cultivada.

4. REVISÃO DE LITERATURA

4.1 A cultura do gergelim no Brasil

O gergelim (*Sesamum indicum* L.), da família Pedaliácea, é uma das plantas oleaginosas mais antigas e usadas pela humanidade. Considera-se a África o continente de origem do gergelim, porque ali existe a maioria das espécies silvestres do gênero *Sesamum*, ao passo que na Ásia se encontra uma riqueza de formas e variedades das espécies cultivadas (EMBRAPA, 2007a).

Dada sua resistente à seca, o gergelim pode produzir com um mínimo de 300 mm de pluviosidade bem distribuída durante o ciclo da cultura, sendo a faixa ótima entre 500 e 650 mm (GRILO JR e AZEVEDO, 2013).

A exploração comercial do gergelim no Nordeste brasileiro teve início no ano de 1986, dado a esse fator comercial, houve aumento das áreas cultivadas nessa região, ocorrendo graças a projetos de pesquisa e fomento desenvolvido pela EMBRAPA ALGODÃO nos estados do Ceará, Rio Grande do Norte e Paraíba. Esses projetos tinham como iniciativa apresentar a cultura do gergelim como alternativa à redução drástica do cultivo do algodão, ocasionado principalmente pelo ataque do bicudo, o que levou a queda da produtividade, além da falta de modernização nos plantios nordestinos. Entretanto, na região Centro-Sul, essa cultura vem sendo cultivada desde os anos 50 (NASCIMENTO, 2019).

O gergelim sempre foi visto como uma cultura secundária no Brasil, porém com a demanda crescente do gergelim por indústrias no ramo alimentar, químico e de cosméticos, observa-se escassez do produto tanto no mercado interno, quanto no internacional (VASCONCELOS et al., 2010).

4.2 Fases fenológicas da cultura do gergelim

A fenologia é uma atividade descritiva, que fornece informações úteis para se avaliar a adaptabilidade das espécies e para se definir metas preliminares nas operações de manejo das plantas (REGO e LAVARONTI, 1995).

A cultivar “BRS Seda”, por exemplo, inicia a floração aos 35 dias e com ciclo precoce de 85 a 90 dias EMBRAPA (2009), enquanto o gergelim “BRS Anahí‟ inicia sua floração aos 39 dias e ciclo médio de 90 dias (EMBRAPA, 2015).

A fenologia da cultura do gergelim é distribuída em quatro fases distintas (Tabela 1).

Tabela 1. Fenologia da cultura do gergelim.

Fase Fenológica	Estádio Vegetativo
Fase I	Início da emergência
Fase II	Início do crescimento e desenvolvimento
Fase III	Início da floração e frutificação
Fase IV	Início da maturação / colheita

Fonte: EMBRAPA ALGODÃO, 2007.

5. SENSORIAMENTO REMOTO AGRÍCOLA

5.1 Mapeamento de alvos agrícolas

O sensoriamento remoto consiste na interação dos alvos terrestre (vegetação, solo, água, etc..) com a radiação eletromagnética, sendo essas interações divididas em três componentes: a refletida, a absorvida e a transmitida (JOSHI et al., 2016).

No sensoriamento remoto aplicado à agricultura, o alvo de maior interesse nos estudos é a vegetação, o que viabiliza o monitoramento do crescimento e o desenvolvimento das culturas, além de permitir detectar condições de estresse que limitam a produtividade. Desta forma, esta tecnologia permite obter informações espaciais e temporais detalhadas sobre o comportamento dos alvos nos diferentes ambientes (FERNANDES, 2009).

As culturas agrícolas apresentam um ciclo fenológico bem definido, o monitoramento desta dinâmica necessita de um acompanhamento periódico. Em regiões com grandes dimensões o sensoriamento remoto possui grande utilidade, uma vez que possibilita acompanhar a dinâmica agrícola, o caráter global, multiespectral e repetitivo (SANCHES et al., 2005).

A aplicação do sensoriamento remoto voltado para agricultura vem mostrando seu potencial com trabalhos realizados em mapeamento e monitoramento de culturas (MANABE, 2014).

5.1.1 Comportamento espectral da vegetação de ciclo anual

De acordo com Formaggio e Sanches (2017), um sistema de sensoriamento remoto é composto, basicamente, de uma fonte de radiação eletromagnética (REM) + um sistema plataforma/sensor + objetos sensoriados. O sol é a mais importante fonte natural de REM e, ao interagir com a superfície da Terra, proporciona a ocorrência de diversos fenômenos físicos (FITZ, 2008). Dessa maneira, a resposta espectral da refletância da vegetação refere-se à capacidade da planta em refletir a radiação eletromagnética nos diferentes comprimentos de onda do espectro. (FERNANDES, 2009).

A reflectância é um parâmetro essencial na interpretação das imagens, e de acordo com Meneses, et al., (2019), é praticamente impossível interpretar imagens de sensores remotos caso não se conheça o comportamento espectral da reflectância dos diferentes tipos e classes de rocha, solo, vegetação e água.

Um dos princípios básicos do sensoriamento remoto é que a extração das informações a partir dos produtos gerados pelos sensores remotos é, em geral, baseada nos peculiares comportamentos de reflectância de cada alvo em distintas regiões do espectro eletromagnético (FORMAGGIO e SANCHES, 2017).

As áreas plantadas com as mesmas culturas durante anos sucessivos criam padrões úteis no processo de interpretação para a identificação de alvos agrícolas. Isto ocorre porque há uma tradição regional de cultivo, e fatores como solo, clima e infraestrutura, fazem com que, numa determinada área, ao longo dos anos, seja plantado um mesmo conjunto de culturas (ORTIZ et al., 2010).

O sensoriamento na região espectral compreendida entre o visível e o infravermelho de ondas curtas (400 nm a 2.500 nm) é baseado em medições remotas da REM refletida pelos alvos da superfície terrestre. Assim, as informações coletadas pelos sensores necessitam ser interpretadas com base no pressuposto de que diferentes alvos apresentam distintos comportamentos espectrais (FORMAGGIO e SANCHES, 2017).

A energia eletromagnética no intervalo espectral correspondente ao infravermelho próximo é encontrada no fluxo solar ou mesmo em fontes convencionais de iluminação (lâmpadas incandescentes), enquanto as energias eletromagnéticas correspondentes ao intervalo espectral do infravermelho médio e distante (também denominadas de radiação termal) são provenientes da emissão eletromagnética de objetos terrestres (MORAES, 2002).

As folhas dominam a interação da radiação eletromagnética com as plantas para essa interpretação a partir de observações remotas baseia-se principalmente nas características da folhagem e sua exibição em copas vegetais (WIEGAND e RICHARDSON, 1987).

A radiação incidente será espalhada por componentes das plantas, como as folhas, as hastes, as flores etc., e, dessa maneira, uma parte dessa radiação espalhada deixará o dossel na direção para cima (FORMAGGIO e SANCHES, 2017).

A reflectância espectral de um dossel vegetal depende da interação da radiação eletromagnética com os parâmetros biofísicos e fisiológicos e, por conseguinte, contém uma grande quantidade de informação. No entanto, sendo estes dados volumosos e largamente variando em diferentes comprimentos de onda, deve ser comprimido em índices espectrais para aplicações úteis (GOVIND et al., 2005). A Figura 01 exemplifica o espectro eletromagnético.

Os índices de vegetação espectral referem-se a muitas características do dossel vegetal (índice de área foliar verde, biomassa verde e percentual de cobertura) e agora são reconhecidos como fornecendo boas estimativas da quantidade de tecido fotossintético ativo no dossel a qualquer momento durante a temporada (WIEGAND e RICHARDSON, 1987).

As abordagens baseadas na reflexão espectral têm o potencial de fornecer uma avaliação rápida das diferenças genotípicas na produção de biomassa, índice foliar verde e conteúdo relativo de água, e de prever o rendimento (AJAYI et al., 2016).

As células das folhas possuem dimensões grandes quando comparadas com os comprimentos de onda da REM. Esse fato tem relevância quando consideramos a interação da radiação eletromagnética com as folhas, uma vez que essa interação depende da relação entre essas dimensões e as dos comprimentos de onda. E no interior da folha que ocorre todo o processo espectral de absorção e transmitância, bem como parte da reflectância, a qual também ocorre na superfície. Em outras palavras, os processos de interação da REM com a matéria se realizam nos átomos e moléculas. (MENESES, 2019).

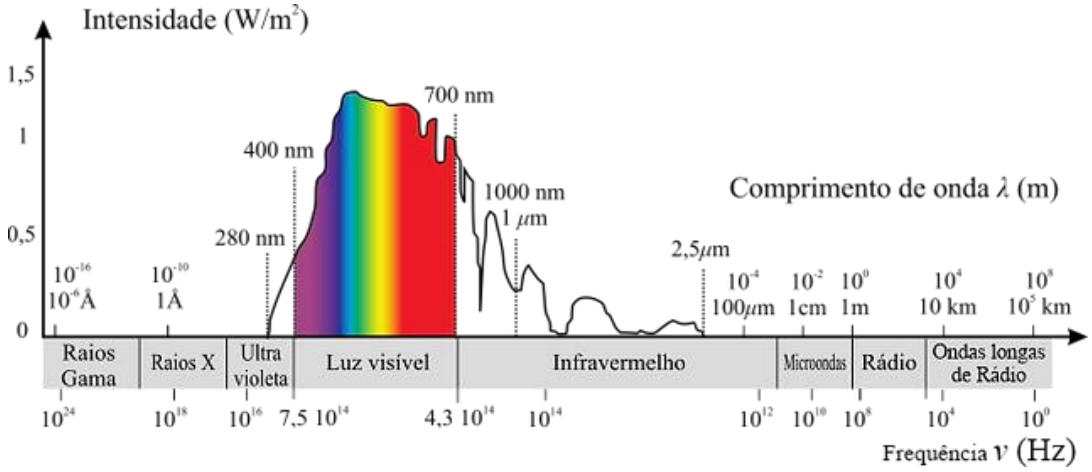


Figura 1. Espectro eletromagnético.
Fonte: Inpe, 2018.

A assinatura espectral do material foliar vivo é dominada pela clorofila, a estrutura celular e o conteúdo de água, que influencia a quantidade de radiação refletida nas regiões do espectro eletromagnético visível (VIS: 350-700 nm), infravermelho próximo (NIR: 700-1.300 nm), e infravermelho de onda curta ou média (SWIR/MIR: 1.300- 2.500 nm) (AJAYI et al., 2016).

Para Jin et al., (2018), o padrão sazonal da vegetação fornece a base para o estudo do mapeamento da cobertura do solo por meio de imagens de sensoriamento remoto de séries temporais. Em comparação com outros tipos de vegetação, como pastagens e florestas, as terras cultivadas têm características próprias nas fases de semeadura, crescimento e colheita. Ao mesmo tempo, com um ou mais picos estreitos, as terras agrícolas são caracterizadas por perfis de NDVI mais irregulares, cujos valores de pico mudam constantemente e substancialmente em comparação com os picos de pastagens e florestas.

As combinações destas assinaturas espectrais são utilizadas para desenvolver uma série de índices de vegetação que podem ser utilizados para interpretar vários parâmetros vegetais (AJAYI et al., 2016). De acordo com Meneses (2019), as imagens aéreas ou orbitais permitem que os padrões de reflectância apresentados pelos dosséis vegetais possam ser minimamente previstos. As variáveis que afetam a reflectância dos dosséis, as quais nem sempre são tão facilmente mensuráveis, podem ser classificadas de acordo com os seguintes fatores: **a) iluminação:** ângulos azimutais e de elevação da luz solar; **b)**

sensor: resoluções espacial, espectral, radiométrica e temporal – visada a nadir e fora do nadir; **c) vegetação:** tipos de vegetação e formas do dossel – índices de área folear e de distribuição angular foliar; **d) solo:** classes de solo e umidade; **e) topografia:** formas de relevo.

O arranjo destas respostas espectrais pode ser utilizado para monitorizar a distribuição espacial das culturas. O índice de vegetação normalizada e outros índices espectrais de vegetação têm sido amplamente utilizado para mapear a estrutura da planta, com base nos diferentes espectros nas regiões do visível e infravermelho próximo e o infravermelho de ondas curtas (ZHAO et al., 2020).

Os valores espectrais entre os valores do visível e infravermelho próximo variam com os pigmentos foliares, e o infravermelho de ondas curtas está relacionado com a concentração de água e azoto e recentemente diversas pesquisas (SILVA JUNIOR et al., 2014; SILVA JUNIOR, et al., 2017; AZEVEDO et al., 2021; RYU; OH; CHO, 2021; WIJEWARDANA et al., 2019; ZHONG et al., 2021; (L. Wang et al., 2021); XIONG et al., 2021 ;HUO; PERSSON; LINDBERG, 2021; LIU et al., 2021) tem obtivo resultados promissores explorando as respostas espectrais da vegetação.

6. SENSORES ORBITAIS

6.1 Panorama geral dos sensores de órbita heliossíncrona

Conforme Fitz (2008), os sensores podem ser entendidos como dispositivos capazes de captar a energia refletida ou emitida por uma superfície qualquer e registrá-la na forma de dados digitais (gráficos, imagens, dados numéricos etc.). Os sensores podem ser classificados de formas diferentes de acordo com a origem da fonte de energia que este utiliza, caracterizando-os como ativos ou passivos.

Fitz (2008) ainda define que os sensores ativos são aqueles que possuem uma fonte de energia própria; já os sensores passivos, por sua vez, não possuem fonte própria de energia e necessitam de fontes externas para a captação da reflexão dos alvos, como a energia solar.

Outra forma de classificar os sensores é em função do produto gerado, sendo os não imageadores que traduzem os dados coletados sob a forma de

gráficos e dados digitais diversos; por outro lado, os sensores *imageadores* são aqueles que traduzem a informação coletada na forma de uma imagem, semelhante a uma fotografia (FITZ, 2008).

A órbita heliossíncrona ou também chamada de órbita polar é geralmente uma órbita de médias ou baixas altitudes (aproximadamente 700 a 800 km) em comparação com o órbitas geo-síncronas. Consequentemente, o período de órbita é menor, o que normalmente varia de 90- 103 minutos. Portanto, os satélites nas órbitas polares fazem mais de uma revolução ao redor a terra em um único dia (KUMAR, 2018). A Figura 2 mostra as órbitas polar e geoestacionária dos satélites.

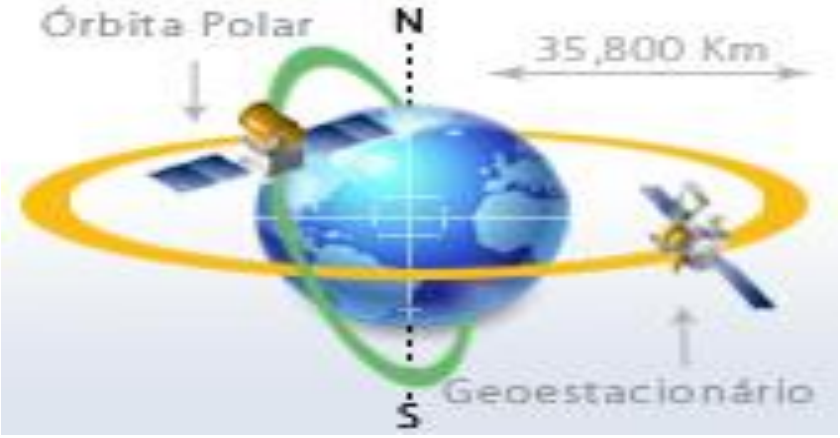


Figura 2. Ilustração do trajeto típico de satélites em órbita polar e geoestacionário. Fonte: Inpe, 2021.

Um ciclo orbital é concluído quando o satélite refaz seu caminho, ou seja, quando o ponto nadir de o satélite passa sobre o mesmo ponto da superfície da Terra pela segunda vez. O ciclo orbital não precisa ser o mesmo que o revisite o período (KUMAR, 2018).

O período de revisita é o tempo decorrido entre duas visualizações sucessivas da mesma área por um satélite. Usando sensores direcionáveis, um instrumento transmitido por satélite pode visualizar áreas fora do nadir antes e depois que a órbita passa sobre um alvo. Em vista desta capacidade de visualização fora do nadir dos satélites, o tempo de revisita pode ser menor do que o ciclo orbital (KUMAR, 2018).

6.1.1 Landsat-8 (OLI)

Desde 1972, os satélites Landsat têm coletado dados consistentemente sobre nosso planeta para o benefício dos EUA e do mundo (NASA, 2021). A superfície terrestre da Terra é longa o suficiente para ser útil para monitorar tendências e avaliar como o uso da terra pode ser mudado; e dado a isso, as aplicações dos dados Landsat-8 incluem agricultura, mapeamento de terras e detecção de mudanças para usos em águas doces e costeiras, neve e gelo (KNIGHT e KVARAN, 2014).

O satélite Landsat-8 tem operado com sucesso em órbita desde que foi lançada em 11 de fevereiro de 2013, e a verificação inicial do observatório Landsat-8 ocorreu conforme planejado até o final de maio de 2013 sem problemas significativos. A Agência Espacial Norte Americana transferiu as operações de observatório para o Serviço Geológico dos Estados Unidos (USGS) que desde então, tem gerido tais operações. Existem dois imageadores a bordo do Landsat-8: o Operational Land Imager (OLI) e o Thermal Sensor Infravermelho (TIRS) (MORFITT et al., 2015).

O Operational Land Imager (OLI) mede as porções visível, infravermelho próximo e infravermelho de onda curta do espectro, tendo suas imagens resoluções espaciais pancromáticas de 15 metros (49 pés) e multiespectrais de 30 metros (98 pés) ao longo de uma faixa de 185 km (115 milhas) de largura, cobrindo amplas áreas da paisagem da Terra, fornecendo resolução suficiente para distinguir características como centros urbanos, fazendas, florestas e outros usos da terra. A Terra inteira ficará à vista uma vez a cada 16 dias devido à órbita quase polar do Landsat-8 (NASA, 2021).

Ainda conforme a NASA (2021), o sensor OLI usa arranjos de detectores longos, com mais de 7.000 detectores por banda espectral, alinhados ao longo de seu plano focal para visualizar através da faixa. Este design de "vassoura" resulta em um instrumento mais sensível, fornecendo informações aprimoradas da superfície do solo com menos peças móveis. Com uma relação sinal-ruído melhorada em comparação com os instrumentos Landsat anteriores, o OLI é mais confiável e oferece desempenho aprimorado.

Nos estudos de Chen e Zhu (2020), usaram 210 imagens dos satélites Landsat-7 ETM+ e Landsat-8 OLI, no período de 8 de abril de 2013 a 26 de

dezembro 2018, para a avaliação a consistência destes satélites no monitoramento aprimorado da matéria orgânica dissolvida em águas complexas, no Rio Sacramento e na Baía de São Francisco do Norte, no estado da Califórnia nos Estados Unidos da América. Os autores concluíram que o Landsat-8 possui alta consistência para derivar baixa concentração da matéria orgânica dissolvida.

Ahmadian et al., (2016), estudaram alguns parâmetros biofísicos (biomassa fresca, índice de área foliar e água na vegetação) das culturas agrícolas: trigo de inverno, cevada e canola, em uma área localizada no nordeste da Alemanha, avaliando e comparando dados reais de campo (*in-situ*) com dados oriundos do processamento de imagens Landsat-7 e Landsat-8 usando o índice de vegetação. De resultado, os autores concluíram que os dados de ambos os satélites apresentaram boa concordância com os dados *in-situ*.

Contudo, ainda conforme Ahmadian et al., (2016), os índices de vegetação derivados do Landsat-8 OLI foram geralmente mais consistentes com os dados *in-situ* do que o Landsat-7 ETM+. Também foi observado que o Landsat-8 OLI apresentou maior capacidade de capturar a pequena variabilidade dos índices de vegetação devido a relação sinal-ruído mais alto e faixa radiométrica mais ampla; além disso, o fenômeno de saturação ocorreu mais cedo para o Landsat 7 ETM+ do que para o Landsat-8 OLI. Isto posto, os autores inteiram que o Landsat-8 OLI apresentou maior desempenho em estimar os parâmetros biofísicos para safras.

Chemura e Mutanga (2016), conduziram um estudo no Distrito 19 de Chipinge, na sub-bacia hidrográfica de Mossurize, no sudeste do Zimbábue. A área é predominantemente composta por fazendas de café em grande escala e o sistema de cultivo é o Café-Sol, o que significa que as plantações são cultivadas exclusivamente, não misturando com árvores de sombra. Este estudo tive por objetivo avaliar o potencial do satélite Landsat-8 OLI usando o algoritmo *Random Forest,* na discriminação e mapeamento das plantações de café de diferentes idades em uma paisagem agrícola heterogênea em comparação com o Landsat-7ETM+.

Neste experimento, Chemura e Mutanga (2016), concluíram que a utilização do Landsat-8 OLI é ideal para mapear esta classe em paisagens heterogêneas, apresentando precisão geral de 70,3% para o sensor OLI e 51,8%

para o sensor ETM+. Ademais ao epílogo, os dados Landsat-8 OLI quando combinados com o classificador Randon Forest foram capazes de produzir mapas temáticos específicos para a idade do cafezal, sendo úteis para tomada de decisões de manejo em nível de campo.

Nos estudos conduzidos por Rawat et al., (2019), foi desenvolvida uma abordagem integrada para estimar a umidade superficial de terras agrícolas na região de Bathinda, no estado de Punjab na Índia. O estudo teve efeito comparativo em avaliar através do processamento de imagens, o quão os dados dos satélites Sentinel-1A e Landsat-8 OLI são satisfatórios quando comparados com dados das medições da umidade do solo *in situ.* Tanto as aquisições das imagens dos satélites quanto a medição da umidade do solo *in situ* foram realizadas na mesma data: 24 de janeiro, 25 de fevereiro e 13 de março de 2018.

Dessa maneira, Rawat et al., (2019), concluíram que o Landsat-8 apresentou ótimos resultados (R^2=0,98) quando implementado o Normalized Difference Vegetation Index (NDVI) para remover a influência que a cobertura vegetal tem sobre o solo. Por fim, os pesquisadores ratificaram que a abordagem teve melhor desempenho quando os valores de NDVI são inferiores a 0,55; quando o índice é superior, não houve qualquer resultado, especialmente para a cultura do trigo.

7. PROCESSAMENTO DIGITAL DE IMAGENS

7.1 Correção atmosférica de imagens

A radiação solar é afetada por fenômenos de absorção e emissão durante a sua trajetória descendente desde o sol até à superfície da Terra e durante a trajetória ascendente detectada por sensores de satélite. Isto leva à distorção das propriedades radiométrica terrestres (reflectância) registadas pelas imagens de satélite (LÓPEZ SERRANO et al., 2016).

Essa distorção, também chamada de espalhamento, para Lorenzzeti (2015), tem sua origem no processo de espalhamento causado pelos componentes atmosféricos, tais como moléculas dos gases constituintes, as partículas, os aerossóis, a poeira e os produtos resultante da queima de combustíveis fósseis.

Assim, Lorenzzeti (2015) expõe que o processo de espalhamento da luz na atmosfera é muito importante para o sensoriamento remoto, pois além de reduzir o fluxo radiante na direção de propagação, faz a radiância observada pelo satélite no campo instantâneo de visada do sensor conter fluxo difuso, que vem diretamente da atmosfera, ou é proveniente de fora do pixel em observação.

Ponzoni et al., (2012), esclarece que cada sensor, em cada banda espectral, tem seu próprio critério para discretizar os valores de radiância medidos na escala específica de sua resolução radiométrica (8 bits, 10 bits, 16 bits etc.). Assim, as imagens resultantes, ainda que obtidas por um mesmo sensor, mas em diferentes bandas, não apresentam necessariamente compatibilidade entre os números digitais (ND).

Dessa maneira, como forma de solucionar essa limitação, faz-se necessária a conversão dos números digitais para valores físicos, mediante o conhecimento de algumas características tanto do sensor que gera as imagens como das condições ambientais nas quais as imagens foram geradas.

Os índices radiométricos são instrumentos importantes para o monitoramento de florestas, vegetação urbana, água ou qualquer outra superfície terrestre que utilize imagens de satélite. Um subconjunto de índices radiométricos são índices de vegetação que são uma parte importante de qualquer pesquisa sobre vegetação e que se tornaram uma das principais fontes de informação para monitoramento da vegetação (LUKA et al., 2019).

Assim, na geração das imagens índices de vegetação é importante a transformação dos números digitais para valores físicos, como radiância, de modo a obter valores comparáveis com os trabalhos disponíveis na literatura (PONZONI et al., 2012). Além disso, este autor salienta que a não conversão dos números digitais das imagens em valores físicos na elaboração de imagens índices de vegetação pode levar a erro grave, pois os números digitais não estão em uma mesma escala radiométrica nas diferentes bandas, o que implica que um determinado valor de número digital, em uma determinada imagem de uma banda espectral específica, não corresponde à mesma intensidade de radiação medida ou representada pelo esmo valor de número digital em uma imagem de outra anda espectral.

Assim, Cooley et al., (2002) expõe que o algoritmo de correção atmosférica FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) é um pacote de software desenvolvido pelo Air Force Research Laboratory, Space Vehicles Directorate (AFRL/VS), Hanscom AFB e Spectral Sciences, Inc. (SSI) para apoiar as análises de sensores de imagem hiperespectrais e multiespectrais de infravermelho visível a ondas curtas (Vis - SWIR).

O objetivo do FLAASH é eliminar os efeitos atmosféricos causados por espalhamento molecular e particulado e absorção das medições de radiância no sensor, a fim de recuperar os valores de refletância na superfície. Dessa forma, o algoritmo FLAASH utiliza o código de transferência radiativa MODTRAN4 e executa a correção para o efeito de adjacência (mistura de pixel devido ao espalhamento de superfície) bem como computa a visibilidade média da cena, operando o algoritmo na faixa espectral entre 0,4μm e 2,5 μm, este método é feito pixel a pixel (FELDE et al., 2003). Este algoritmo de máscara de nuvem FLAASH é baseado na combinação de testes de brilho, equilíbrio de cores e baixa coluna de água nas regiões visível e infravermelho próximo (BOREL et al., 1999).

8. CLASSIFICADORES DIGITAIS NA DETECÇÃO DE ALVOS AGRÍCOLAS

A classificação digital é um processo de reconhecimento de padrões e de objetos homogêneos, representados em um conjunto de pixels, ao qual se aplica o mapeamento de áreas pertencentes a uma única classe de objetos (GAROFALO et al., 2015). Ainda, segundo estes autores, os algoritmos responsáveis pela classificação digital são denominados “classificadores”, e a classificação pode ser feita pixel a pixel ou por regiões/objetos. As classificações pixel a pixel baseiam-se principalmente em parâmetros estatísticos (média, variância etc.) dos pixels de treinamento, a exemplo dos perfis espectrais, para o agrupamento dos pixels remanescentes nas classes predeterminadas (GAROFALO et al., 2015).

Uma gama de classificadores aplicados à classificação supervisionada pixel a pixel tem sido amplamente empregada na classificação de imagens remotamente situadas, e os classificadores Mahalanobis, “maximum likelihood”, “minimum distance” e “parellelepiped” têm sido os mais utilizados, todos de natureza paramétrica (SISODIA et al., 2014).

Neste contexto, trabalhos têm avaliado a acurácia desses classificadores, para a definição de áreas homogêneas a partir de imagens de resolução espacial moderada (LAURIN et al., 2012).

As classificações por região ou orientadas a objetos, do inglês "object-based image analysis" (GEOBIA), utilizam, além de informação espectral de cada pixel, a informação espacial que envolve a relação de pixels circunvizinhos, tais como forma e textura. Assim, o classificador simula a análise de um fotointérprete, ao delimitar áreas homogêneas nas imagens, a partir das características espectrais e espaciais dos objetos que constituem as classes de interesse (GAROFALO et al., 2015).

O modelo paramétrico ou comumente chamado de clássico, emprega parâmetros, como média e matriz de covariância, para definir as regras do classificador para cada classe, o que demanda certa homogeneidade nas propriedades espectrais das classes (ANDREOLA e HAERTEL, 2009).

Quando se fala de classificadores contextuais ou não paramétricos, para Semolini (2002), estes diferem-se dos modelos paramétricos na medida em que podem ser empregados com distribuições que não obedeçam aos parâmetros da curva normal. Na última década, pesquisadores da área do sensoriamento remoto e ciência de dados têm um objetivo comum no uso de aprendizado de máquina avançado em algoritmos para classificação de imagens.

Nos últimos anos, diversos trabalhos (MOUNTRAKIS et al., 2011; POURSANIDIS; et al., 2015; WAHIDIN et al., 2015; GAROFALO et al., 2015) têm demonstrado que as técnicas envolvendo a classificação não paramétrica com o uso de classificadores como Máquinas de Suporte de Vetores (Support Vector Machines – SVM), Random Forest e Redes Neurais Artificiais (Multilayer Perceptron) têm-se destacado na classificação orientada a objetos uma vez que atributos como forma, textura e nitidez apresentam natureza não paramétrica.

8.1 Análise via Machine Learning

Classificadores são utilizados para reconhecimento de classes/padrões, utilizados para rotular a que classe determinado pixel, ou objeto, pertence. O que difere um classificador do outro é a forma como são criadas as fronteiras de decisão a partir das amostras de treinamento (MASTELLA e VIEIRA, 2018). De acordo com

estes autores, a implementação de um robusto sistema de classificação é crucial para o sucesso de mapeamento e acurácia temática, e que a acurácia de um mapa depende de muitos fatores, incluindo nível de detalhes (esquemas de classificação) e a variabilidade das categorias a ser mapeada.

Moayedi et al., (2020) esclarece que a extração e o mapeamento de classes de uso e cobertura da terra são geralmente realizados por meio de algoritmos de classificação de imagens supervisionadas. Esses modelos de classificação supervisionada são usados para prever classes de uso e cobertura da terra de pixels não rotulados na imagem com base na informação espectral e contextual de pixels de amostra ou dados de referência que foram fornecidos com base no conhecimento do usuário das várias classes de uso e cobertura da terra existentes na imagem.

As duas abordagens para treinar e rotular o pixel são conhecidas na literatura como classificação supervisionada e não-supervisionada. Na classificação supervisionada, o analista define suas amostras de treinamento previamente. Assim, o valor espectral de cada pixel no ponto de treinamento é usado para definir o limite de decisão para tal classe. O algoritmo, então, classifica todos os pixels, ou objetos, de acordo com alguma regra de decisão. As amostras da área devem ser representativas às respectivas classes (MASTELLA e VIEIRA, 2018).

8.1.1 Support Vector Machines

Support Vector Machines (SVM) é um conjunto de métodos de aprendizagem supervisionada propostos por Vapnik em 1995 que funcionam minimizando o erro de classificação para maximizar a margem geométrica entre as classes (WALEED et al., 2020). O SVM é um algoritmo de aprendizagem supervisionada de máquinas que se baseia sobre a teoria da aprendizagem estatística (CHAKOURI et al., 2020). A Figura 3 ilustra a forma como a SVM classifica o conjunto de dados.

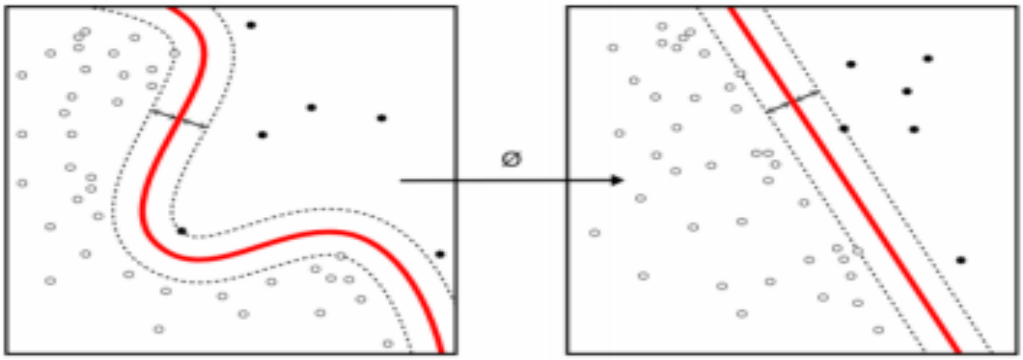


Figura 3. Capacidade do SVM de classificar conjuntos de dados que não são linearmente separáveis.
Fonte: Waleed et al., 2020.

Em um problema de reconhecimento de padrão de duas classes onde as classes são linearmente separáveis, o SVM seleciona o limite de decisão linear que deixa a maior margem entre as duas classes. A margem é definida como a soma das distâncias ao hiperplano dos pontos mais próximos das duas classes (PAL, 2005).

O classificador funciona encontrando o hiperplano certo para separar os pontos de dados em classes obrigatórias. Uma vez que o hiperplano é determinado, as amostras de teste são previstas para estar em qualquer lado do plano. Matematicamente, o hiperplano é dado pela Equação 1:

(1)

$$\mathbf{w} * \mathbf{x} + \mathbf{b} = \mathbf{0}$$

onde: **x** é um vetor de entrada N-dimensional, **w** é um vetor de peso descrito como w = (w1, w2, w3 ... wn), e **b** representa o viés do modelo e é uma quantidade escalar descrita como a distância perpendicular do hiperplano à origem (WALEED et al., 2020). A equação do SVM para a classificação de classes de informação distintas é dada pela equação 2 (JAMAL-EDDINE et al., 2018):

(2)

$$\mathbf{f(x)} = \mathbf{sign}(\sum_{\mathbf{i}}^{\mathbf{N}} \boldsymbol{\alpha}\mathbf{i}\ \mathbf{yi} * \mathbf{K(x, xi)} + \mathbf{b})$$

onde: Xi = vetor de suporte para o caso de treinamento i; Yi = rótulo da classe de informação; K (x, xi) = função do núcleo; B = constante de interação; N = número de vetores de suporte, e; αi = são constantes determinadas pela maximização.

Para Song et al., (2012), em experimentos conduzidos no Condado de Yanqing de Pequim, China, onde a agricultura predomina o uso do solo, principalmente o cultivo de milho e soja; objetivou-se a comparação detalhada do desempenho entre os classificadores Redes Neurais Artificiais e Máquina de Vetor de Suporte, utilizando imagens do sensor SPOT-5.

Song et al., (2012), concluíram que ambos os classificadores se bem treinados, não apresentam diferença significativa quanto a precisão da classificação, contudo, a Máquina de Vetor de Suporte apresentou um desempenho ligeiramente melhor. A análise do efeito do tamanho do conjunto de treinamento destacou que a Máquina de Vetor de Suporte possui grande tolerância em um pequeno conjunto de treinamento e evita o problema de treinamento insuficiente dos classificadores redes neurais artificiais.

Em estudos realizados por Pal e Mather (2005), utilizando imagens multiespectrais (Landsat-7 ETM +) nas áreas de estudo no leste da Inglaterra, localizada próximo a cidade de Littleport, para classificar os dados da imagem em sete tipos de cobertura do solo (trigo, batata, beterraba, cebola, ervilha, alface e feijão); e, nas áreas no centro da Espanha, na região de La Mancha Alta, foram usadas imagens hiperespectrais (DAIS) para classificar oito tipos de cobertura do solo (trigo, corpo d'água, lago de sal seco, vegetação hidrofílica, vinhas, solo descoberto, pastagens e área construída).

Este estudo de Pal e Mather (2005), comparou os classificadores Verossimilhança, Máquina de Vetor de Suporte e Redes Neurais Artificiais. Os resultados demonstraram valores de acurácia de 87.90% e índice Kappa de 0.87 para o classificador Máquina de Vetor de Suporte, sendo superior aos classificadores Verossimilhança (82.9% e Kappa =0.80) e Redes Neurais Artificiais (85.10% e Kappa =0.83). Os autores sugerem que a Máquina de Vetor de Suporte pode alcançar alta precisão de classificação com dados dimensionais elevados, mesmo se o tamanho de o conjunto de dados de treinamento seja pequeno.

8.1.2 Random Forest

O classificador Random Forest é um método de aprendizagem em conjunto, comumente utilizado no caso da classificação por cobertura terrestre, tirando partido de imagens de satélite multiespectral e hiperespectrais. O Random Forest produz várias árvores (Figura 4) de acordo com os padrões aleatórios da base de dados de treino (JAMALI, 2019).

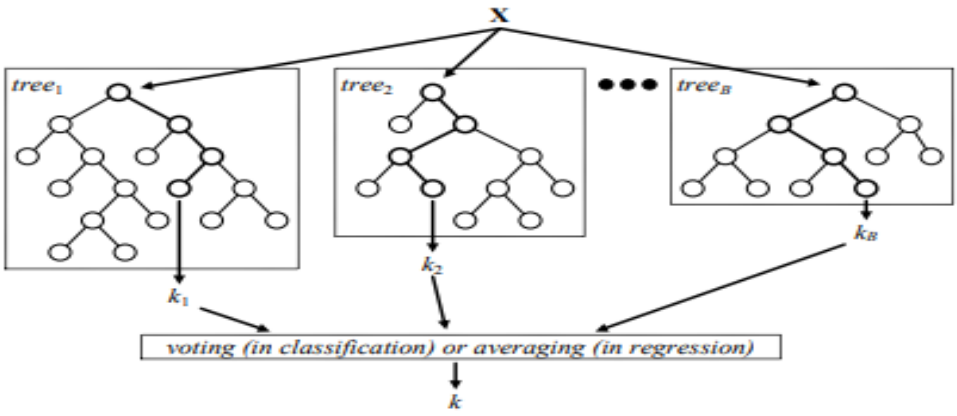


Figura 4. Arquitetura do modelo de Random Forest.

Fonte: Verikas et al., 2016.

Este classificador consiste em uma combinação de classificadores de árvore onde cada classificador gerado é usado em um vetor aleatório amostrado independentemente da entrada vetor, e cada árvore lança um voto de unidade para a classe mais popular para classificar uma entrada vetor (PAL, 2005). O Random Forest é muito popular entre os algoritmos de classificação devido ao seu alto desempenho (WALLED et al., 2020).

Ele utiliza características selecionadas aleatoriamente ou uma combinação de características em cada nó para crescer uma árvore. O ensacamento é um método para gerar um conjunto de dados de treinamento através do desenho aleatório com exemplos de N de substituição, onde N é o tamanho do conjunto de treinamento original (PAL, 2005).

Desse maneira, Jamali, (2019) denota que este classificador um preditor que consiste em uma coleção de árvores de regressão de base aleatória

$\{r_{n,}(X, \theta_n, D_n)$, $m \geq 1\}$. Essas árvores aleatórias são integradas para criar a estimativa de regressão agregada como $\mathrm{r_n}\ (\mathrm{X}, \mathrm{D_n}) = E_\theta[\mathrm{r_n}(\mathrm{X}, \theta, \mathrm{D_n})]$, onde:
E_θ = define a expectativa em relação à parâmetros aleatórios, condicionalmente no X e no conjunto de dados $\mathrm{D_n}$.

A medida de seleção de atributos que será utilizada na indução da árvore de decisão será o critério o Índice de Gini. O classificador Random Forest utiliza o Índice de Gini como uma medida de seleção de atributos, que mede a impureza de um atributo em relação às classes. Para um determinado conjunto de treinamento T, selecionando um caso (pixel) ao acaso e dizendo que ele pertence a alguma classe Ci, o Índice Gini (Pal, 2005), pode ser calculado pela equação 3:

(3)

$$\Sigma \sum_{j \neq i} \left(\frac{f(Ci, T)}{|T|} \right) * \left(\frac{f(Cj, T)}{|T|} \right)$$

onde: $\boldsymbol{f}\ (\boldsymbol{Ci, T})/|\boldsymbol{T}|$ é a probabilidade de que o caso selecionado pertença à classe Ci.

Nos testes conduzidos por Waleed et al., (2020), entre diversos classificadores para detecção automática de oliveiras (Olea europaea) no território espanhol, na comunidade de Castilla La Mancha, que abrange a província de Toledo, região produtora de azeitonas, as amostras de dados que representaram as taxas de treinamento para testes de 50 a 50%, 60 a 40% e 70 a 30%, respectivamente. Os autores observaram que ao aumentar o treinamento na proporção de 50 % a 70%, os classificadores apresentaram aumento nos resultados da classificação.

Na conclusão dos estudos de Waleed et al., (2020), os classificadores Naive Bayesian, Support Vector Machines (SVM), Redes Neurais Artificiais - Multilayer Perceptron e Random Forest, demonstraram uma precisão geral de 79,5%, 90,1%, 92,9% e 96,6%, respectivamente, a uma proporção de 50 a 50% de treinamento. Já na proporção de 70% de treinamento amostrais, resultou em maior precisão, com Naive Bayesian a 80,1%, Support Vector Machines em 92,3%, Multilayer Perceptron em 94% e destaque para o Random Forest com 97,52%.

Para Jamali (2019), em seu estudo comparativo em termos de precisão entre oito algoritmos de aprendizado de máquinas: Random Forest (RF), Decision

Table, Decision Table or Naive Bayes (DTNB), J48, Lazy IBK, Multi-layer Perceptrom, NN ge e Simple Logistic, para classificação de imagens Landsat-8 OLI/TIRS do município de Sari, na província de Mazandaran, no norte do Irã, entre a costa sul do Mar Cáspio. Sari foi escolhido por causa da complexidade da composição da natureza, que incluem planícies, florestas tropicais e pradarias.

De resultado, Jamali (2019) conclui que todos os algoritmos obtiveram excelentes resultados, com destaque para o algoritmo NN ge e Random Forest, com valores de precisão geral (100) e (100); erro médio absoluto (0) e (0.0074) e, erro do quadrado médio da raiz (0) e (0.0476), respectivamente.

Jin et al., (2018), em sua pesquisa, escolheram o Condado de Pingyin, na província de Shandong, por ser uma das maiores áreas agrícolas da China, dominada principalmente por milho, soja trigo e batata, sendo em sua maioria, cultivadas em consorciação. Na pesquisa, foram utilizadas imagens Landsat 5 – TM, no intervalo de abril a outubro do ano de 2009, para avaliar a adição do índice de vegetação com diferença normalizada (NDVI), série temporal, matriz de ocorrência de nível de cinza (GLCM) e variáveis texturais, usando o algoritmo Random Forest para o mapeamento da cobertura da terra na região agrícola complexa.

Os resultados de Jin et al., (2018), demonstraram que adicionar imagens multi-temporais, variáveis texturais, (GLCM) e variáveis de séries temporais NDVI, foi possível obter valores de precisão geral de 89%, e coeficiente Kappa (k) de 0,81. Dessarte, os autores concluíram que para a classificação de regiões complexas, o resultado foi significativo, uma vez que é necessária uma grande quantidade de treinamento de fontes de dados.

9. MATERIAL E MÉTODOS

9.1 Área de estudo

A área de estudo, de acordo com a Figura 5, compreende o município de Canarana no estado de Mato Grosso, localizado na região Centro-Oeste do Brasil, município este situado entre as coordenadas geográficas 12°36'10" a 13°46'56"S e 51°22'51" e 53°06'04" W.

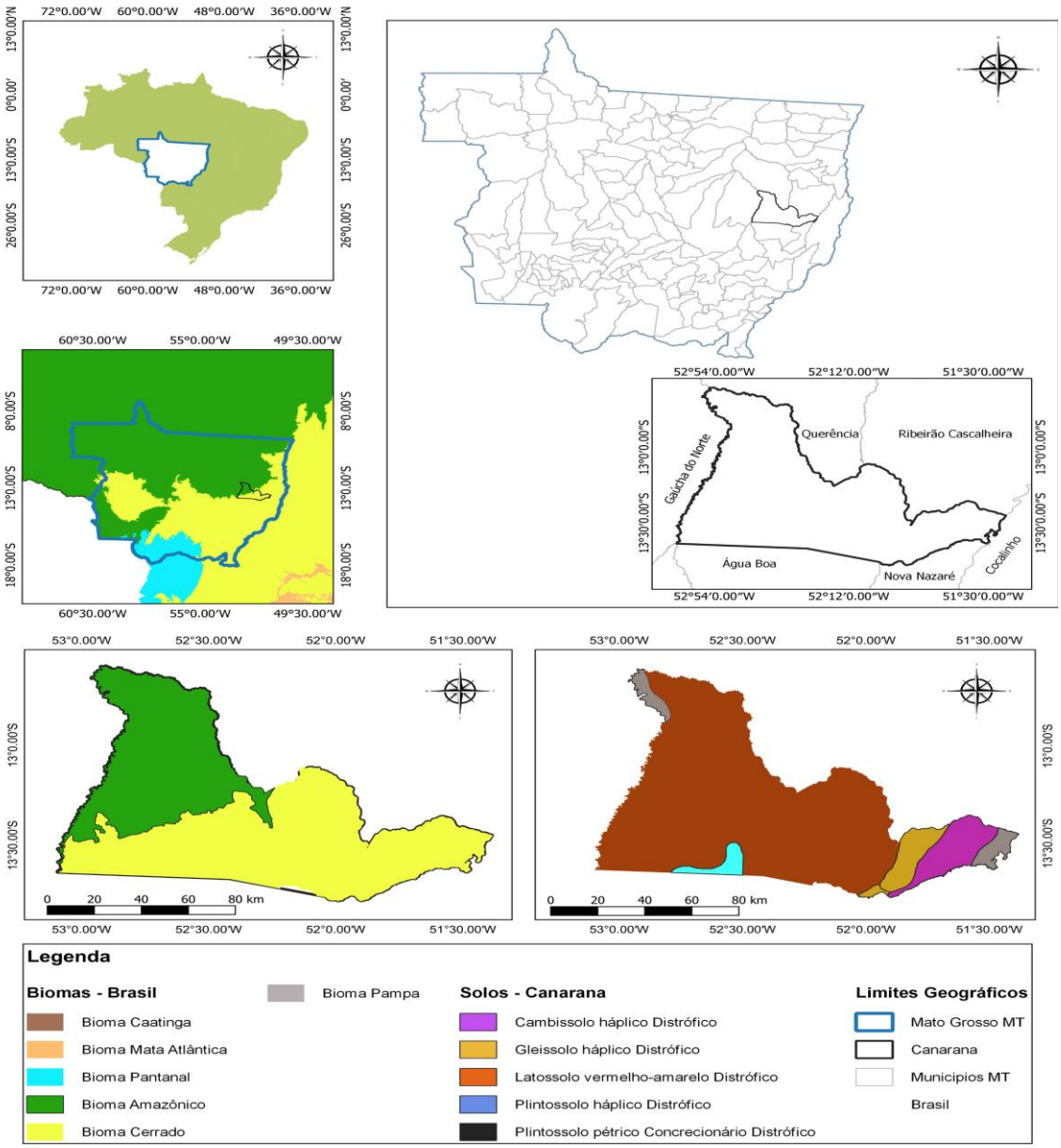


Figura 5. Localização geográfica/ distribuição dos biomas e tipologia dos solos da área de estudo.

Este município está inserido nos biomas do Cerrado e Amazônico e seu território é de aproximadamente 10.855 km² (IBGE, 2020), sendo escolhido para esta pesquisa por sua aptidão agrícola para as mais variadas culturas anuais em segunda safra, sobretudo o gergelim. A altitude média varia entre 200 a 400 metros, predominando de acordo com a classificação climática de Köppen o clima tropical

com inverno seco, apresentando temperatura média entre 24° a 26°C (ALVARES et al., 2013).

Possui estação chuvosa no verão, de novembro a abril, e nítida estação seca no inverno, de maio a outubro. A temperatura média do mês mais frio é superior a 18°C, e as precipitações são superiores a 750 mm anuais, atingindo 1800mm (EMBRAPA, 2021). Na área de estudo, os solos são caracterizados conforme o Sistema de Classificação Brasileiro (SANTOS et al., 2018), como: Cambissolos, Gleissolos, Plintossolos e Latossolos, sendo o último o mais predominante.

9.2 Aquisição das Imagens

A metodologia empregada neste estudo é ilustrada pelo fluxograma apresentado na Figura 6. As imagens foram adquiridas no website do United States Geological Survey - Earth Explorer (USGS). Estas imagens foram obtidas em formato GeoTIF – DATUM WGS 84 (*World Geodetic System 1984*).

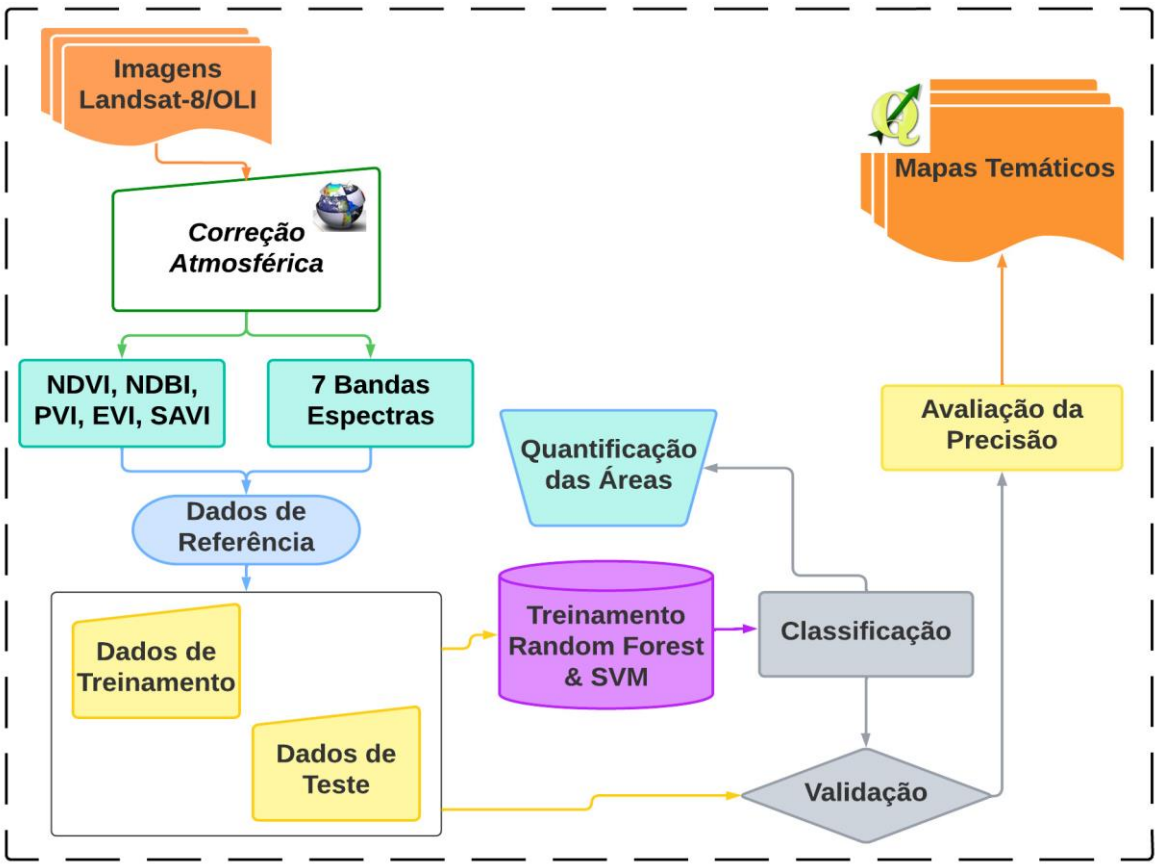


Figura 6. Fluxograma para mapeamento e estimativa da cultura do gergelim.

A escolha da data das imagens foi determinada pelo período de clímax do desenvolvimento vegetativo da cultura do gergelim, de acordo com o calendário agrícola de plantio e colheita de grãos no Brasil. Conforme a Conab (2020b), para o ano-safra 2021/22, o período de clímax do desenvolvimento vegetativo ocorreu no mês de maio do ano de 2022. Outro fator preponderante foi a escolha de imagens em condições reduzidas de cobertura nuvens (<10%), devido ao fato que a presença de nuvens interfere significativamente na reflectância do dossel. Dessa maneira, as imagens selecionadas dataram dos dias julianos 125 e 132, para as órbitas/pontos – 224 e 225/069, respectivamente às datas.

9.3 Pré-processamento

O pré-processamento é o conjunto de operações primárias necessárias à análise prévia da imagem para erradicar ou reduzir os erros radiométricos, atmosféricos e geométricos (MOAYEDI et al., 2020). Para isso, antes da utilização de imagens Landsat-8 OLI na fase de classificação, é necessária a conversão da imagem do número digital para valores de reflectância.

Dessa forma, a conversão dos níveis digitais das imagens para radiância foi utilizado o software ENVI 5.1®, dispondo do algoritmo *radiometric calibration*. Nesta etapa, foram aplicados os parâmetros pré-estabelecidos pelo software para transformação dos níveis de cinza das imagens em radiância. Com as imagens devidamente ajustadas em radiância, no passo seguinte foi realizada a correção atmosférica com o algoritmo FLAASH.

Assim, a correção atmosférica foi realizada utilizando o software ENVI 5.1®, dispondo do algoritmo FLAASH, e para tanto, foi necessário previamente, inserir parâmetros de entrada, dispostos nas Tabelas 2 e 3, para as imagens do satélite Landsat-8/OLI. Esses dados de entrada foram obtidos no arquivo de texto metadata (MTL), adquiridos juntamente com as imagens em formato GeoTiff.

Tabela 2.Parâmetros de entrada do modelo para correção atmosférica da cena órbita/ponto 224-069.

Órbita/Ponto	Parâmetros	Valores de Entrada
224-069	Tipo de sensor	Multiespectral
	Latitude do pixel central	13º 00' 52,20" S
	Longitude do pixel central	51º 55' 33,61"
	Tipo de modelo da atmosfera gasosa	Tropical
	Tipo de modelo de aerossóis	Rural
	Visibilidade (km)	40
	Altitude média (km)	0,372

Fonte: United States Geological Survey - USGS, 2022.

Tabela 3. Parâmetros de entrada no modelo para correção atmosférica da cena órbita/ponto 225-069.

Órbita/Ponto	Parâmetros	Valores de Entrada
225-069	Tipo de sensor	Multiespectral
	Latitude do pixel central	13º 00' 05,64" S
	Longitude do pixel central	53º 29' 18,54"
	Tipo de modelo da atmosfera gasosa	Tropical
	Tipo de modelo de aerossóis	Rural
	Visibilidade (km)	40
	Altitude média (km)	0,390

Fonte: United States Geological Survey - USGS, 2022.

No entanto, para Ponzoni et al., (2012), ainda não é possível a caracterização espectral de um objeto existente na superfície terrestre, uma vez que intrínsecos aos valores de reflectância bidirecional aparente encontra-se os efeitos da atmosfera. Para que tal caracterização seja possível, é necessário

eliminar ou minimizar os efeitos da atmosfera sobre os valores de reflectância bidirecional aparente.

Deste modo, calcular os índices de vegetação é imprescindível. Para isso, foi necessário dispor das informações do sensor OLI do satélite Landsat 8, conforme Tabela 4.

Tabela 4. Características do sensor Operational Land Imager satélite Landsat 8.

Bandas	Região do Espectro Eletromagnético	Frequência	Resolução Espacial
Banda 1	Coastal Aerosol	0.43 - 0.45 µm	30 m
Banda 2	Blue	0.45 - 0.51 µm	30 m
Banda 3	Green	0.53 - 0.59 µm	30 m
Banda 4	Red	0.64 - 0.67 µm	30 m
Banda 5	Near-Infrared (NIR)	0.85 - 0.88 µm	30 m
Banda 6	SWIR 1	1.57 - 1.65 µm	30 m
Banda 7	SWIR 2	2.11 - 2.29 µm	30 m
Banda 8	Panchromatic (PAN)	0.50 - 0.68 µm	15 m
Banda 9	Cirrus	1.36 - 1.38 µm	30 m

Fonte: United States Geological Survey - USGS, 2014.

9.4 Índices de Vegetação

O índice de vegetação (IV), definido como a combinação aritmética de duas ou mais bandas relacionadas às características espectrais da vegetação, tem sido amplamente utilizado para o monitoramento fenológico, classificação e derivação biofísica da vegetação (MATSUSHITA et al., 2007).

Para essa finalidade, foram realizados os cálculos dos índices de vegetação. O *Normalized Difference Vegetation Index* (NDVI), proposto por Rouse Jr et al., (1974), tem sido amplamente utilizado para identificar e extrair vegetação de imagens de detecção remota. O NDVI é matematicamente expresso pela reflectância das bandas do infravermelho próximo (NIR) e vermelho (RED), usando a Equação 4:

(4)

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

O valor do NDVI varia entre -1 a 1. Seus valores são positivos para vegetação, incluindo culturas agrícolas, arbustos, gramíneas e florestas, e, próximos de zero ou negativos, para áreas não vegetadas, incluindo rochas, areia e superfícies de concreto (ZHENG et al., 2021).

O *Soil-Adjusted Vegetation Index* (SAVI) é usado para corrigir o Índice de Vegetação por Diferença Normalizada (NDVI) devido a influência da luminosidade do solo em áreas onde a cobertura vegetal é baixa. O SAVI derivado da Refletância da Superfície Landsat é calculado como uma razão entre os valores RED e NIR com um fator de correção de brilho do solo (L) definido como 0,5 para acomodar a maioria dos tipos de cobertura do solo (USGS, 2015). O SAVI é calculado pela Equação 5:

(5)

$$\mathbf{SAVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED} + \mathrm{L}} * (1 + \mathrm{L})$$

O *Normalized Difference Built-up Index* (NDBI) foi desenvolvido por Zha et al., (2003), para identificar áreas urbanas e construídas, utilizando o infravermelho próximo (NIR – 0,85 a 0,88 μm) e infravermelho médio (SWIR – 1,57 a 1,65 μm). O índice é baseado na adição do número digital das áreas construídas nos intervalos do NIR e SWIR (MORATO et al., 2007). É dado pela Equação 6:

(6)

$$\mathbf{NDBI} = \frac{\mathrm{NIR} - \mathrm{SWIR}}{\mathrm{NIR} + \mathrm{SWIR}}$$

Contudo, Huete et al., (2002), pondera que o *Enhanced Vegetation Index* (EVI) utiliza informações na região espectral do infravermelho próximo, região do vermelho e a reflectância na faixa do azul para indicar o vigor vegetativo das plantas, e este índice permite um melhor monitoramento da vegetação através da redução de substrato do dossel e da influência da atmosfera.

De acordo com o USGS (2015), o EVI é semelhante ao Normalized Difference Vegetation Index (NDVI) e pode ser usado para quantificar o verde da vegetação. No entanto, o EVI corrige algumas condições atmosféricas e ruídos de fundo do dossel e é mais sensível em áreas com vegetação densa. Ele incorpora um valor “L” para ajustar o fundo do dossel, valores “C” como coeficientes de resistência atmosférica e valores da faixa azul (B). Esses aprimoramentos permitem o cálculo do índice como uma razão entre os valores R e NIR, enquanto reduzem o ruído de fundo, o ruído atmosférico e a saturação na maioria dos casos. O EVI é calculado pela Equação 7:

(7)

$$\mathbf{EVI} = \mathrm{G} * \frac{((\mathrm{NIR} - \mathrm{RED})}{(\mathrm{NIR} + \mathrm{C1} * \mathrm{RED} - \mathrm{C2} * \mathrm{BLUE} + 1))}$$

onde: NIR = reflectância na faixa do infravermelho próximo; RED = reflectância na faixa do vermelho; BLUE = reflectância na faixa do azul; C1 = coeficiente de correção dos efeitos atmosféricos para a banda do vermelho (6); C2 = coeficiente de correção dos efeitos atmosféricos para a banda do azul (7,5); L = fator de correção para a interferência do solo (1); G = fator de ganho (2,5).

O *Perpendicular Vegetation Index* (PVI) é baseado em distância e foi projetado para assumir que o espaço perpendicular do pixel da linha do solo é linearmente dependente da área de vegetação; tenta eliminar a diferença em fundo de dossel, calculado pela Equação 8:

(8)

$$\mathbf{PVI} = (\sin(\mathrm{a}) * \mathrm{NIR} - (\cos(\mathrm{a}) * \mathrm{R})$$

onde: (a) é o ângulo entre a linha do solo e o eixo do NIR.

9.5 Dados de Referência

Para reduzir a redundância e a autocorrelação espacial, Ge et al., (2020), sustentam que as amostras (dados de referência) baseadas em pixels para os algoritmos de aprendizado de máquina devem ser selecionadas por meio de visitas de campo, coletando o maior número possível de amostras, valendo-se do

conhecimento prévio do observador, dos variados usos do solo da área amostral. Partindo desse posicionamento, os dados de referência (pontos de controle em coordenadas geográficas) foram cuidadosamente coletados in loco no município de Canarana (Figura 7). A coleta dos dados foi planejada e executada de forma abranger a área de estudo de maneira latitudinal e longitudinal, visando percorrer a área de maneira a viabilizar a coleta do maior número de dados de referência o mais distribuído possível.

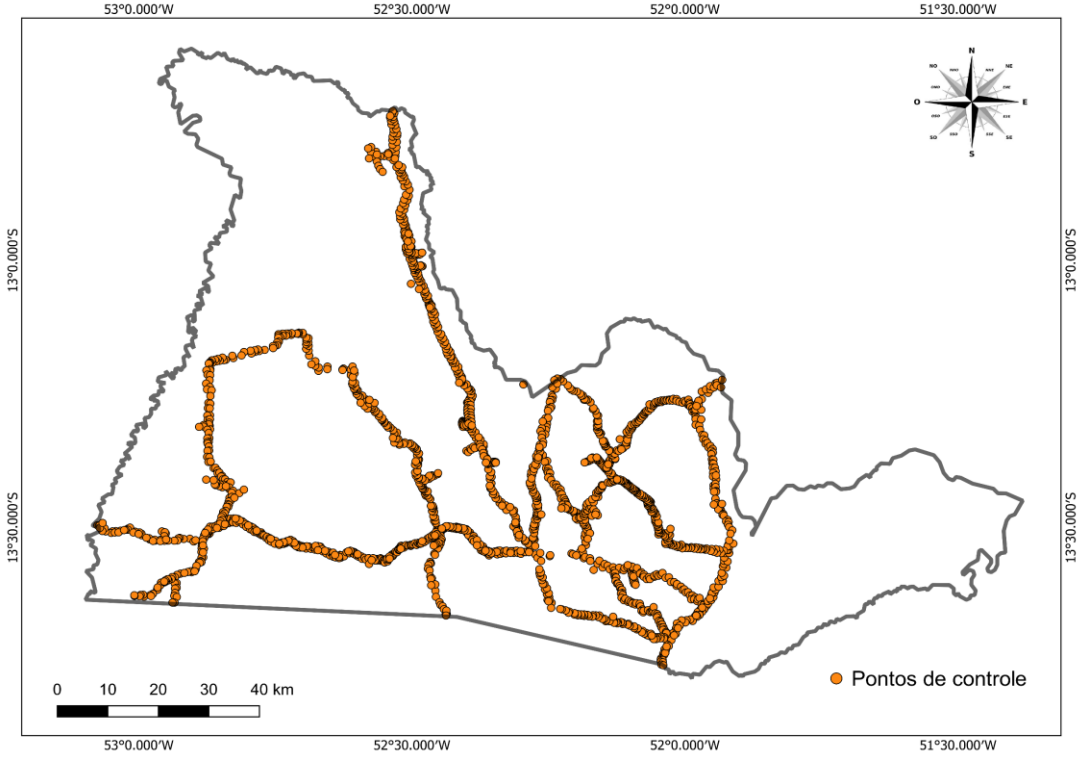


Figura 7. Distribuição dos pontos de controle coletados na área de estudo.

A identificação do uso e cobertura do solo foi realizada visualmente, e, utilizando o aplicativo Android Locus Map 3 Classic – instalado em um dispositivo do tipo Tablet Samsung A16, os dados de referência foram registrados e, ao final, foram coletados 1746 pontos, totalizando 6 classes de uso do solo com registro fotográfico de algumas áreas.

Esses dados, conforme proposto por Moayedi et al., (2020), subsidiam as informações contextuais de pixels das amostras que serão coletadas com base

no conhecimento do pesquisador das várias classes de uso e cobertura do solo existentes na imagem.

9.6 Treinamento e Teste dos Algoritmos

Todo o processo de treinamento e teste foi desenvolvido de forma supervisionada no software Quantum GIS 3.16. Para tal, na fase de treinamento, as imagens multiespectrais do satélite Landsat 8/OLI foram combinadas com os produtos dos índices de vegetação no objetivo de melhorar a qualidade do treinamento, atenuando as prováveis interferências que o fundo do dossel exerce na reflectância do alvo.

A partir das amostras de treinamento, há uma distribuição de probabilidade. Após a remoção de algumas amostras, são verificadas as mesmas distribuições de probabilidade repetidamente, para que sejam estabelecidos, a posteriori, vetores de suporte em um hiperplano n-dimensional (HUANG et al., 2002). Dessa forma, na fase treinamento do SVM, foi adotado como parâmetro de dados entrada a função Kernel, conforme Pal e Mather, (2005) propuseram em caso de dados de entrada em um espaço de recursos de alta dimensão. A função de base radial Kernel, amplamente utilizada, de acordo com Löw et al., (2013), apresenta melhores resultados em comparação com os outros kernels (linear, polinomial ou gaussiano) no mapeamento do tipo de cultura.

Na etapa de treinamento e teste realizada no Quantum GIS, tanto para o Random Forest quanto para o SVM, as amostras foram estratificadas conforme o percentual definido pelo operador. Corroborando com estudos de Jamali (2020) e Moayedi et al., (2020), da totalidade de pontos coletados, 70% foram destinados para fase de treinamento e teste, e, 30% para fase de validação, conforme a Tabela 5.

Tabela 5. Distribuição do conjunto amostral de pontos de controle, de acordo com as classes de uso e ocupação do solo, nas fases de treinamento/teste* e validação** da classificação.

Nº da Classe	Classes	Número de Amostras (pontos de controle)		
		Treinamento/Teste	Validação	Total

1	Gergelim	173	74	**247**
2	Algodão	13	5	**18**
3	Água	33	14	**47**
5	Milho 2ª safra	508	217	**725**
6	Outros usos do Solo	246	105	**351**
7	Floresta Nativa	251	107	**358**
Total		**1.224**	**522**	**1.746**

*Treinamento/teste: corresponde a 70% do conjunto amostral total em cada classe.
**Validação: corresponde a 30% do conjunto amostral total em cada classe.

9.7 Avaliação e Validação da Acurácia

Uma avaliação de precisão começa com a definição da população-alvo, que é a área ou região representada pelo mapa de cobertura da terra. As unidades individuais ou elementos desta população são definidos como pixels ou polígonos, dependendo da representação do mapa (STEHMAN e CZAPLEWSKI, 1998).

Para uma boa avaliação da acurácia temática dos produtos do sensoriamento orbital, são necessários três passos: 1) modelo ou esquema de amostragem; 2) modelo de resposta (uso da melhor regra de decisão), e; 3) estimativa e análise da acurácia temática (índices derivados da Matriz de Erros - ME) (STEHMAN e CZAPLEWSKI, 1998).

Além do mais, para uma boa avaliação da precisão dos produtos temáticos do sensoriamento orbital de uso e mudança de terra, de acordo com Olofsson et al., (2013), devem ser incluídos: 1) estimativas de precisão de mudança; 2) estimativas de área de mudança de terra que ajuste a área de “contagem de pixels” do mapa para erro de classificação, e; 3) intervalos de confiança associados às estimativas dos parâmetros de precisão e mudança de área.

A avaliação de precisão aborda questões relacionadas o quão bem os locais de mudança mapeada correspondem ao real as áreas de mudança (OLOFSSON et al., 2014). A acurácia temática é uma verificação para constatar se

o mapa está correto no tocante ao rótulo das classes, além de aumentar a qualidade da informação do mapa, identificando e corrigindo as fontes de erros. Assim, a avaliação da acurácia determina a qualidade do mapa criado a partir de imagens e produtos do sensoriamento remoto, qualitativa ou quantitativamente, indicando e mensurando os erros do mapa (MASTELLA e VIEIRA, 2018).

A avaliação dos mapas temáticos gerados a partir da classificação proveniente do aprendizado de máquina foi realizada utilizando-se amostras independentes das utilizadas para a construção dos modelos. Para Silva Junior (2014), esta abordagem permite a avaliação entre todos os processos de classificações.

O conjunto de amostras identifica o uso e cobertura do solo da área de estudo, categorizado em seis classes distintas (Tabela 6). A categorização foi realizada agrupando os usos do solo que não possuem relevância à pesquisa e/ou importância comercial à área de estudo e, atribuindo classes únicas e próprias à cultura do gergelim, algodão e milho 2ª safra.

Tabela 6. Definição das classes de uso e ocupação do solo e suas descrições.

Número da Classe	Nome da Classe	Descrição da Classe
1	Gergelim	Cultura do gergelim
2	Algodão	Cultura do algodão
3	Água	Lagos, lagoas e cursos d'água
4	Milho 2ª safra	Cultura do milho
5	Outros usos do solo	Área de expansão agrícola, área em pousio, área em regeneração, área urbanizada, solo exposto
6	Floresta nativa	Cerrado e Floresta Amazônica

A cultura do gergelim por se tratar do objeto alvo deste estudo, recebeu classificação própria para viabilizar a descrição e análise do resultado da classificação de maneira mais concisa.

A cultura do algodão e milho 2ª safra também recebeu classificações únicas por se tratar de culturas de relevante interesse do ponto de vista comercial para a área de estudo. Uma vez que estas culturas são relevantes, conhecer e

estimar a área cultivada torna-se uma importante informação para a elaboração de políticas públicas.

No entanto, não possuindo relevância significativa ao objetivo da pesquisa, os recursos hídricos (lagos, lagoas e cursos d'água) e demais usos do solo também se encontram unificados em classes próprias e distintas, uma vez que estimar a área destes não possui relevância.

Para embasar a adoção do número de classes, bem como a quantidade de amostras (Tabela 8), adotou-se o que Mastella e Vieira (2018) assumem como um mínimo de 50 amostras para cada classe de mapa, para mapas de menos de 1 milhão de acres (aproximadamente 406.000 hectares) em tamanho, e menos de 12 classes. Contudo, com exceção das classes água e algodão, onde não foi possível obter mais amostras por não terem sido identificados outras áreas, as demais classes possuem o número de amostras recomendado pelos autores.

A classificação proposta para a imagem do Landsat 8 OLI baseada nos algoritmos de aprendizagem de máquinas, Random Forest e Support Vector Machine, foi avaliada usando os coeficientes de exatidão global (EG) e índice Kappa (k) e extraídos os erros e a precisão do ponto de vista do produtor e do usuário, conforme proposto por Silva Junior (2014), expressos, respectivamente, nas Equações 9 e 10.

(9)

$$EG = \left(\frac{\sum_{i=1}^{k} x_{ii}}{N} \right)$$

(10)

$$\kappa = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} (x_{i+} . x_{+i})}{N^2 - \sum_{i=1}^{k} (x_{i+} . x_{+i})}$$

onde: $\sum_{i=1}^{k} X$ii = soma dos elementos da matriz em sua diagonal principal; κ = valor estimado Kappa; k = número de linhas; Xii = número de observações na linha i e coluna i; Xi+ = soma dos elementos da matriz em sua diagonal principal; Xi+ = soma total das observações para as linhas; X+i = soma total das observações para as colunas, e; N = número de observações totais.

Por fim, para avaliar qual classificador obteve melhor performance, foi considerada a equivalência entre os dois coeficientes de exatidão provenientes das classificações diferentes $(k1 = k2)$, a estatística do teste foi dada pelas equações 11, 12 e 13.

(11)

$$Z = \frac{(\hat{\kappa}_1 - \hat{\kappa}_2) - (\kappa_1 - \kappa_2)}{\sqrt{\sigma^2(\hat{\kappa}_1) + \sigma^2(\hat{\kappa}_2)}} \sim N(0,1)$$

(12)

$$Z = \frac{\hat{\kappa} - \kappa}{\sqrt{\sigma^2(\hat{\kappa})}} \sim N(0,1)$$

(13)

$$\sigma^2(\hat{\kappa}) = \frac{1}{n}\left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta)_2^2}{(1-\theta_2)^4}\right]$$

onde:

$$\theta_1 = \frac{\sum_{k=1}^{c} x_{kk}}{n}; \theta_2 = \frac{\sum_{k=1}^{c} x_{k+} x_{+k}}{n^2}; \theta_3 = \frac{\sum_{k=1}^{c} x_{kk}(x_{k+} + x_{+k})}{n^2}; \theta_4 = \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} x_{ij}(x_{i+} + x_{+j})}{n^3}.$$

Para tanto, foi adotado como hipótese nula (H0) a uma igualdade entre as classificações e o inverso para as suas diferenças (H1), a um nível de 0,05 de significância ($p<0,05$) (SILVA JUNIOR, 2014).

10. RESULTADOS E DISCUSSÃO

A classificação resultou em mapas temáticos (Figuras 8 e 9) onde é possível visualizar as classes de uso e ocupação do solo classificadas. Observou-se que os classificadores apresentaram resultados com algumas divergências. Na Figura 8, o Random Forest em algumas áreas distinguiu melhor as classes de uso do solo quando comparado ao SVM que, em determinadas áreas, classificou erroneamente a classe algodão no interior de áreas de gergelim.

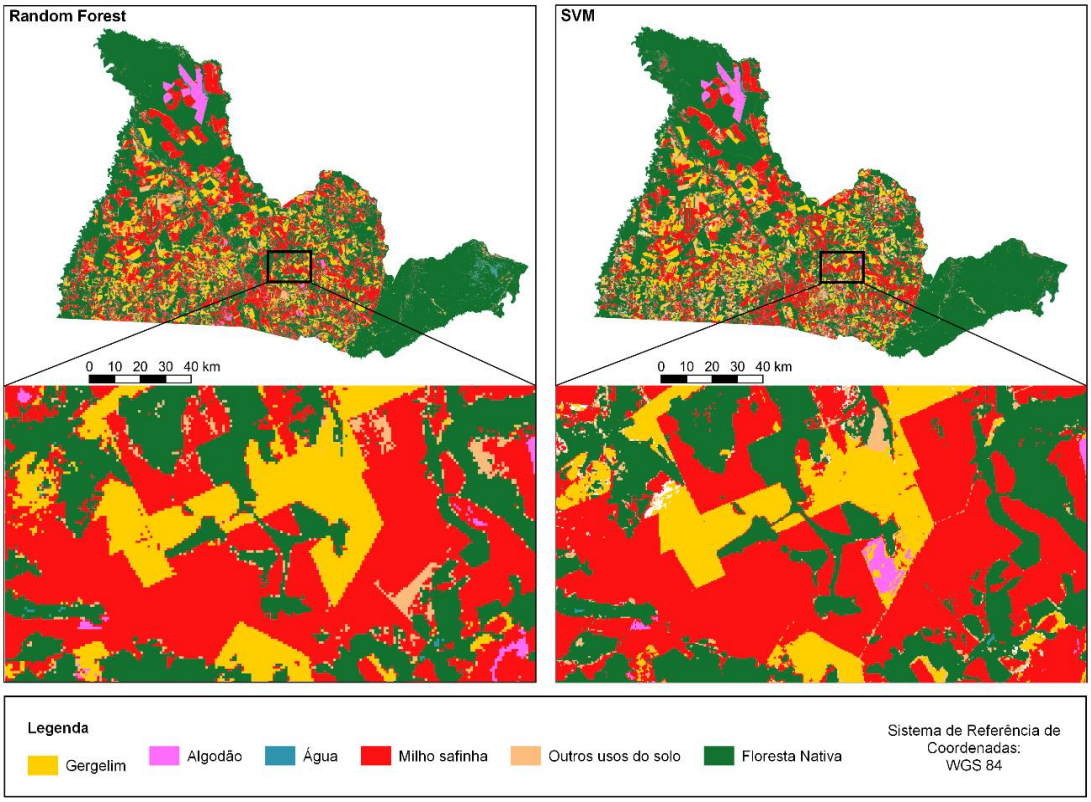


Figura 8. Resultado da classificação para o ano-safra 2021/2022 no município de Canarana-MT.

O efeito “sal e pimenta” observado nas imagens, conforme Resende et al., (2012), ocorreu pelo fato que o pixel não é maior ou igual aos objetos pesquisados e a área de interesse possui grande heterogeneidade. Assim, notou-se que o Random Forest detectou (Figura 9) melhor as variações dos pixels, sobretudo para a classe gergelim, ao passo que o SVM às considerou homogêneas.

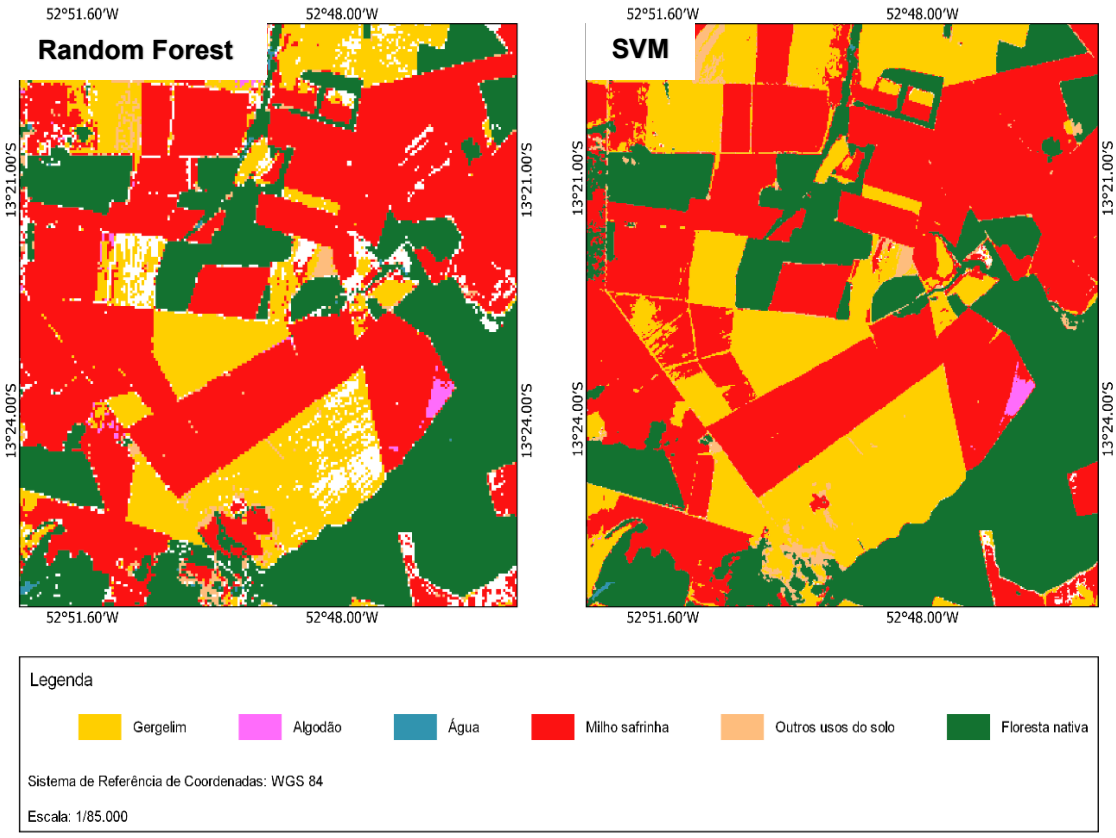


Figura 9. Resultado da classificação para os classificadores para o ano-safra 2021/2022 no município de Canarana-MT.

A Tabela 7 demonstra os resultados da validação dos mapas proveniente dos classificadores. Foi observado na matriz de confusão que ambos os classificadores na classe gergelim, confundiram pixels para as mesmas classes. Essa confusão ocorreu em virtude da proximidade dos valores dos pixels entre as classes. No entanto, essa confusão foi menor no Random Forest quando comparado ao SVM. A menor confusão entre os pixels no Random Forest ocorreu pela maneira que este classificador aborda a fase de treinamento, criando quantas

árvores de decisão necessárias forem, realizando testes no conjunto de dados supervisionados, o que não ocorre SVM. Este por sua vez, cria hiperplanos para separar os pontos de dados na classificação, que nesta situação demonstrou-se menos efetivo.

Tabela 7. Matrizes de confusão (número de pixels) das tabulações cruzadas, corretas e incorretas do conjunto de validação, para os classificadores.

Classificador		Gergelim	Algodão	Água	Milho 2ª safra	Outros usos do solo	Floresta nativa	Total
Random Forest	**Gergelim**	1047	1	0	22	63	0	**1133**
	Algodão	1	359	0	1	0	0	**361**
	Água	0	0	173	0	0	47	**220**
	Milho 2ª safra	27	4	0	2874	81	63	**3049**
	Outros usos do solo	96	0	2	99	911	296	**1404**
	Floresta nativa	0	0	59	12	103	13881	**14055**
	Total	**1171**	**364**	**234**	**3008**	**1158**	**14287**	**20222**
Support Vector Machine	**Gergelim**	981	15	0	48	89	0	**1133**
	Algodão	2	359	0	0	0	0	**361**
	Água	0	0	159	0	0	61	**220**
	Milho 2ª safra	135	0	0	2772	7	135	**3049**
	Outros usos do solo	156	0	0	209	282	757	**1404**
	Floresta nativa	0	0	70	65	7	13913	**14055**
	Total	**1274**	**374**	**229**	**3094**	**385**	**14866**	**20222**

Foi observado (Tabela 8) que o classificador Random Forest apresentou melhores resultados, com exatidão global de 95% e índice Kappa de 0,90, quando comparado com o SVM (91% e k=0,81). Ademais, apesar do Random Forest ter apresentado desempenho superior ao SVM, Sim e Wright, (2005) salientam que resultados de Exatidão Global e Kappa superiores a 85% e 0,80, respectivamente, performam satisfatoriamente. Dessa maneira, ambos os classificadores foram eficientes. Trabalhos de classificação de uso e ocupação do solo em paisagens complexas (LU et al., 2012; JAMAL-EDDINE et al., 2018; SURESH e LAL, 2020; ZAGAJEWSKI et al., 2021) obtiveram resultados semelhantes, 85% de Exatidão Global e Kappa de 0,80.

Tabela 8. Resultado da acurácia da classificação.

Classificador	**Acurácia da Classificação**	
	Exatidão Global (%)	Índice Kappa
Random Forest	95	0,90
Support Vector Machine	91	0,81

A utilização dos índices de vegetação para melhorar o desempenho da classificação (Figura 10), sobretudo em relação à cultura do gergelim, foi relevante, haja visto que foi observado na área de estudo que diversas lavouras estavam com dosséis mal formados e em muitas áreas o solo exposto era expressivo (Figura 11).

A adoção dos índices de vegetação neste trabalho vem de encontro com a proposição de Meneses (2019), que afirma que em dosséis não completamente fechados (mal formado), parte do fluxo de radiação incidente pode atravessá-los sem ser interceptada e atingir o solo. A refletância que retorna do solo se soma à do dossel, alterando os valores de reflectância da vegetação. A cor do solo é fundamental na refletância do dossel.

Além disso, cabe frisar que Ponzoni et al., (2012) esclarece que a baixa reflectância das folhas na região do visível é decorrente da absorção solar pela ação dos pigmentos fotossintetizantes, enquanto a alta reflectância na região do infravermelho próximo decorre do espalhamento (reflectância e transmitância) da radiação no interior das folhas em função de sua estrutura celular. Portanto, a reflexão

da radiação eletromagnética pelas folhas depende da sua composição química e estrutura interna.

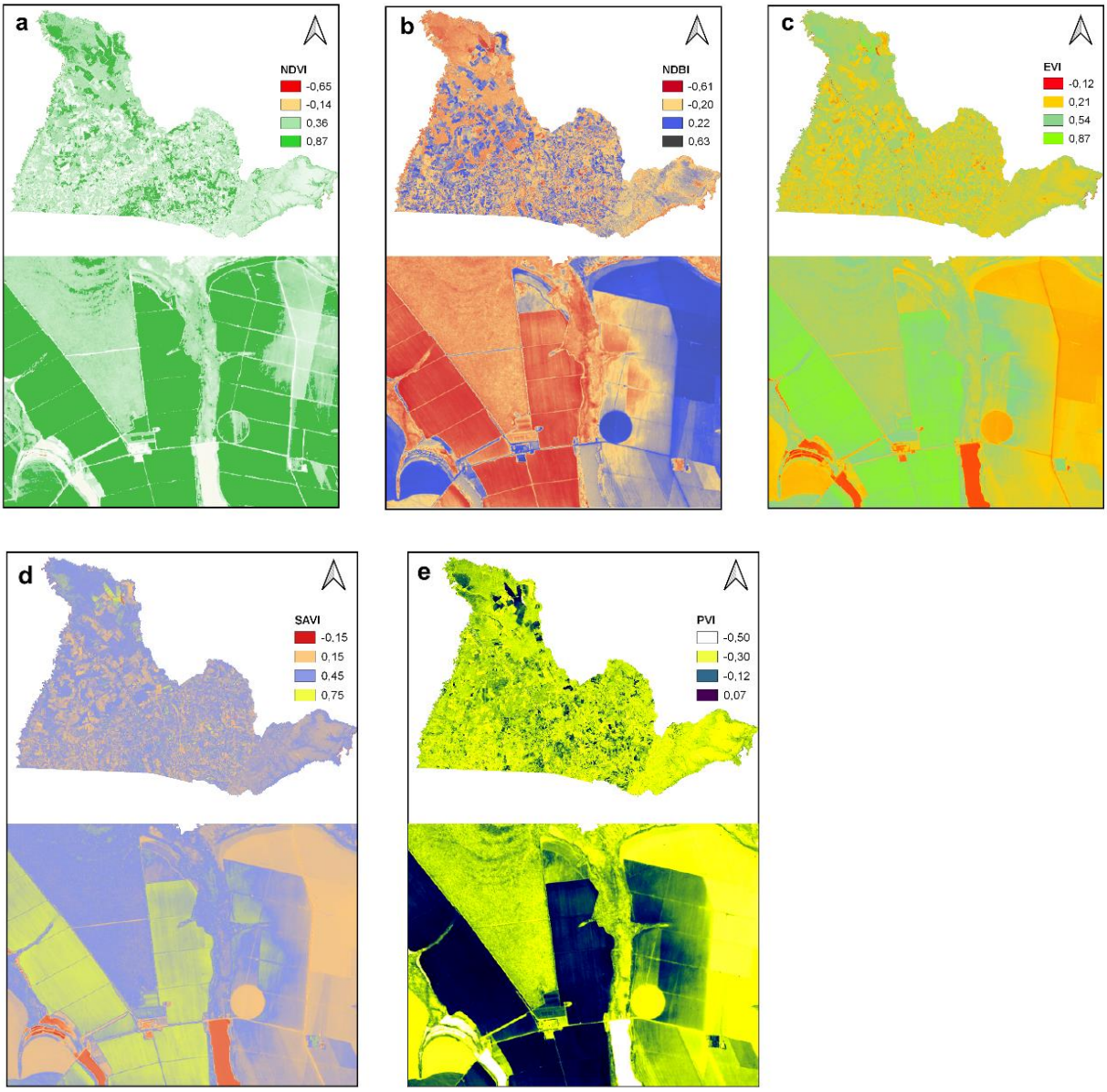


Figura 10. Índices de vegetação para área de estudo. a) NDVI; b) NDBI; c) EVI; d) SAVI; e) PVI.

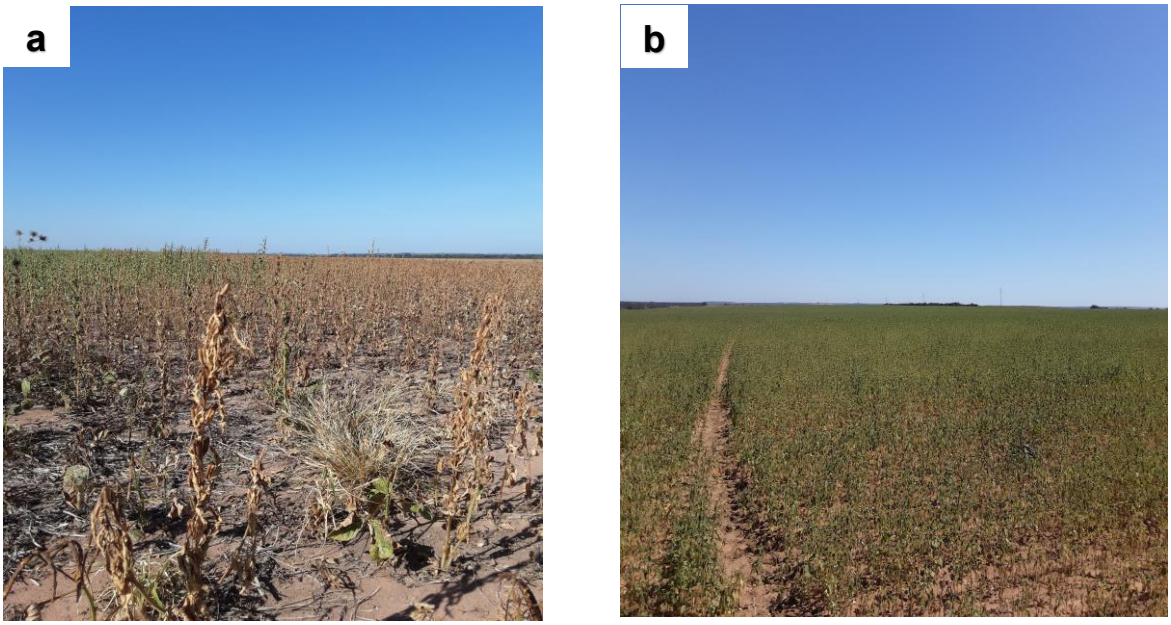


Figura 11. Cultura do gergelim com dosséis mal formados (a, b).

Tratando-se da validação do mapeamento realizado, Grossi et al., (2009) define que o erro de comissão consiste na inclusão de uma área dentro de uma classe à qual ela não pertence na realidade. Por outro lado, um erro de omissão consiste na exclusão de uma área da classe à qual ela pertence de fato. A precisão do usuário ou do consumidor, refere-se ao erro de omissão, enquanto a precisão do produtor refere-se ao erro de comissão. Partindo desta definição, observando a matriz de erros e as precisões do produtor, do usuário e índice Kappa para a mensuração da precisão das classes, notou-se que ocorreram erros consideráveis tanto no Random Forest quanto para o SVM (Tabela 9).

Tabela 9. Precisão do produtor, do usuário e índice Kappa.

Classificador	Precisão	Gergelim	Algodão	Água	Milho 2ª safra	Outros usos do solo	Floresta nativa
Random Forest	Precisão do Produtor	0,89	0,99	0,74	0,96	0,79	0,97
	Erro do Produtor	0,11	0,01	0,26	0,04	0,21	0,03
	Precisão do Consumidor	0,92	0,99	0,79	0,94	0,65	0,99
	Erro do Consumidor	0,08	0,01	0,21	0,06	0,35	0,01

	Coeficiente Kappa (k)	0,92	0,99	0,78	0,93	0,63	0,96
Support Vector Machine	Precisão do Produtor	0,77	0,96	0,69	0,90	0,73	0,94
	Erro do Produtor	0,23	0,04	0,31	0,10	0,27	0,06
	Precisão do Consumidor	0,87	0,99	0,72	0,91	0,20	0,99
	Erro do Consumidor	0,13	0,01	0,28	0,09	0,80	0,01
	Coeficiente Kappa (k)	0,86	0,99	0,72	0,89	0,19	0,96

Observou-se que para os classificadores, os erros mais contundentes estão nas classes outros usos do solo e água. Os erros de comissão observados para o Random Forest e SVM indicam que, 21% e 26%; 27% e 31%, respectivamente, das áreas reais destas classes foram classificadas como outras classes.

No entanto para o erro de omissão, os resultados denotaram que, 35% e 21%; 80% e 28%, respectivamente, das áreas que deveriam ter sido classificadas como outros usos do solo e água não foram, e com isso, apresentaram índice Kappa de 0,63 e 0,78; 0,19 e 0,72, respectivamente. Já os melhores resultados referentes aos erros de comissão, omissão e índice Kappa, foram apresentados nas classes algodão e floresta nativa, com 1% e 4%; 1% e 1% e; 0,99 e 0,96; 0,99 e 0,96, respectivamente ao Random Forest e SVM, notando-se que os classificadores apresentaram resultados semelhantes.

Tratando-se da classe gergelim, o Random Forest classificou 11% da área real como outras classes, e o SVM 23%; no entanto, 89% e 77%, respectivamente, foram classificados corretamente. Para o erro de omissão e Kappa, os resultados foram de 8% e 13% e, 0,92 e 0,86, respectivamente. Portanto, observa-se que o Random Forest apresentou melhores resultado quando comparado ao SVM para classificação da cultura do gergelim. A classe milho 2ª safra demonstrou resultados similares a classe gergelim, podendo-se observar o Kappa de 0,93 para o Random Forest e 0,89 para o SVM. No entanto, não se identifica tal proximidade ao que se refere ao erro de omissão, obtendo resultados de 6% e 9%; e para o erro de comissão, 4% e 10%, respectivamente.

De maneira semelhante, Waleed et al., (2020) comparou o desempenho de quatro diferentes algoritmos, dentre eles, o Random Forest e SVM na classificação de áreas cultivadas de oliveira (*Olea europaea*) na Espanha, e observou que o Random Forest apresentou acurácia global de 97,5%, erro de comissão de 4% e erro omissão de 1,5%; já o SVM apresentou acurácia global de 92,3%, erro de comissão de 19%, e erro e omissão de 0,3%, demonstrando que o desempenho daquele foi superior a este.

Assim, se observou que ambos os classificadores subestimaram a área das classes, no entanto, o SVM apresentou resultados inferiores quando comparado ao Random Forest. Contudo, Grossi et al., (2009) afirmam que esses erros são ocasionados pela ambiguidade espectral entre as classes de uso e cobertura da terra em ambientes de paisagem complexa. Para tanto, a avaliação do classificador que obteve melhor performance, em relação ao teste de hipóteses, pode ser visto na Tabela 10.

Tabela 10. Teste de hipótese comparando os resultados de exatidão entre os índices - ano safra 2021/2022.

Kappa¹ x Kappa²	Z	valor-p
Random Forest x Support Vector Machine	17,30	0,0000*

^{ns}Não significativo; *Significativo a 0,05 de probabilidade.

Analisando o resultado do teste Z, foi possível verificar que o algoritmo Random apresentou índice Kappa superior ao SVM. Contudo, haja visto que as amostras usadas foram as mesmas para ambos os algoritmos, esse resultado possivelmente decorreu de uma confusão nas áreas durante a fase de treinamento. Por fim, foi gerada a quantificação da estimativa das áreas classificadas (Tabela 11) e exposta na figura 12, a classificação realizada pelos algoritmos especificamente para a classe gergelim. Diante das informações, mesmo implementando as técnicas de tratamento dos dados nas fases de pré-processamento e treinamento, não foi possível atenuar todos os erros de classificação.

Tabela 11. Áreas estimadas para as classes de uso do solo no município de Canarana-MT.

Classificador	Gergelim	Algodão	Água	Milho 2ª safra	Outros usos do Solo	Floresta Nativa
Random Forest	95.395,45	15.592,12	5.861,74	203.649,04	35.413,03	601.213,20
Support Vector Machine	111.604,86	13.177,21	3.140,09	203.162,22	47.246,56	638.696,17

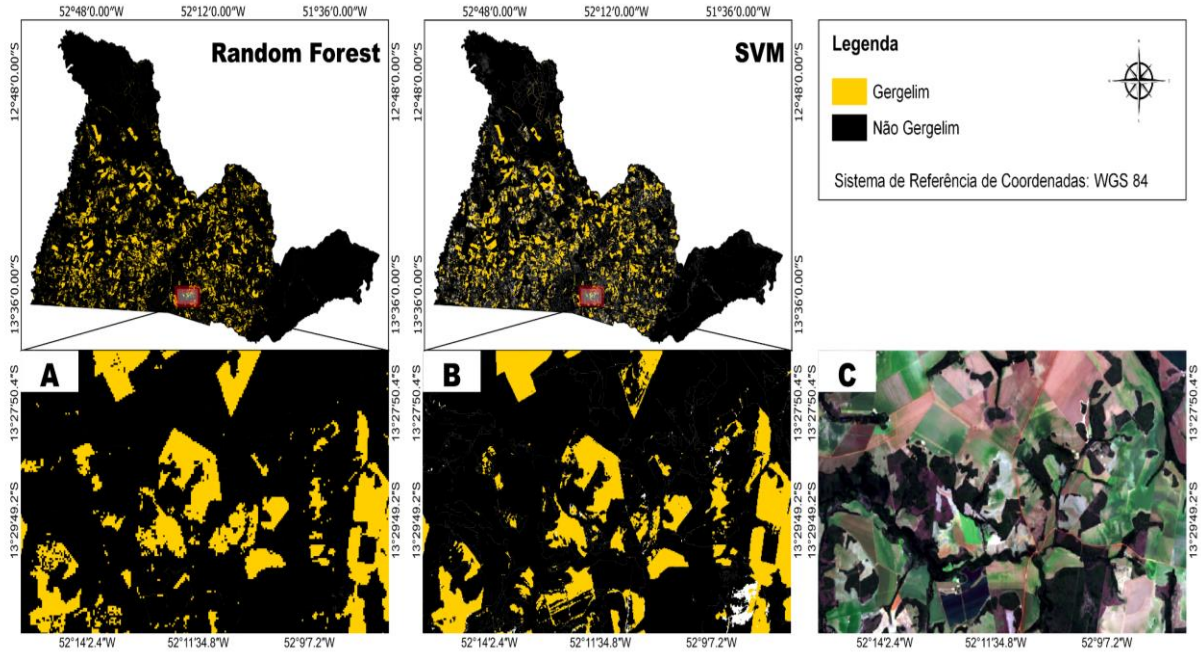


Figura 12. Classificação dos algoritmos para a classe gergelim: a) Random Forest; b) SVM; c) Composição falsa cor RGB.

Nota-se na figura 12 que, nas mesmas porções de áreas ou polígonos classificados, os algoritmos apresentaram resultados distintos. Esse resultado é fruto da maneira que os algoritmos trataram os dados de input na fase de treinamento, o que acabou ocasionando diferenças nas estimativas das áreas. O fator preponderante à essas diferenças é a complexidade da paisagem, que varia desde pastagens degradadas, solos expostos, floresta nativa até culturas agronômicas. É conhecido

que a vegetação apresenta ao longo do seu desenvolvimento vegetativo, diferentes comportamentos espectrais, o que a depender de fatores fisiológicos, podem assemelhar-se a outros tipos de vegetação, gerando confusão aos algoritmos na fase de treinamento e, consequentemente erros nas estimativas.

11. CONCLUSÃO

O algoritmo Random Forest mostrou-se superior na estimativa da área cultivada de gergelim na safra 2021/2022, no município de Canarana-MT. A maneira que este classificador estrutura o modelo de dados de treinamento, gerando diversas árvores de decisão, resultou em dados mais precisos quando comparado ao SVM. Considerando que ambos os classificadores utilizaram o mesmo conjunto de dados de treinamento e imagens, foi concluído que o SVM nestas condições se mostrou mais limitado. No entanto, ambos os algoritmos possuem potencial para produzir dados confiáveis.

A precisão dos mapas temáticos produzidos pelos classificadores revelou-se satisfatória. A utilização das imagens do satélite Landsat-8/OLI disponibilizadas gratuitamente, com resolução temporal e espacial adequadas para o monitoramento agrícola, possibilitou acompanhar remotamente o desenvolvimento vegetativo da cultura do gergelim, obtendo assim, informações em grande escala com precisão, tornando o processo de machine learning para obtenção de dados confiáveis menos dispendioso.

12. REFERÊNCIAS BIBLIOGRÁFICAS

Ahmadian, N., Ghasemi, S., Wigneron, J. P., & Zölitz, R. (2016). **Comprehensive study of the biophysical parameters of agricultural crops based on assessing Landsat 8 OLI and Landsat 7 ETM+ vegetation indices**. GIScience and Remote Sensing, 53(3), 337–359. https://doi.org/10.1080/15481603.2016.1155789.

Ajayi, S., Reddy, S. K., Gowda, P. H., Xue, Q., Rudd, J. C., Pradhan, G., Liu, S., Stewart, B. A., Biradar, C., & Jessup, K. E. (2016). **Spectral Reflectance Models for Characterizing Winter Wheat Genotypes**. Journal of Crop Improvement, 30(2), 176–195. https://doi.org/10.1080/15427528.2016.1138421.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. D. M., & Sparovek, G. (2013). **Köoppen climate classification map for Brazil**. Meteorologische Zeitschrift, 22(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Andreola, R., Haertel, V. (2009). **Support Vector Machines na Classificação de Imagens Hiperespectrais**. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto. Instituto de Pesquisas Espaciais - INPE, 1, 6765–6764.

Asghar, A., Majeed, M. N., Akhtar, M. N. (2014). **A review on the utilization of sesame as functional food**. American Journal of Food and Nutrition, April, 21–34. https://doi.org/10.5251/ajfn.2014.4.1.21.34.

Azevedo, O., Parker, T. C., Siewert, M. B., & Subke, J.-A. (2021). **Predicting Soil Respiration from Plant Productivity (NDVI) in a Sub-Arctic Tundra Ecosystem**. Remote Sensing, 13(13), 2571. https://doi.org/10.3390/rs13132571.

Bernardes, T., Adami, M., Formaggio, A. R., Moreira, M. A., França, D. de A., & de Novaes, M. R. (2011). **Imagens mono e multitemporais Modis para estimativa da área com soja no Estado de Mato Grosso**. Pesquisa Agropecuaria Brasileira, 46(11), 1530–1537. https://doi.org/10.1590/S0100-204X2011001100015.

Borel, Cristoph C.; Villeneuve, Pierre V.; Clodium, William B.; Szymenski, John J. & Davis, A. B. (1999). **Practical Atmosperic Correction Algorithms for a Multi-Spectral Sensor From the Visible Through the Termal Spectral Regions**. https://digital.library.unt.edu/ark:/67531/metadc702684/m1/1/.

Camps-Valls, G. (2009). **Machine learning in remote sensing data processing**. 2009 IEEE International Workshop on Machine Learning for Signal Processing, 1–6. https://doi.org/10.1109/MLSP.2009.5306233.

Chakouri, M., Lhissou, R., El Harti, A., Maimouni, S., & Adiri, Z. (2020). A**ssessment of the image-based atmospheric correction of multispectral satellite images for geological mapping in arid and semi-arid regions**. Remote Sensing Applications: Society and Environment, 20. https://doi.org/10.1016/j.rsase.2020.100420.

Chemura, A., & Mutanga, O. (2016). **Developing detailed age-specific thematic**

maps for coffee (Coffea arabica L .) in heterogeneous agricultural landscapes using random forests applied on Landsat 8 multispectral sensor. Geocarto International, 6049(June). https://doi.org/10.1080/10106049.2016.1178812.

Chen, J., & Zhu, W. (2020). **Consistency evaluation of landsat-7 and landsat-8 for improved monitoring of colored dissolved organic matter in complex water**. Geocarto International, 0(0), 1–12. https://doi.org/10.1080/10106049.2020.1734872.

CONAB. (2020a). **Acompanhamento da Safra Brasileira - SAFRA 2019/20** (Março 2020). Companhia Nacional de Abastecimento, 7(6), 1–89.

CONAB. (2020b). **Calendário de plantio e colheita de Grãos no Brasil**. https://blog.chbagro.com.br/calendario-de-plantio-e-colheita-para-cana-de-acucar-e-graos.

Cooley, T., Anderson, G. P., Felde, G. W., Hoke, M. L., Ratkowski, A. J., Chetwynd, J. H., Gardner, J. A., Adler-Golden, S. M., Matthew, M. W., Berk, A., Bernstein, L. S., Acharya, P. K., Miller, D., & Lewis, P. (2002). **FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation**. IEEE International Geoscience and Remote Sensing Symposium, 3, 1414–1418. https://doi.org/10.1109/IGARSS.2002.1026134.

Delhommelle, J. (2018). **Machine learning.** Molecular Simulation, 44(11), 865. https://doi.org/10.1080/08927022.2018.1471777.

Elleuch, M., Besbes, S., Roiseux, O., Blecker, C., & Attia, H. (2007). **Quality characteristics of sesame seeds and by-products**. Food Chemistry, 103, 641–650. https://doi.org/10.1016/j.foodchem.2006.09.008.

EMBRAPA. (2007a). **A Cultura do Gergelim**. Empresa Brasileira de Pesquisa Agropecuária. Embrapa Informação Tecnológica. https://www.embrapa.br/busca-de-publicacoes/-/publicacao/276375/a-cultura-do-gergelim.

EMBRAPA. (2007b). **Característica da Cultivar - BRS Seda**. https://www.embrapa.br/busca-de-solucoes-tecnologicas.

EMBRAPA. (2008). **Produção de Gergelim Orgânico nas Comunidades de Produtores Familiares de São Francisco de Assis do Piauí**. In Embrapa Algodão. https://www.infoteca.cnptia.embrapa.br/bitstream/doc/277831/1/DOC190.pdf.

EMBRAPA. (2009). **Característica da Cultivar - BRS Seda**. https://www.embrapa.br/busca-de-solucoes-tecnologicas.

EMBRAPA. (2015). **Gergelim - BRS Anahí**. 83. https://ainfo.cnptia.embrapa.br/digital/bitstream/item/132000/1/Folder-Gergelim-Nair-Curvas1.pdf.

EMBRAPA. (2021). **Clima - Classificação de Köoppen**. Empresa Brasileira de Pesquisas Agropecuárias. https://www.cnpf.embrapa.br/pesquisa/efb/clima.htm.

Felde, G. W., Anderson, G. P., Cooley, T. W., Matthew, M. W., Adler-Golden, S. M., Berk, A., & Lee, J. (2003). **Analysis of Hyperion Data with the FLAASH Atmospheric Correction Algorithm**. International Geoscience and Remote Sensing Symposium (IGARSS), 1(C), 90–92. https://doi.org/10.1109/igarss.2003.1293688.

Lobato Fernandes, J. (2009). **Monitoramento da cultura de cana-de-açucar no Estado de São Paulo por meio de imagens spot vegetation e dados meteorologicos**. Dissertação (Mestrado - Faculdade de Engenharia Agrícola). Universidade Estadual de Campinas. Campinas. SP. https://doi.org/10.47749/T/UNICAMP.2009.469139.

Fitz, P. R. (2008). **Geoprocessamento sem Complicação (1st ed.)**. Ofina de Textos.

Food and agriculture organization of the united nations. (2017). **Cenário por Demanda de Alimentos**. FAO No Brasil. http://www.fao.org/brasil/noticias/detail-events/en/c/901168/.

Formaggio, A. R., & Sanches, I. D. (2017). **Sensoriamento Remoto em Agricultura (1st ed.)**. Ofina de Textos. https://ebooks.ofitexto.com.br/product/sensoriamento-remoto-em-agricultura%0Ahttps://ebooks.ofitexto.com.br/epubreader/sensoriamento-remoto-em-agricultura.

Garofalo, D. F. T., Messias, C. G., Liesenberg, V., Bolfe, É. L., & Ferreira, M. C. (2015). **Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático**. Pesquisa Agropecuaria Brasileira, 50(7), 593–604. https://doi.org/10.1590/S0100-204X2015000700009.

Ge, G., Shi, Z., Zhu, Y., Yang, X., & Hao, Y. (2020). **Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms**. Global Ecology and Conservation, 22, e00971. https://doi.org/10.1016/j.gecco.2020.e00971.

Govind, A., Bhavanarayana, M., Kumari, J., & Govind, A. (2005). **Efficacy of different indices derived from spectral reflectance of wheat for nitrogen stress detection**. Journal of Plant Interactions, February 2015, 37–41. https://doi.org/10.1080/17429140500396958.

Grilo, J. A., Azevedo, P. V. (2013). **Crescimento, desenvolvimento e produtividade do gergelim brs seda irrigado com água do lençol freático, na agrovila de canudos, em ceará mirim (RN)**. Holos, 2, 19–33. https://doi.org/10.15628/holos.2013.1223.

Grossi, T., Vieira, C., Maria, H., Alves, R., Marin, M., Volpato, L., & Bernardes, T. (2009). VI Simpósio de Pesquisa dos Cafés do Brasil. **Avaliação de classificadores automáticos no mapeamento de áreas cafeeiras da região de guaxupé , minas gerais**.http://www.consorciopesquisacafe.com.br/index.php/component/content/article/40-fixo/547-anais-do-vi-simposio-de-pesquisa-dos-cafes-do-brasil.

Hernandez, Fernando Braz Tangerino.; Franco, Renato Alberto Momesso.; Teixeira, A.

H. de C. (2015). **Uso do Satélite Landsat 8 na Determinação da Produtividade da Água em Bacia Hidrográfica com Predomínio do Uso Agrícola da Cana-de-Açúcar**. Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR, João Pessoa-PB, Brasi, 1, 6381–6388.

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). **An assessment of support vector machines for land cover classification**. International Journal of Remote Sensing, 23(4), 725–749. http://www.tandfonline.com/loi/tres20%5Cnhttp://dx.doi.org/10.1080/01431160110040323%5Cnhttp://www.tandfonline.com/page/terms-and-conditions.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). **Overview of the radiometric and biophysical performance of the MODIS vegetation indices**. Remote Sensing of Environment, 83, 195–213. https://www.sciencedirect.com/science/article/abs/pii/S0034425702000962.

Huo, L., Persson, H. J., & Lindberg, E. (2021). E**arly detection of forest stress from European spruce bark beetle attack , and a new vegetation index : Normalized distance red & SWIR (NDRS)**. Remote Sensing of Environment, 255(December 2020), 112240. https://doi.org/10.1016/j.rse.2020.112240.

Industry Research. (n.d.). **Global sesame seeds market - segmented by geography - growth, trends, and forecast (2019 - 2024)**. https://www.industryresearch.co/global-sesame-seeds-market-14244158.

Instituto Brasileiro de Geografia e Estatística. (2020). **Bases Cartográficas Contínuas**. https://dados.gov.br/dataset?tags=Geoespacial&res_format=ZIP+SHP.

Jamal-eddine, O., Abderrazak, E. H., Rachid, L., Ali, E. M., Naima, B., Rabii, O., Mostafa, B. El, Ghmari, E., Ali, E. M., Naima, B., Rabii, E. O., & El, B. (2018). **Crop type mapping from pansharpened Landsat 8 NDVI data: A case of a highly fragmented and intensive agricultural system**. Remote Sensing Applications: Society and Environment. https://doi.org/10.1016/j.rsase.2018.05.002.

Jamali, A. (2019). **Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran**. SN Applied Sciences, 1(11). https://doi.org/10.1007/s42452-019-1527-8.

Jamali, A., & Rahman, A. A. (2019). **Sentinel-1 Image Classification for City Extraction Based on the Support Vector Machine and Random Forest Algorithms**. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(4/W16), 297–302. https://doi.org/10.5194/isprs-archives-XLII-4-W16-297-2019.

Jannuzzi, P. D. M. (2018). **A importância da informação estatística para as políticas sociais no Brasil : breve reflexão sobre a experiência do passado para considerar no presente**. 35(1), 1–10.

Jansen, L. J. M., & Di Gregorio, A. (2003). **Land-use data collection using the "land cover classification system": Results from a case study in Kenya. Land Use Policy**, 20(2), 131–148. https://doi.org/10.1016/S0264-8377(02)00081-9.

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). **Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong**. International Journal of Remote Sensing, 39(23), 8703–8723. https://doi.org/10.1080/01431161.2018.1490976.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M. R., Kuemmerle, T., Meyfroidt, P., Mitchard, E. T. A., Reiche, J., Ryan, C. M., & Waske, B. (2016). **A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring**. Remote Sensing, 8(1), 1–23. https://doi.org/10.3390/rs8010070.

Junior, C. A. da S. (2014). **Estimativa e discriminação de áreas de soja [Glycine max L .] no estado do Paraná com dados mono e multitemporais do sensor MODIS**. Dissertação (Mestrado - Centro de Ciências Agrárias). Universidade Estadual de Maringá. Maringá. PR. http://repositorio.uem.br:8080/jspui/handle/1/1235.

Knight, E. J., & Kvaran, G. (2014). **Landsat-8 operational land imager design, characterization and performance**. Remote Sensing, 6(11), 10286–10305. https://doi.org/10.3390/rs61110286.

KUMAR, D. N. (2018). **Basic Concepts of Remote Sensing**. Remote Sensing-Remote Sensing Systems. http://nptel.ac.in/courses/105108077/module2/lecture6.

Laurin, G. V., Liesenberg, V., Chen, Q., Guerriero, L., Del Frate, F., Bartolini, A., Coomes, D., Wilebore, B., Lindsell, J., & Valentini, R. (2012). **Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa**. International Journal of Applied Earth Observation and Geoinformation, 21(1), 7–16. https://doi.org/10.1016/j.jag.2012.08.002.

Leite, A. P., Santos, G. R., & Santos, J. É. O. (2017). **Análise temporal dos índices de vegetação NDVI e SAVI na Estação Experimental de Itatinga utilizando imagens Landsat 8**. Revista Brasileira de Energias Renováveis, 6(4). https://doi.org/10.5380/rber.v6i4.45830.

Liu, J., Maeda, E. E., Wang, D., & Heiskanen, J. (2021). **Sensitivity of Spectral Indices on Burned Area Detection using Landsat Time Series in Savannas of Southern Burkina Faso**. Remote Sensing, 13(13), 2492. https://doi.org/10.3390/rs13132492.

López-serrano, P. M., Corral-rivas, J. J., & Díaz-varela, R. A. (2016). **Evaluation of Radiometric and Atmospheric Correction Algorithms for Aboveground Forest Biomass Estimation Using Landsat 5 TM Data**. Remote Sensing, 1–19. https://doi.org/10.3390/rs8050369.

Lorenzzetti, J. A. (2015). **O espalhamento da radiação eletromagnética**. Princípios

Físicos de Sensoriamento Remoto (1st ed., pp. 133–152). Editora Edgar Blücher.

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). **Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines**. ISPRS Journal of Photogrammetry and Remote Sensing, 85, 102–119. https://doi.org/10.1016/j.isprsjprs.2013.08.007.

Lu, D., Batistella, M., Li, G., Moran, E., Hetrick, S., Freitas, C. da C., Dutra, L. V., & Sant'Anna, S. J. S. (2012). **Land use/cover classification in the brazilian amazon using satellite images**. Pesquisa Agropecuaria Brasileira, 47(9), 1185–1208. https://doi.org/10.1590/S0100-204X2012000900004.

Luka, R., Mario, M., & Damir, M. (2019). **Contemporary comparative assessment of atmospheric correction influence on radiometric indices between Sentinel-2A and Landsat 8 imagery**. Geocarto International, 0(0). https://doi.org/10.1080/10106049.2019.1590465.

Manabe, V. D. (2014). **Metodologia Para Mapeamento Da Expansão De Cana-De-Açúcar no Estado De Mato Grosso Por Meio De Séries Temporais Ndvi/Modis**. Dissertação (Mestrado - Faculdade de Engenharia Agrícola). Campinas. SP. https://bdtd.ibict.br/vufind/Record/UNICAMP-30_916f487f259d3a28a8f9479483c24468.

Mastella, A. F., & Vieira, C. A. (2018). **Acurácia temática para classificação de imagens utilizando abordagens por pixel e por objetos**. Revista Brasileira de Cartografia, 70(5), 1618–1643. https://doi.org/10.14393/rbcv70n5-44559.

Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). **Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest**. Sensors, 7(11), 2636–2651. https://doi.org/10.3390/s7112636.

Meneses, Paulo Roberto; Almeida, Tati de; Baptista, G. M. de M. (2019). **Reflectâncias dos Materiais Terrestres - Análise e Interpretação.** 1.ed. Oficina de Textos. São Paulo-SP.

Meneses, P. R., De Almeida, T., Roberto, P., Tati, M., Almeida, D., Nuno De Castro, A., & Rosa, S. (2012). **Introdução ao Processamento de Imagens de Sensoriamento Remoto**. Abril.

Ministério da agricultura, P. E. A. (2020). **O Brasil vai exportar gergelim para Índia**. https://www.gov.br/pt-br/noticias/agricultura-e-pecuaria/2020/01/brasil-vai-exportar-gergelim-para-india.

Moayedi, H., Jamali, A., Gibril, M. B. A., Kok Foong, L., & Bahiraei, M. (2020). **Evaluation of tree-base data mining algorithms in land used/land cover mapping in a semi-arid environment through Landsat 8 OLI image; Shiraz, Iran. Geomatics, Natural Hazards and Risk**, 11(1), 724–741.

https://doi.org/10.1080/19475705.2020.1745902.

Moraes, E. C. De. (2002). **Fundamentos De Sensoriamento Remoto**. In Inpe-8984-Pud/62. Instituto de Pesquisas Especias - INPE. http://mtc-m12.sid.inpe.br/col/sid.inpe.br/sergio/2005/06.14.12.18/doc/CAP1_ECMoraes.pdf.

Morato, G., Shinji, F., Marcos, K., Martines, R., Remoto, S., Universidade, H., Av, U. S. P., & Prestes, L. (2007). **Efeito do Padrão de Distribuição dos Pontos na Interpolação do NDBI**. Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, 2003, 2863–2869. http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.15.20.38.36/doc/2863-2869.pdf.

Morfitt, R., Barsi, J., Levy, R., Markham, B., Micijevic, E., Ong, L., Scaramuzza, P., & Vanderwerff, K. (2015). **Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit**. Remote Sensing, 7(2), 2208–2237. https://doi.org/10.3390/rs70202208.

Mountrakis, G., Im, J., & Ogole, C. (2011). **Support vector machines in remote sensing: A review**. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Nascimento, V. C. (2019). **Avaliação da Qualidade de Sementes de duas Cultivares de Gergelim em Diferentes Espaços**. https://repositorio.ufpb.br/jspui/handle/123456789/16596.

National Aeronautics and Space Administration (NASA). **Landsat**. Goddard Media Studios. Retrieved July 2, 2021, from https://svs.gsfc.nasa.gov/Gallery/Landsat.html.

National Aeronautics and Space Administration (NASA). **Operational Land Imager - OLI**. Retrieved June 11, 2021, from https://landsat.gsfc.nasa.gov/landsat-8/operational-land-imager.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). **Good practices for estimating area and assessing accuracy of land change**. Remote Sensing of Environment, 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.

Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). **Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation**. Remote Sensing of Environment, 129, 122–131. https://doi.org/10.1016/j.rse.2012.10.031.

Ortiz, M. J., Formaggio, A. R., & Epiphanio, J. C. N. (2010). **Classification of croplands through integration of remote sensing, GIS, and historical database**. International Journal of Remote Sensing, 18(1), 95–105. https://doi.org/10.1080/014311697219295.

Paiva, A. M.; Aragão, C. C. D. (2020). **Políticas Públicas: como fortalecer o**

agronegócio no estado do Rio de Janeiro? Revista Científica Multidiciplinar - Núcleo Do Conhecimento, 88–119. https://www.nucleodoconhecimento.com.br/administracao/fortalecer-o-agronegocio.

Pal, M. (2005). **Random Forest Classifier for Remote Rensing Classification**. International Journal of Remote Sensing, 26(1), 217–222. https://doi.org/10.1080/01431160412331269698.

Pal, M., & Mather, P. M. (2005). **Support vector machines for classification in remote sensing**. International Journal of Remote Sensing, 26(5), 1007–1011. https://doi.org/10.1080/01431160512331314083.

Ponzoni, F. J., shimabukuro, Y. E., & kuplich, T. M. (2012). **Sensoriamento Remoto da Vegetação (2ª ed)**. Ofina de Textos. São Paulo-SP.

Potgieter, A. B., Apan, A., Hammer, G., & Dunn, P. (2010). E**arly-season crop area estimates for winter crops in NE Australia using MODIS satellite imagery**. ISPRS Journal of Photogrammetry and Remote Sensing, 65(4), 380–387. https://doi.org/10.1016/j.isprsjprs.2010.04.004.

Poursanidis, D., Chrysoulakis, N., & Mitraka, Z. (2015). **Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urbanland cover mapping**. International Journal of Applied Earth Observation and Geoinformation, 35(PB), 259–269. https://doi.org/10.1016/j.jag.2014.09.010.

Rawat, K. S., Singh, S. K., & Ray, R. L. (2019). **An integrated approach to estimate surface soil moisture in agricultural lands**. Geocarto International, 0(0), 1–19. https://doi.org/10.1080/10106049.2019.1678674.

Rego, G M; lavaronti, O. J. (1995). **Atividades Fenológicas de Imbuia (Ocotea porosa (NEES ET MARTIUS ex NEES) em Áreas de Floresta Ombrófila Mista, no Estado do Paraná** (EMBRAPA); p. 1–18. Empresa Brasileira de Pesquisa Agropecuária. Embrapa Informação Tecnológica. https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/314095/1/AtividadesFenologicas.pdf.

Resende, M. R., Bernucci, L. L. B., & Quintanilha, J. A. (2012). **Classificação híbrida: pixel a pixel e baseada em objetos para o monitoramento da condição da superfície dos pavimentos rodoviários**. Boletim de Ciências Geodésicas, 18(3), 397–420. https://doi.org/10.1590/s1982-21702012000300004.

Rouse, J. W., J., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). **Monitoring Vegetation Systems in the great Plains With Erts**. NASA. Goddard Space Flight Center 3d ERTS-1 Symp., 309–317. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf.

Rudorff, B. F. T., de Aguiar, D. A., da Silva, W. F., Sugawara, L. M., Adami, M., & Moreira, M. A. (2010). **Studies on the rapid expansion of sugarcane for ethanol**

production in São Paulo state (Brazil) using Landsat data. Remote Sensing, 2(4), 1057–1076. https://doi.org/10.3390/rs2041057.

Ryu, J., Oh, D., & Cho, J. (2021). **Simple method for extracting the seasonal signals of photochemical reflectance index and normalized difference vegetation index measured using a spectral reflectance sensor**. Journal of Integrative Agriculture, 20(7), 1969–1986. https://doi.org/10.1016/S2095-3119(20)63410-4.

Sanches, I. D., Epiphanio, J. C. N., & Formaggio, A. R. (2005). **Culturas Agrícolas em Imagens Multitemporais do Satélite Landsat**. Agric. São Paulo, 52, 83–96. http://www.iea.sp.gov.br/out/publicacoes/pdf/asp-1-05-6.pdf.

Santos, H. G. J., Anjos, P. K. T., Oliveira, L. H. C., Lumbreras, V. A. de., Coelho, J. F., Almeida, M. R., Araújo Filho, J. A. de., Oliveira, J. C. de., (2018). **Sistema Brasileiro de Classificação de Solos**. In Embrapa Solos (5th, revista ed.). https://ainfo.cnptia.embrapa.br/digital/bitstream/item/199517/1/SiBCS-2018-ISBN-9788570358004.pdf.

Semolini, R. (2002). **Support vector machines, inferencia transdutiva e o problema de classificação**. Dissertação (Mestrado - Faculdade de Engenharia Elétrica e de Computação). Universidade Estadual de Campinas. Campinas. SP. https://doi.org/10.47749/T/UNICAMP.2002.272543.

Silva junior, C. A. ., Nanni, M. R. ., Teodoro, P. E., Silva, G. F. C. (2017). **Vegetation indices for discrimination of soybean areas: A new approach**. Agronomy Journal, 109(4), 1331–1343. https://doi.org/10.2134/agronj2017.01.0003.

Silva Junior, C. A. da, Frank, T., Rodrigues, T. C. S. (2014). **Discriminação de áreas de soja por meio de imagens EVI/MODIS e análise baseada em geo-objeto**. Revista Brasileira de Engenharia Agrícola e Ambiental, 18(1), 44–53. https://doi.org/10.1590/s1415-43662014000100007.

Sim, J., & Wright, C. C. (2005). **The kappa statistic in reliability studies: Use, interpretation, and sample size requirements**. Physical Therapy, 85(3), 257–268. https://doi.org/10.1093/ptj/85.3.257.

Sisodia, P. S., Tiwari, V., & Kumar, A. (2014). **A comparative analysis of remote sensing image classification techniques.** Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014, 1418–1421. https://doi.org/10.1109/ICACCI.2014.6968245.

Song, X., Duan, Z., & Jiang, X. (2012). **Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image**. International Journal of Remote Sensing, 33(10), 3301–3320. https://doi.org/10.1080/01431161.2011.568531.

Stehman, S. V, & Czaplewski, R. L. (1998). **Design and Analysis for Thematic Map Accuracy Assessment - an application of satellite imagery**. Remote Sensing of

Environment, 64(January), 331–344. http://www.ingentaconnect.com/content/els/00344257/1998/00000064/00000003/art00010%5Cnhttp://dx.doi.org/10.1016/S0034-4257(98)00010-8.

Suresh, S., & Lal, S. (2020). **A metaheuristic framework based automated Spatial-Spectral graph for land cover classification from multispectral and hyperspectral satellite images**. Infrared Physics and Technology, 105(December 2019), 103172. https://doi.org/10.1016/j.infrared.2019.103172.

United States Geological Survey - USGS. (2014, December 1). **Landsat-8 Missions**. USGS - Landsat-Missions/Landsat-8. https://www.usgs.gov/landsat-missions/landsat-8.

United States Geological Survey - USGS. (2015). **Landsat Enhanced Vegetation Index**. https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index.

United States Geological Survey - USGS. (2022). **EarthExplorer.** https://earthexplorer.usgs.gov/.

Vasconcelos, R. A., Gondim, T. M. de S., Arriel, N. H. de C. (2010). **Características Vegetativas e Reprodutivas e Genótipos de Gergelim no Cariri Cearense**. Congresso Brasileiro de Mamona, 4 & Simpósio Internacional de Oleaginosas Energéticas, 1, 2010, João Pessoa. Inclusão Social e Energia: Anais... Campina Grande: Embrapa Algodão, 2010. p. 1599-1605.

Verica, W. R. (2018). **Mapeamento semiautomático por meio de padrão espectro-temporal de áreas agrícolas e alvos permanentes com evi/modis no Paraná**. 2018. 116 f. Dissertação (Mestrado - Programa de Pós-Graduação em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná. Cascavel. PR. https://10.47749/T/UNICAMP.2002.272543.

Wahidin, N., Siregar, V. P., Nababan, B., Jaya, I., & Wouthuyzen, S. (2015). **Object-based Image Analysis for Coral Reef Benthic Habitat Mapping with Several Classification Algorithms**. Procedia Environmental Sciences, 24, 222–227. https://doi.org/10.1016/j.proenv.2015.03.029.

Waleed, M., Um, T. W., Khan, A., & Khan, U. (2020). **Automatic detection system of olive trees using improved K-means algorithm.** Remote Sensing, 12(5), 1–16. https://doi.org/10.3390/rs12050760.

Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., Wang, R., Sun, F., & Wu, X. (2018). **Evaluating the performance of Sentinel-2, Landsat 8 and Pléiades-1 in mapping mangrove extent and species**. Remote Sensing, 10(9). https://doi.org/10.3390/rs10091468.

Wang, L., Wang, J., & Qin, F. (2021). **Feature Fusion Approach for Temporal Land Use Mapping in Complex Agricultural Areas**. Remote Sensing, 13(13), 2517. https://doi.org/10.3390/rs13132517.

Wiegand, C. L., & Richardson, A. J. (1987). **Spectral components analysis rationale, and results for three crops**. International Journal of Remote Sensing, 8(7), 1011–1032. https://doi.org/10.1080/01431168708954744.

Wijewardana, C., Alsajri, F. A., Irby, J. T., Krutz, L. J., Golden, B., Henry, W. B., Gao, W., & Reddy, K. R. (2019). **Physiological assessment of water deficit in soybean using midday leaf water potential and spectral features**. Journal of Plant Interactions, 14(1), 533–543. https://doi.org/10.1080/17429145.2019.1662499.

Xiong, Y., Li, Y., Xiong, S., Wu, G., & Deng, O. (2021). **Multi-scale spatial correlation between vegetation index and terrain attributes in a small watershed of the upper Minjiang River**. Ecological Indicators, 126, 107610. https://doi.org/10.1016/j.ecolind.2021.107610.

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). **Comparison of Random Forest, Support Vector Machines, and Neural Networks for Post-Disaster Forest Species Mapping of the Krkonoše/Karkonosze Transboundary Biosphere Reserve**. Remote Sensing, 13(13), 2581. https://doi.org/10.3390/rs13132581.

Zha, Y., Gao, J., & Ni, S. (2003). **Use of normalized difference built-up index in automatically mapping urban areas from TM imagery**. International Journal of Remote Sensing, 24(November 2012), 583–594. https://doi.org/10.1080/01431160210144570.

Zhao, J., Zhong, Y., Hu, X., Wei, L., & Zhang, L. (2020). **A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions**. Remote Sensing of Environment, 239(December 2019), 111605. https://doi.org/10.1016/j.rse.2019.111605.

Zheng, Y., Tang, L., & Wang, H. (2021). **An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI**. Journal of Cleaner Production, 328, 129488. https://doi.org/10.1016/j.jclepro.2021.129488.

Zhong, L., Hu, L., Yu, L., Gong, P., & Biging, G. S. (2016). A**utomated mapping of soybean and corn using phenology**. ISPRS Journal of Photogrammetry and Remote Sensing, 119, 151–164. https://doi.org/10.1016/j.isprsjprs.2016.05.014.

Zhong, R., Wang, P., Mao, G., Chen, A., & Liu, J. (2021). **Spatiotemporal variation of enhanced vegetation index in the Amazon Basin and its response to climate change**. Physics and Chemistry of the Earth, Parts A/B/C, 123(1088), 103024. https://doi.org/10.1016/j.pce.2021.103024.

APÊNDICE A





Áreas de solo exposto (I, II).

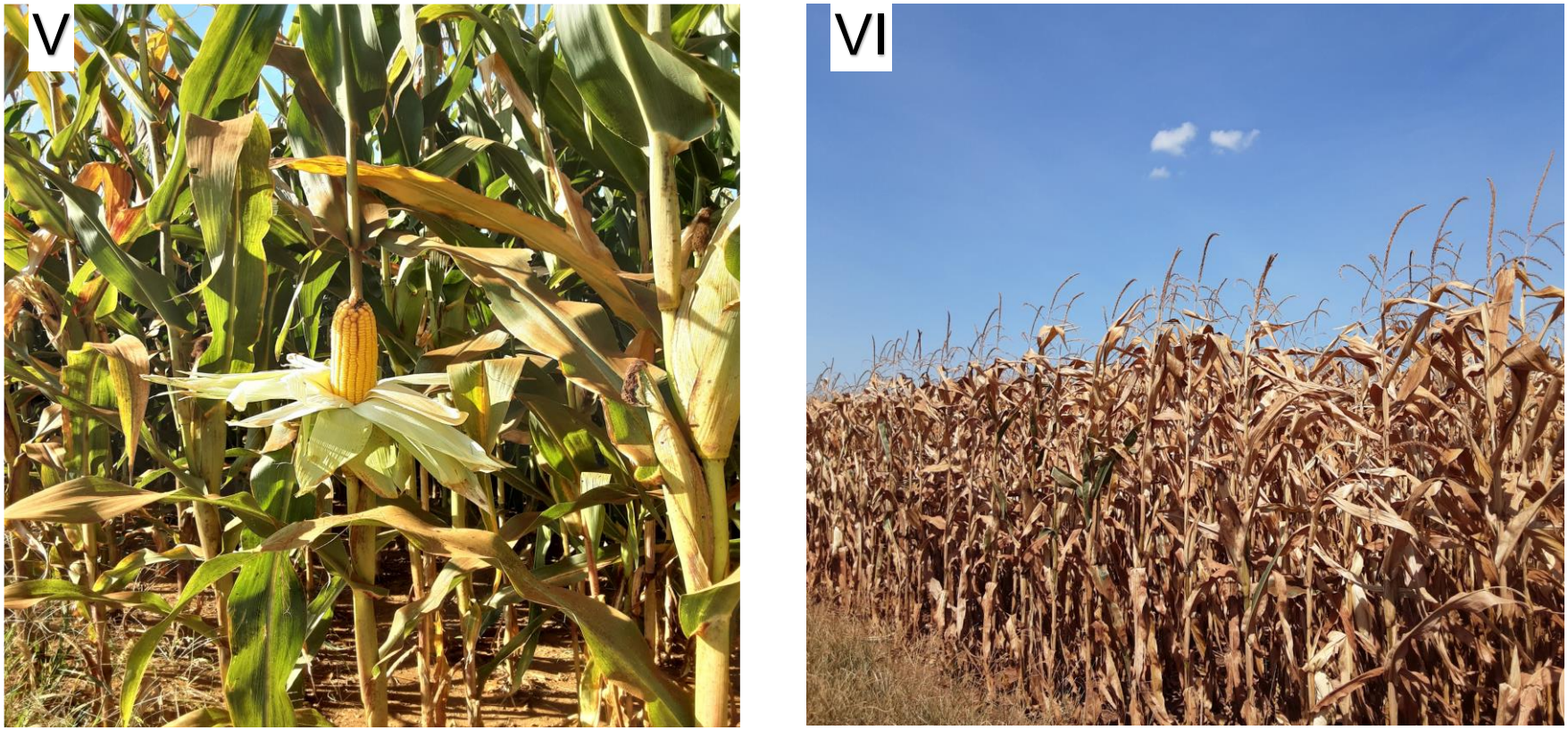
APÊNDICE B





Recursos hídricos (III, IV).

APÊNDICE C



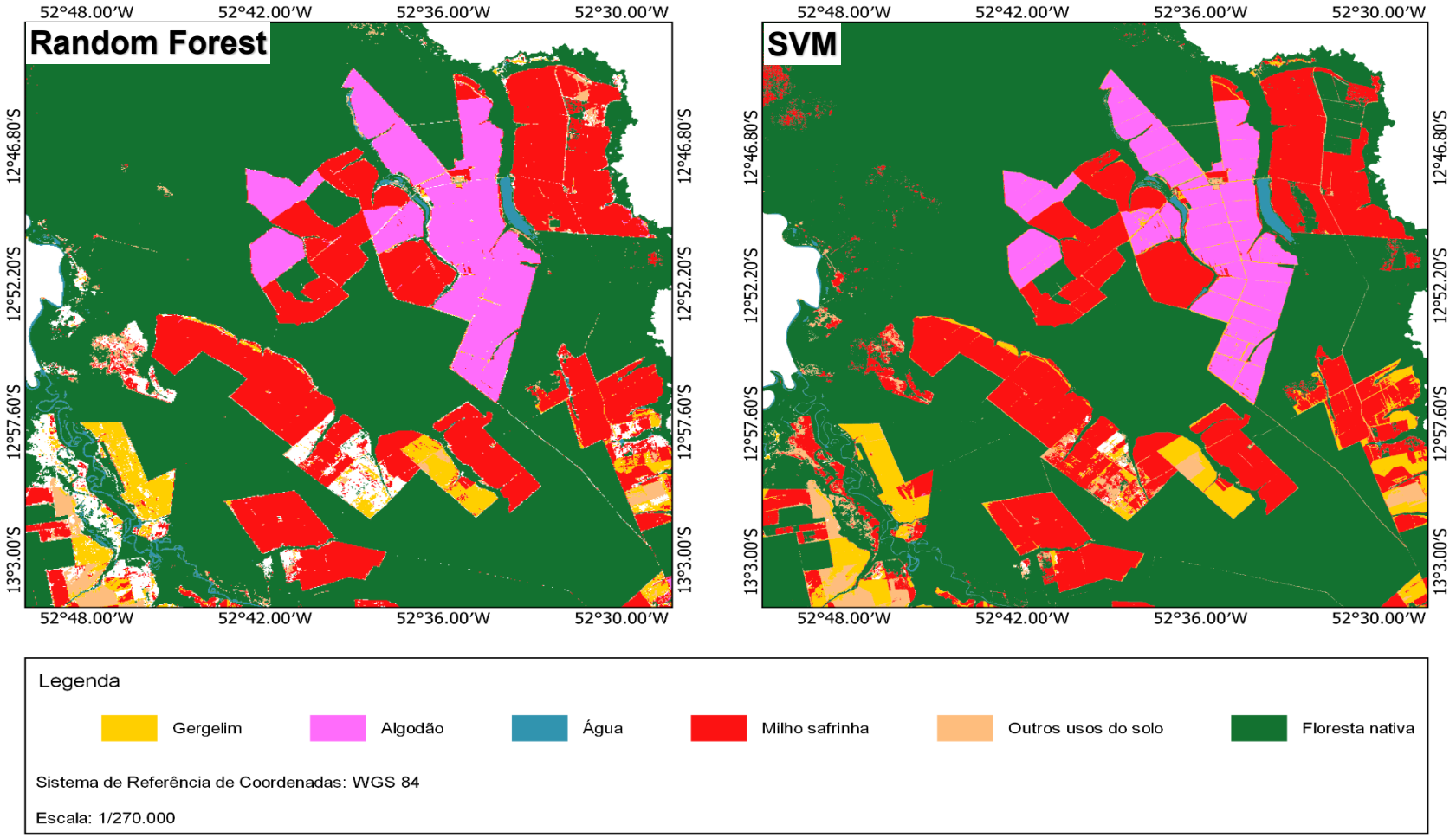
Cultura do milho 2ª safra: **V)** estádio de enchimento de grãos; **VI)** estádio de maturação.

APÊNDICE D



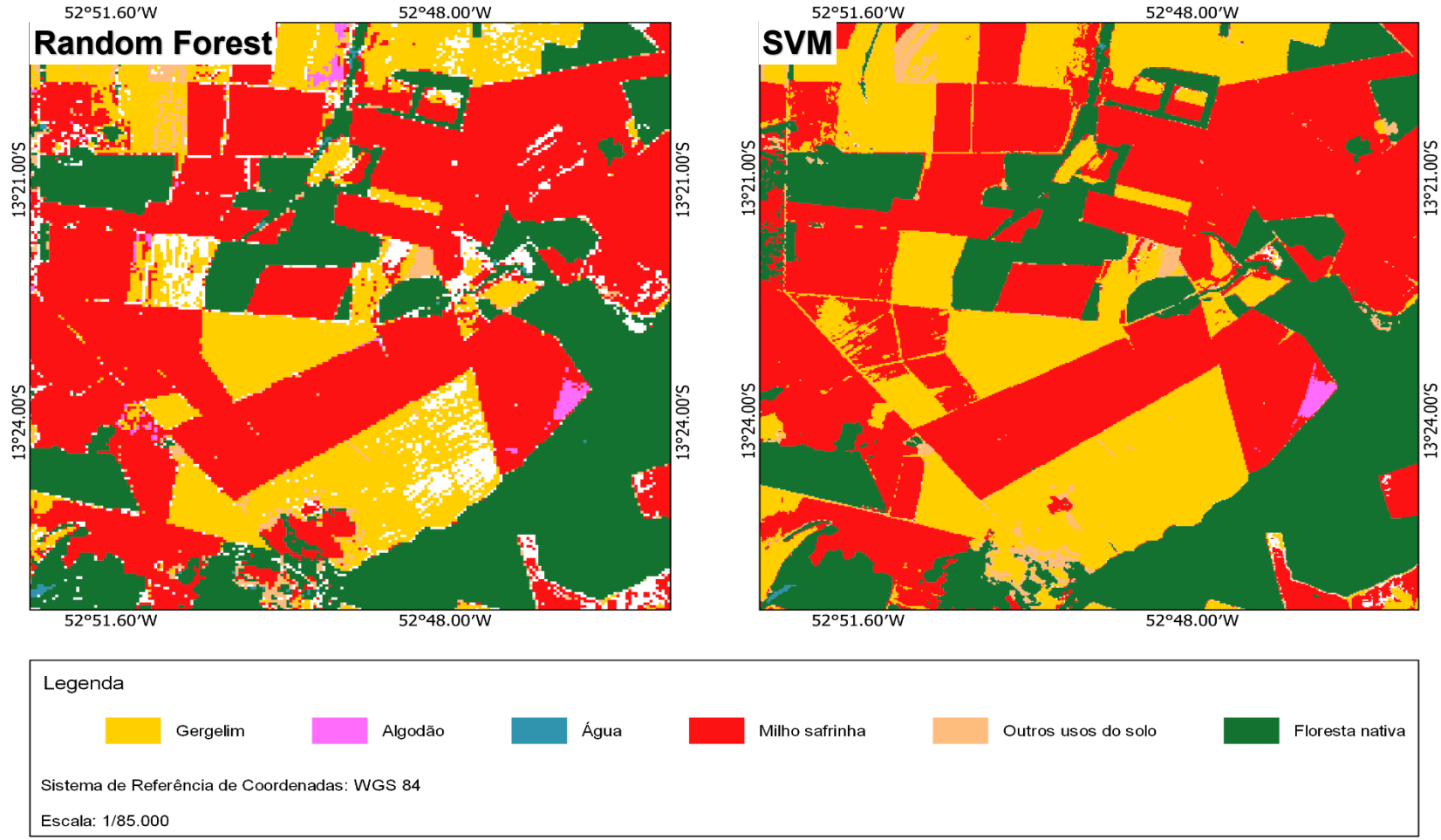
Cultura do gergelim: **VII)** estádio de florescimento; **VIII)** estádio de maturação.

APÊNDICE E



Resultado dos algoritmos de classificação: Random Forest; SVM.

APÊNDICE F



Resultado dos algoritmos de classificação: Random Forest; SVM.

APÊNDICE G

Erro médio apresentado pelos classificadores.

Classificador		Gergelim	Algodão	Água	Milho 2ª safra	Outros usos do solo	Floresta Nativa
Random Forest	**Gergelim**	0,0518	0,0000	0,0000	0,0011	0,0031	0,0000
	Algodão	0,0000	0,0178	0,0000	0,0000	0,0000	0,0000
	Água	0,0000	0,0000	0,0086	0,0000	0,0000	0,0023
	Milho 2ª safra	0,0013	0,0002	0,0000	0,1421	0,0040	0,0031
	Outros usos do solo	0,0047	0,0000	0,0001	0,0049	0,0450	0,0146
	Floresta nativa	0,0000	0,0000	0,0029	0,0006	0,0051	0,6864
SVM	**Gergelim**	0,0485	0,0007	0,0000	0,0024	0,0044	0,0000
	Algodão	0,0001	0,0178	0,0000	0,0000	0,0000	0,0000
	Água	0,0000	0,0000	0,0079	0,0000	0,0000	0,0030
	Milho 2ª safra	0,0067	0,0000	0,0000	0,1371	0,0003	0,0067
	Outros usos do solo	0,0077	0,0000	0,0000	0,0103	0,0139	0,0374
	Floresta nativa	0,0000	0,0000	0,0035	0,0032	0,0003	0,6880