

UNIVERSIDADE FEDERAL DE VIÇOSA

GABRIEL DUMBÁ MONTEIRO DE CASTRO

**MODELO PARA MONITORAMENTO REMOTO DA FERRUGEM DO CAFEEIRO
UTILIZANDO *MACHINE LEARNING***

VIÇOSA – MINAS GERAIS

2022

GABRIEL DUMBÁ MONTEIRO DE CASTRO

**MODELO PARA MONITORAMENTO REMOTO DA FERRUGEM DO CAFEIEIRO
UTILIZANDO *MACHINE LEARNING***

Relatório final, apresentado a Universidade Federal de Viçosa, como parte das exigências para a obtenção do título de Engenheiro Agrícola e Ambiental.

Orientador: Domingos Sárvio Magalhães Valente

VIÇOSA – MINAS GERAIS

2022

GABRIEL DUMBÁ MONTEIRO DE CASTRO

**MODELO PARA MONITORAMENTO REMOTO DA FERRUGEM DO CAFEIEIRO
UTILIZANDO *MACHINE LEARNING***

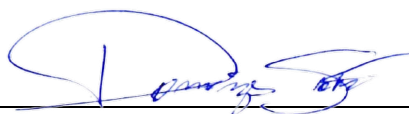
Relatório final, apresentado a Universidade Federal de Viçosa, como parte das exigências, para obtenção do título de Engenheiro Agrícola e Ambiental.

APROVADO: 16 de dezembro de 2022.

Assentimento:



Gabriel Dumbá Monteiro de Castro
Autor



Domingos Sarvio Magalhães Valente
Orientador

RESUMO

A ferrugem do café (*Hemileia vastatrix*) é considerada a principal doença desse cultivo no Brasil. O fungo patógeno encontra condições ambientais favoráveis na maioria das regiões cafeeiras do país, sendo uma ameaça constante para os produtores brasileiros. O monitoramento da ferrugem do cafeeiro é importante, pois possibilita identificar os possíveis focos de incidência da doença e auxilia no controle e manejo dos cafezais. Objetivou-se com esse trabalho o desenvolvimento de um modelo para monitoramento remoto da ferrugem do cafeeiro utilizando técnicas de *Machine Learning*. Foi utilizado como base de informações para concepção do modelo um banco de dados sobre temperatura do ar, precipitação mensal e incidência de ferrugem de quatro campos experimentais da Empresa de Pesquisa Agropecuária de Minas Gerais, localizados nos municípios de Três Pontas, São Sebastião do Paraíso, Machado e Patrocínio, no estado de Minas Gerais. Com intuito de acrescentar informações ao *dataset* do trabalho, foram obtidos dados sobre a refletância espectral da vegetação dos campos experimentais, por meio do cálculo de 18 índices vegetativos, utilizando imagens de satélites da coleção Sentinel-2. O processamento das imagens e cálculo dos índices foram realizados utilizando a plataforma de processamento em nuvem *Google Earth Engine*. O desenvolvimento do modelo foi realizado mediante a elaboração de um *script* em linguagem de programação *python* na plataforma da *Google Research – Colaboratory*. Com o *dataset* completo, foi reservado uma parcela dos dados para treinamento do modelo e outra para teste. O monitoramento foi então desenvolvido com base no algoritmo de *Machine Learning - Random Forest* que classificou as áreas como infestadas ou saudáveis. Os atributos adotados para a classificação foram ranqueados por importância na concepção do modelo. A temperatura mensal e o índice *Enhanced Vegetation Index* (EVI) foram os atributos que apresentaram maior importância. A performance do modelo foi avaliada por meio das métricas de desempenho: acurácia geral, precisão, *recall*, especificidade e *area under curve* (AUC), alcançando os valores de 77,3%; 38,5%; 71,4%; 78,4% e 0,749; respectivamente. Este estudo indica que o modelo desenvolvido possui robustez e capacidade generalizada para o monitoramento remoto da infestação da ferrugem do cafeeiro.

Palavras-chave: Ferrugem do café; *Hemileia vastatrix*; *Random Forest*.

ABSTRACT

Coffee Rust (*Hemileia vastatrix*) is considered the primary coffee disease in the world. The pathogenic fungus can find favorable environmental conditions in different countries, constantly threatening coffee producers. The monitoring of coffee rust is crucial, because identifies possible outbreaks of disease incidence and helps in the control and management of coffee plantations. The objective of this work was to develop a model for remote monitoring of coffee rust using Machine Learning techniques. A database of monthly temperature, monthly precipitation and rust incidence of four experimental fields of the Agricultural Research Company of Minas Gerais, located in the municipalities of Três Pontas, São Sebastião do Paraíso, Machado e Patrocínio, in the state of Minas Gerais, was used to develop the model. In order to add information to the work's dataset, data on the spectral reflectance of the vegetation in the experimental fields were obtained by calculating 18 vegetative indices, using satellite images from the Sentinel-2 collection. The Image processing and index calculation were performed using the cloud processing platform - Google Earth Engine. The development of the model was carried out through the elaboration of a script in python programming language on the platform of Google Research - Collaboratory. With the dataset complete, a portion of the data was reserved for training the model and another for testing. The monitoring was then developed based on the Machine Learning - Random Forest algorithm that classified the areas as infested or healthy. The 22 attributes adopted for the classification were ranked according to their importance in the model's conception. The monthly temperature and the Enhanced Vegetation Index (EVI) were the attributes that showed the greatest importance. The model's performance was evaluated using the following performance metrics: overall accuracy, precision, recall, specificity and area under curve (AUC), reaching values of 77.3%, 38.5%, 71.4%, 78.4%, and 0.749, respectively. This study indicates that the developed model has robustness and generalized capability for the remote monitoring of coffee rust infestation.

Keywords: Coffee rust; *Hemileia vastatrix*; *Random Forest*.

SUMÁRIO

1	INTRODUÇÃO	6
2	MATERIAL E MÉTODOS	9
2.1	Banco de Dados e Localização	9
2.2	Coleção de Imagens de Satélite	11
2.3	Índices Vegetativos	11
2.4	O Algoritmo <i>Random Forest</i>	13
2.5	Desenvolvimento do Modelo	15
2.6	Avaliação do Desempenho do Modelo	16
3	RESULTADOS E DISCUSSÃO	19
3.1	Dataset.....	19
3.2	Performance do Modelo.....	20
3.3	Importância dos Atributos.....	23
3.4	Métricas de Desempenho.....	25
3.5	Considerações Finais	27
4	CONCLUSÃO	29

1 INTRODUÇÃO

O plantio do café (*Coffea arabica* e *Coffea robusta*) é de grande relevância econômica, social e cultural no Brasil e no mundo. Segundo Carvalho et al. (2018), estima-se que 25 milhões de famílias, em mais de 50 nações em desenvolvimento produzem e comercializam o café. O atual levantamento da safra de 2022 no Brasil estima uma produção por volta de 50 milhões de sacas de café beneficiadas (CONAB, 2022). O cultivo do café continua sendo um importante gerador de renda para o país, contribuindo com mais de 756 bilhões de reais para o Produto Interno Bruto em 2021 (CNCAFE, 2022). Indiscutivelmente, o Brasil depende do cultivo eficiente do café para o avanço da economia do país.

O sucesso do cultivo de café no Brasil está atrelado a fatores naturais como principalmente o clima e o controle da infestação de pragas e doenças. Dentre as doenças que ocorrem no café, destaca-se a ferrugem do cafeeiro por representar uma ameaça significativa ao longo de toda safra, o fungo patogênico (*Hemileia vastatrix*) encontra condições ambientais favoráveis na maioria das regiões cafeeiras do país (POZZA; CARVALHO; CHALFOUN, 2010). O problema se torna ainda maior onde ainda há amplo uso de cultivares suscetíveis a doença. A infestação por ferrugem pode causar perdas de produtividade de até 50% se nenhuma medida de controle for empregada (KUSHALAPPA; ESKES 1989; ZAMBOLIM, 2016). Desse modo, a detecção prévia nas lavouras e o manejo adequado da doença são etapas fundamentais no cultivo do café para garantir boa produtividade e gerar maiores retornos econômicos.

A evolução da ferrugem do cafeeiro nas lavouras está associada à combinação de três fatores, o ambiente (condições climáticas), o hospedeiro (cafeeiro) e o patógeno, havendo interação entre eles (MORAES, 1983). Os fatores climáticos favoráveis à doença são temperatura na faixa de 20°C a 25°C e umidade em níveis adequados à germinação dos esporos, condição favorecida pelas chuvas frequentes, principalmente as finas, pelo orvalho noturno e por ambientes sombrios (Pereira et al., 2008). Assim, os níveis de inoculação do fungo no Brasil geralmente aumentam entre novembro e abril, período em que ocorrem maiores temperaturas e chuvas constantes. Devido a isso, observa-se maiores infecções nos meses seguintes, de junho a agosto.

Atualmente no estado de Minas Gerais, o controle da ferrugem do cafeeiro é realizado em área total com o uso de fungicidas protetores sistêmicos nas folhas ou com aplicação via solo. As operações de pulverização agroquímicas são baseadas no calendário, começando em novembro/dezembro e continuando até abril (EMBRAPA, 1999). É importante destacar que a

distribuição espacial e temporal da incidência da ferrugem do café é heterogênea na lavoura, podendo supor que a estratégia atual de controle da ferrugem em área total possa ser substituída pelo manejo localizado nos focos de incidência da doença (ALVES et al., 2009).

Com o avanço das geotecnologias e do sensoriamento remoto, o manejo de pragas e doenças tornou-se cada vez mais eficiente e preciso, possibilitando o direcionamento do controle apenas nos locais necessários (QUEIROZ et al., 2021). Para aplicação dessas técnicas, um dos principais recursos utilizados é a adoção de índices de vegetação em que, por meio da refletância espectral das culturas, medem características como vigor vegetativo, potencial produtivo e podem fornecer informações a respeito da infestação de pragas e doenças. (BARROS; FARIAS; MARINHO, 2020).

Resumidamente, índices de vegetação são operações matemáticas envolvendo duas ou mais bandas espectrais, permitindo intercomparações espaciais e temporais da atividade fotossintética e do estado da vegetação (HUETE et al., 2002). O princípio da utilização dos índices vegetativos consiste no fato de que as energias refletidas de bandas específicas no espectro eletromagnético estão diretamente relacionadas com a atividade fotossintética do dossel, bem como a suposição de que o uso de duas ou mais bandas espectrais pode minimizar substancialmente as principais fontes de ruído que afetam a resposta da vegetação (FERREIRA; FERREIRA; FERREIRA, 2008).

Nos últimos anos, o uso em conjunto de imagens de satélites, índices vegetativos e algoritmos de aprendizado de máquina (*Machine Learning*) vem sendo utilizados para detecção de pragas e doenças em diferentes culturas (ZHENG; HUANG; CUI, 2018; LIU et al., 2020; MARIN et al., 2021). O princípio básico de estudos dessa natureza é fornecer aos algoritmos *Machine Learning* informações a respeito da infestação da cultura por meio de um *dataset* (banco de dados) e assim, através do processo de aprendizado de máquina, treinar o modelo para que consiga prever uma infestação.

Dentre os algoritmos de *Machine Learning* adotados nos estudos de detecção de pragas e doenças, o classificador *Random Forest* é o mais amplamente utilizado (MISRA; LI; HE, 2019). O *Random Forest* é um algoritmo de aprendizagem proposto por Breiman (2001), que consiste em um conjunto de árvores de decisão e vetores aleatórios independentes e identicamente distribuídos. Por meio do *Random Forest*, Fletcher e Reddy (2016) desenvolveram um modelo para discriminação da planta daninha caruru (*Amaranthus viridis*) na cultura da soja. Chemura, Mutanga e Dube (2017) utilizaram o *Random Forest* e outros métodos de *Machine Learning* para discriminar os níveis de infecção da ferrugem do cafeeiro em escala hiperespectral. A adoção do *Random Forest* tem apresentado resultados

satisfatórios na literatura científica, desempenhando bem quando comparado com outros métodos de classificação (MISRA; LI; HE, 2019).

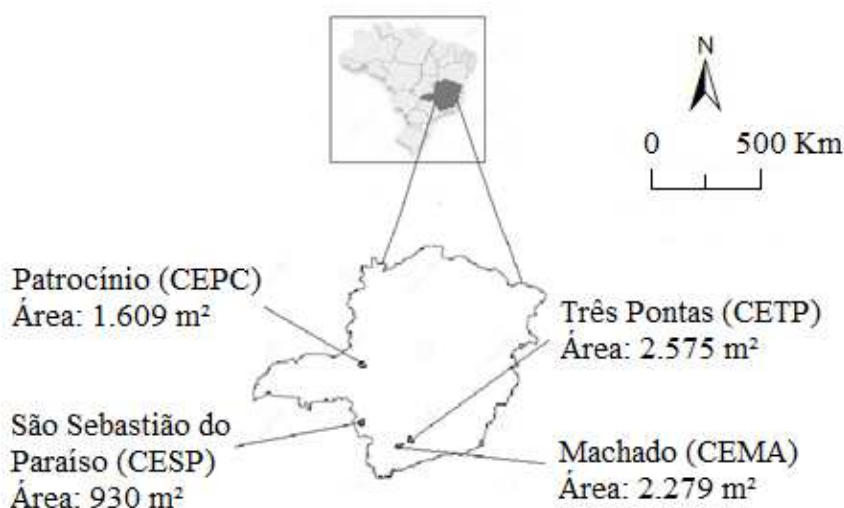
Partindo da hipótese de que é possível detectar alterações no comportamento espectral das plantas devido à incidência da ferrugem do cafeeiro por meio do sensoriamento remoto orbital. O objetivo com este trabalho foi desenvolver um modelo utilizando índices vegetativos e o algoritmo de *Machine Learning – Random Forest*, para monitoramento remoto da incidência da ferrugem do cafeeiro por imagens de satélites.

2 MATERIAL E MÉTODOS

2.1 Banco de Dados e Localização

Para servir como base de informações no desenvolvimento do modelo, foi utilizado um banco de dados sobre a incidência da ferrugem do cafeeiro de quatro campos experimentais da Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), localizados nos municípios de Machado (CEMA), São Sebastião do Paraíso (CESP), Patrocínio (CEPC) e Três Pontas (CETP) no estado de Minas Gerais (Figura 1). O acesso ao banco de dados foi cedido voluntariamente para desenvolvimento do trabalho e consta no repositório da Empresa de Pesquisa Agropecuária de Minas Gerais (SILVA, 2021).

Figura 1 – Localidade e área dos campos experimentais onde foram coletados os dados



Fonte: O autor.

O banco de dados contém informações sobre a porcentagem de infestação da ferrugem do cafeeiro em cada centro experimental, registradas a campo ao longo dos anos de 2019, 2020 e 2021. Foi também obtido os dados climatológicos de temperatura média e precipitação das regiões. Todos os dados foram registrados mês a mês ao longo dos anos, com indicação da data em que foi feita a vistoria a campo.

Com intuito de simplificar o desenvolvimento do modelo resumiu-se os dados de infestação de ferrugem em duas classes: infestado, quando o registro de infestação no campo experimental foi maior ou igual a 40%; e saudável, quando o registro de infestação no campo experimental foi menor que 40%. O valor de fronteira de 40% foi determinado segundo o manual de recomendações da cultura de café no Brasil, que apontam que porcentagens de infestação de ferrugem superiores a 40% já indicam significativas perdas produtivas devido a eventual desfolha causada pela doença (MATIELLO et al., 2010). Com essa simplificação todas as análises posteriores tornam-se mais objetivas e a proposta de monitoramento da ferrugem no modelo se torna uma classificação binária das áreas por meio do apontamento de áreas infestadas ($\geq 40\%$ de infestação) ou saudáveis ($< 40\%$ de infestação).

O tipo de cultivar, as características edafoclimáticas e as coordenadas geográficas dos campos experimentais onde foram coletados os dados estão apresentados na Tabela 1. Os dados de precipitação foram obtidos a partir da base *Climate Hazards Group InfraRed Precipitation with Stations* (CHIRPS) (FUNK et al., 2015) e a temperatura a partir da base *Latest Climate Reanalysis Produced by ECMWF / Copernicus Climate Change Service - ERA 5*.

Tabela 1 – Características edafoclimáticas, tipo de cultivar, e coordenadas geográficas dos campos experimentais

Característica	CEMA	CESP	CEPC	CETP
Cultivar	Cauaí 99	Catuaí 99	Rubi	Mundo Novo
Elevação (m)	970	880	997	916
P. média anual (mm)	816	857	845	720
T. média anual (°C)	21.5	22	21.7	21.4
Coordenadas geográficas	45° 28' 59.452"O 21° 20' 37.014"S	47° 7' 20.341"O 20° 54' 42.023"S	46° 59' 21.700"O 18° 59' 28.284"S	45° 28' 59.452"O 21° 20' 37.014"S

CEMA: Centro Experimental de Machado, CESP: Centro Experimental de São Sebastião do Paraíso, CEPC: Centro Experimental de Patrocínio, CETP: Centro Experimental de Três Pontas. P: Precipitação, T: Temperatura do ar. O: Oeste, S: Sul.

Fonte: O autor.

Para elaboração do estudo, levou-se em consideração apenas a parcela central dos campos experimentais, o objetivo desse procedimento foi reduzir os efeitos de bordadura dos talhões que poderiam eventualmente provocar alterações na refletância espectral. Cada parcela estudada compreendeu então as seguintes áreas: CEMA (2.279 m²); CESP (930 m²); CEPC (1.609 m²) e CETP (2.575 m²).

2.2 Coleção de Imagens de Satélite

Mediante análise das coleções de imagens de satélites disponíveis, foi definido no estudo pela utilização da coleção Sentinel-2. A escolha se deu devido a ampla resolução espectral da coleção e apreciável resolução temporal de cinco dias, parâmetros que fornecem melhor aplicabilidade ao modelo desenvolvido. O sensor *Multi-Spectral Instrument* (MSI) incorporado aos satélites dessa coleção obtém informações de refletância espectral da superfície terrestre registrando-as em 13 diferentes bandas do espectro eletromagnético, na Tabela 2 é possível observar a denominações das bandas, o valor de comprimento de onda central, a largura das faixas espectrais e suas respectivas resoluções espaciais (ESA, 2019). A abundância de bandas captadas pela coleção Sentinel-2 possibilita o cálculo dos principais índices vegetativos utilizados em estudos de vegetação.

Tabela 2 - Bandas espectrais e resoluções do Sentinel-2 sensor MSI

Bandas do Sentinel-2	Comprimento de onda central (nm)	Largura de faixa (nm)	Resolução espacial (m)
B1	<i>Coastal aerosol</i>	443	60
B2	<i>Blue</i>	490	10
B3	<i>Green</i>	560	10
B4	<i>Red</i>	665	10
B5	<i>Red-edge 1 (Re1)</i>	705	20
B6	<i>Red-edge 2 (Re2)</i>	740	20
B7	<i>Red-edge 3 (Re3)</i>	783	20
B8	<i>Near infrared (NIR)</i>	842	10
B8A	<i>Red-edge 4 (Re4)</i>	865	20
B9	<i>Water vapor</i>	945	60
B10	<i>Shortwave infrared/cirrus</i>	1375	60
B11	<i>Shortwave infrared 1 (SWIR1)</i>	1610	20
B12	<i>Shortwave infrared 2 (SWIR2)</i>	2190	20

Fonte: ESA (2019).

É importante salientar que, considerando a resolução espacial da coleção Sentinel-2 de 10 metros para as bandas B2, B3, B4 e B8; e 20 metros para as bandas B5, B6, B7 e B8A, as áreas delimitadas dos campos experimentais foram suficientes para atender as análises propostas no trabalho.

2.3 Índices Vegetativos

No intuito de gerar mais dados para desenvolvimento do modelo de detecção remota da ferrugem do cafeeiro, foram utilizados 18 índices vegetativos comumente adotados em estudos de vegetação. Por meio das coordenadas geográficas dos campos experimentais (Tabela 1), foram obtidas imagens de satélites das áreas de estudo referentes aos meses dos anos de 2019 a 2021. Desse modo, com a informação da refletância espectral disponível nas imagens, foi calculado o valor médio de cada um dos 18 índices, em cada um dos campos experimentais. A obtenção das imagens de satélites e cálculo dos índices vegetativos foram realizados por meio da plataforma de processamento em nuvem *Google Earth Engine* (GEE). Os nomes, siglas, fórmulas e referências dos índices vegetativos utilizados no estudo estão apresentados na Tabela 3.

Tabela 3 – Índices vegetativos utilizados no desenvolvimento do modelo

Definição	Descrição da fórmula	Referência
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - RED}{NIR + RED}$	Rouse et al. (1974)
Renormalized Normalized Difference Vegetation Index (RNDVI)	$\frac{NIR - RED}{\sqrt{NIR + RED}}$	Gitelson e Merzlyak (1994)
Normalized Difference Red Edge 1 Index (NDVIRE1)	$\frac{NIR - RE1}{NIR + RE1}$	Gitelson e Merzlyak (1994)
Normalized Difference Red Edge 2 Index (NDVIRE2)	$\frac{NIR - RE2}{NIR + RE2}$	Gitelson e Merzlyak (1994)
Normalized Difference Red Edge 3 Index (NDVIRE3)	$\frac{NIR - RE3}{NIR + RE3}$	Gitelson e Merzlyak (1994)
Green normalized difference vegetation index (GNDVI)	$\frac{NIR - GREEN}{NIR + GREEN}$	Gitelson e Merzlyak (1996)
Green Chlorophyll Index (GCI)	$\frac{NIR}{GREEN} - 1$	Gitelson et al. (2005)
Simplified Canopy Chlorophyll Content Index 1 (SCCCI1)	$\frac{NDVI \cdot RE1}{NDVI}$	Barnes et al. (2000)
Simplified Canopy Chlorophyll Content Index 2 (SCCCI2)	$\frac{NDVI \cdot RE2}{NDVI}$	Barnes et al. (2000)
Simplified Canopy Chlorophyll Content Index 3 (SCCCI3)	$\frac{NDVI \cdot RE3}{NDVI}$	Barnes et al. (2000)
Red Edge 1 Chlorophyll Index (CLRE1)	$\frac{NIR}{RE1} - 1$	Gitelson et al. (2005)
Red Edge 2 Chlorophyll Index (CLRE2)	$\frac{NIR}{RE2} - 1$	Gitelson et al., (2005)

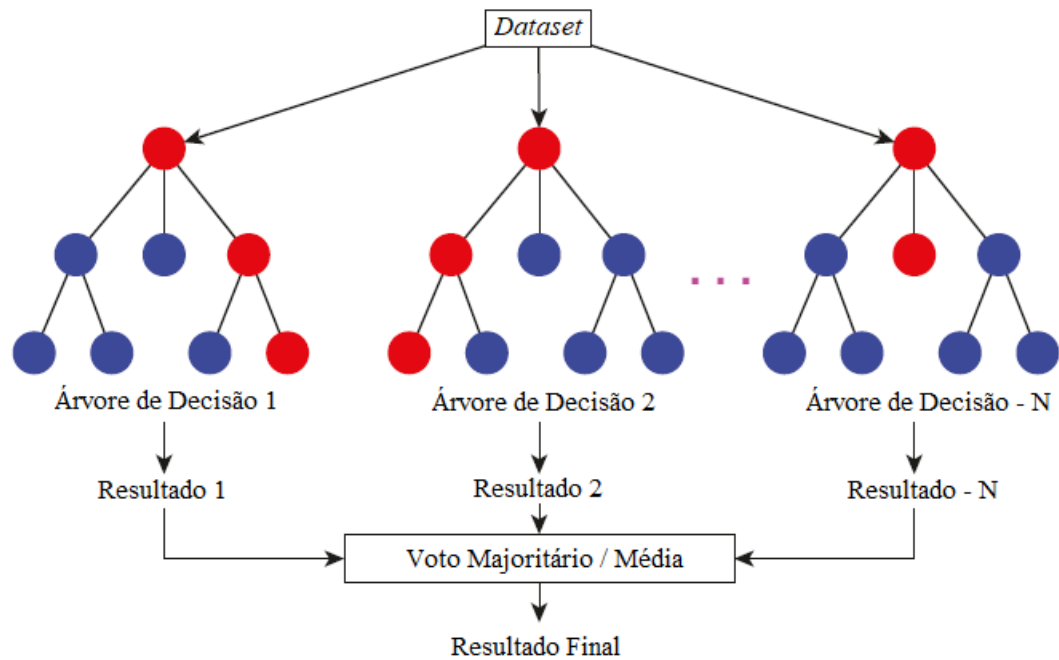
Red Edge 3 Chlorophyll Index (CLRE3)	$\frac{NIR}{RE3} - 1$	Gitelson et al. (2005)
Normalized red-edge1 index (NREDI1)	$\frac{RE2 - RE1}{RE2 + RE1}$	Gitelson e Merzlyak (1994)
Normalized red-edge2 index (NREDI2)	$\frac{RE3 - RE1}{RE3 + RE1}$	Gitelson e Merzlyak (1994)
Normalized red-edge3 index (NREDI3)	$\frac{RE3 - RE2}{RE3 + RE2}$	Gitelson e Merzlyak (1994)
Enhanced Vegetation Index (EVI)	$\frac{2,5 (NIR - RED)}{NIR + 6 RED - 0,5 BLUE + 1}$	Justice et al. (1998)
Simple ratio (SR)	$\frac{NIR}{RED}$	Jordan (1969)

Fonte: O autor.

2.4 O Algoritmo *Random Forest*

O modelo desenvolvido no estudo teve como base a utilização do algoritmo de *Machine Learning – Random Forest*. O algoritmo consiste em um conjunto de árvores de decisão e vetores aleatórios independentes e identicamente distribuídos. O *Random Forest* assume um conjunto de treinamento original com “N” instâncias e cada instância com “M” atributos (GENUER; POGGI; TULEAU-MALOT, 2015). Durante o processo de construção das árvores de decisão o algoritmo atua em dois aspectos: (1) amostrando um novo conjunto de treinamento com substituição a cada iteração, do mesmo tamanho do conjunto original; (2) e ao invés de escolher a melhor divisão entre todos os atributos, “m” atributos são escolhidos aleatoriamente de “M” em cada nó, assim esses “m” atributos são usados para dividir o nó de acordo com o princípio do algoritmo da árvore de decisão, em que “m” << “M”, no qual é mantido constante durante o processo de construção da floresta de decisão. Na Figura 2 é apresentada uma ilustração simplificada do processo de classificação do algoritmo.

Figura 2 – Resumo do processo de classificação do *Random Forest*



Fonte: Adaptada de Khan et al. (2021).

Como o algoritmo não usa todas as amostras para treinamento do modelo de uma só vez, isso torna possível a utilização das amostras restantes (*out of bag data*) para avaliar o erro envolvido em cada uma das amostras (*out of bag error - OOB*). Com essas informações o algoritmo consegue realizar a classificação de importância dos atributos, possibilitando a geração de um *ranking*. O princípio da elaboração desse *ranking* é a comparação da diferença do OOB de cada atributo antes e depois da adição de um ruído, determinando a importância de cada recurso utilizado para a classificação. Sendo assim, cada característica tem sua relevância diretamente proporcional a essa diferença (GENUER; POGGI; TULEAU-MALOT, 2015).

Diante disso, no presente trabalho foi proposto o desenvolvimento de um modelo com análise preditiva da infestação de ferrugem do cafeeiro utilizando o *Random Forest*. A análise preditiva nesse contexto objetiva realizar a estimativa do risco de eventos futuros reincidirem baseada em experiências já vividas, de modo a auxiliar na tomada de decisão atual através da implementação de um algoritmo de aprendizado de máquina usado para a compreensão de dados (KUHN et al., 2013).

2.5 Desenvolvimento do Modelo

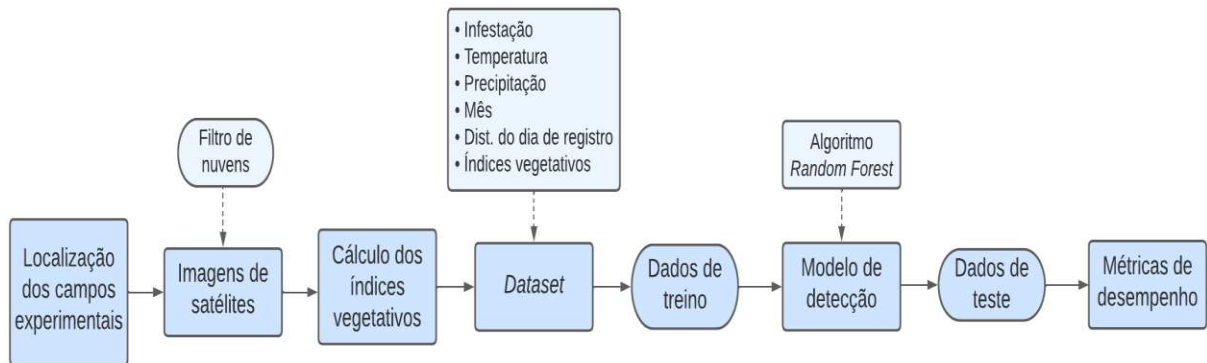
O desenvolvimento do modelo para monitoramento remoto da infestação da ferrugem do cafeeiro se deu primeiramente pela obtenção das imagens de satélites de cada campo experimental, na data mais próxima do registro da infestação a campo, em todos os meses no período de 2019 até 2021. A apuração das imagens disponíveis foi feita com a utilização de um filtro de nuvens permitindo porcentagem de pixel nublado de apenas 10%. Posteriormente, a seleção da imagem com melhor visualização do campo experimental foi feita através da observação de todas as imagens disponíveis. Meses nos quais não foi possível selecionar uma imagem de satélite foram excluídos do estudo.

Após a obtenção das imagens de satélites, foi calculado a média de cada um dos 18 índices vegetativos (Tabela 3), em cada um dos campos experimentais. Gerando assim as informações de refletância espectral da vegetação das áreas de estudo. Esses dados foram adicionados ao banco de dados do estudo, que continha informações a respeito da infestação de ferrugem, precipitação mensal e temperatura mensal de cada campo experimental. A junção de todos os dados formou o *dataset* do modelo.

Outro importante atributo adicionado ao *dataset* foi a diferença entre a data das imagens de satélites selecionadas e a data do registro de infestação feito a campo. A causa dessa diferença foi devida a incidência de nuvens e sombras em alguns casos que impossibilitou a seleção da imagem de satélite exatamente na data em que foi feito o registro da infestação. Casos em que ocorreram essa situação, selecionou-se a imagem de satélite imediatamente mais próxima da data pretendida. A inserção dessa informação no *dataset* teve como objetivo avaliar a influência desse parâmetro e fornecer mais um conjunto de dados para análise no modelo. Apesar do exposto, vale ressaltar que, normalmente a ferrugem no cafeeiro não evolui rapidamente na escala de dias (MORAES, 1983).

Com o *dataset* estabelecido, repartiu-se uma parcela dos dados para treinamento do modelo e outra para teste, os dados referentes aos anos de 2019 e 2020 foram selecionados para treino e os dados referentes ao ano de 2021 para teste. Essa escolha foi feita com objetivo de exemplificar uma aplicação do modelo, na qual se têm o registro de infestação de ferrugem de dois anos passados e deseja-se realizar o monitoramento no ano atual. Na Figura 3 está apresentado o fluxograma com todas as etapas realizadas no desenvolvimento do modelo.

Figura 3 – Fluxograma de desenvolvimento do modelo de monitoramento remoto da ferrugem do cafeeiro



Fonte: Autor.

As etapas de geoprocessamento e o cálculo dos índices vegetativos foram realizadas na plataforma de processamento em nuvem *Google Earth Engine*. A elaboração do modelo de monitoramento da ferrugem do cafeeiro se deu mediante o desenvolvimento de um *script* em linguagem de programação *python* na plataforma da *Google Research – Colaboratory*, utilizando as bibliotecas de processamento de dados *numpy* e *pandas*, e a biblioteca de *Machine Learning* de código aberto *scikit-learn*.

2.6 Avaliação do Desempenho do Modelo

A performance do modelo de monitoramento da ferrugem do cafeeiro foi avaliada com base nas principais métricas de desempenho utilizadas em algoritmos de *Machine Learning*, sendo elas: matriz de confusão, acurácia global, precisão, especificidade, *recall* e *area under curve* (AUC). A obtenção de cada um desses parâmetros ocorreu com base na predição realizada na etapa de teste do modelo.

A Matriz de Confusão é calculada pela função de pesquisa da classificação, ela exibe a distribuição dos registros em termos de suas classes reais e de suas classes previstas. Isso indica a qualidade do modelo atual (COUTO, 2003). A diagonal principal da matriz de confusão representa a concordância entre os dados de classificação e os dados de referência. Portanto, a classificação correta é representada pelas amostras da diagonal principal da matriz de confusão e a classificação incorreta pelas amostras fora da diagonal principal (COUTO, 2003; STEHMAN, 2019). Na Figura 4 está apresentado um exemplo de matriz de confusão para duas classes “Positivo” e “Negativo”, a classificação resulta no enquadramento dos resultados em quatro grupos: Verdadeiro Negativo (VN), dados no qual foi predito e observado a classe Negativo; Verdadeiro Positivo (VP), no qual foi predito e observado a

classe Positivo; Falso Negativo (FN), dados no qual foi predito classe Negativo porém foi observado Positivo; e Falso Positivo (FP), no qual foi predito Positivo porém foi observado Negativo.

Figura 4 – Exemplo de Matriz de Confusão para duas classes

		Predito	
		Negativo	Positivo
Observado	Negativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Fonte: Autor.

A acurácia total ou acurácia global apresentada por Story e Congalton (1986), tem como objetivo analisar a concordância geral entre os dados de classificação e os de referência através da soma da diagonal principal dividida pelo total de amostras, conforme apresentado na Equação 1.

$$Acurácia\ Global = \frac{VN + VP}{VN + FP + FN + VP} \quad (1)$$

A precisão avalia dentre todas as classificações da classe positivo que o modelo fez, quantas estão corretas (SHALEV-SHWARTZ; BEN-DAVID, 2014). Tal métrica pode ser calculada conforme apresentado na Equação 2. Conforme é possível interpretar pela fórmula, a precisão é fortemente influenciada pelos casos de falso positivos.

$$Precisão = \frac{VP}{VP + FP} \quad (2)$$

Por outro lado, temos a métrica especificidade que avalia a capacidade do modelo de detectar resultados negativos, ou seja, dentro todos os casos classificados como negativos, quantos estão corretos (SHALEV-SHWARTZ; BEN-DAVID, 2014). Tal métrica pode ser calculada conforme apresentado na Equação 3.

$$Especificidade = \frac{VN}{VN + FP} \quad (3)$$

O parâmetro *recall* avalia dentre todas as situações de classe positivo observadas, quantas o modelo está predizendo corretamente, ou seja, a capacidade do método de detectar com sucesso resultados classificados como positivos (SHALEV-SHWARTZ; BENDAVID, 2014). Tal métrica pode ser calculada conforme apresentado na Equação 4.

$$Recall = \frac{VP}{VP + FN} \quad (4)$$

A última métrica de desempenho analisada foi a *area under curve* (AUC), esse parâmetro é baseado na *receiver operating characteristics* (ROC) que é uma curva de probabilidade do modelo traçada a partir da taxa verdadeiro-positivo (*recall*), contra a taxa de falsos-positivos (especificidade), (PRATES, 2020). O AUC é adotado como medida para simplificar a curva ROC, resumindo as informações em um único valor que representa a “área sob a curva”. A AUC representa então o grau ou medida de separabilidade da curva (FERRI; HERNÁNDEZ-ORALLO; FLACH, 2011). Tem-se para essa métrica as seguintes relações, quanto mais próximo de 1 está o AUC, melhor o modelo está prevendo as classes propostas. Quando próximo a 0 significa que o modelo não tem capacidade de separação de classe. Quando próximo a 0,5 significa que o modelo está semelhante a uma classificação aleatória sem parâmetros definidos (PRATES, 2020).

O cálculo de todas as métricas de desempenho foi realizado no próprio *script* de desenvolvimento do modelo utilizando funções da biblioteca de *Machine Learning* de código aberto *scikit-learn*.

3 RESULTADOS E DISCUSSÃO

3.1 Dataset

No total 121 imagens de satélites da coleção Sentinel-2 foram selecionadas para o desenvolvimento do modelo. O resumo da classificação dos dados em saudável ou infestado, com base nos dados de infestação de ferrugem registrado nas áreas de estudo, está apresentado para os três anos na Tabela 4.

Tabela 4 – Resumo da imagens de satélites obtidas

Ano	2019	2020	2021	Total
Saudável	19	19	37	75
Infestado	19	20	7	46
Total	38	39	44	121

Fonte: autor.

É possível observar que o ano de 2021 apresentou maior número de imagens disponíveis, com 44 registros, seguido de 2020 e 2019. O padrão dos dados de infestação por ferrugem do cafeeiro foi similar em 2019 e 2020, apresentando quantidade de pontos considerados saudáveis e infestados semelhantes. Já no ano de 2021 o registro de pontos saudáveis foi muito superior que o de infestados. No geral, considera-se a distribuição dos dados totais favorável para a concepção do modelo, apresentando um pouco mais de 1/3 dos dados totais enquadrados na classe infestado, ou também chamado de “positivo”.

O cálculo dos 18 índices vegetativos gerou 2178 dados de refletância das áreas de estudo. O banco de dados com as informações a respeito da infestação de ferrugem, precipitação mensal, temperatura mensal e distância do dia de registro contribuiriam com 605 dados. No total 2783 dados foram utilizados para desenvolvimento do modelo. Na tabela 5 está apresentado o resumo do *dataset* por ano e origem dos dados.

Tabela 5 – Resumo do *dataset* utilizado para desenvolvimento do modelo

Ano	2019	2020	2021	Total
Índices Vegetativos	684	702	792	2178
Banco de dados	190	195	220	605
Total	874	897	1012	2783

Fonte: autor.

Conforme determinado, para a etapa de treinamento do modelo foram utilizados os dados referentes aos anos de 2019 e 2020, totalizando 1771 dados de treino. Para a etapa de teste foram utilizados os dados referentes ao ano de 2021 totalizando 1012 dados de teste. Novamente a distribuição de um pouco mais de 1/3 dos dados para teste e um pouco menos que 2/3 para treinamento é coerente com as proporções utilizadas em trabalhos dessa natureza (MEIRA; RODRIGUES; MORAES; PEREIRA; EGIDIO, 2009; THAMADA; GIROLAMO NETO; MEIRA 2013).

3.2 Performance do Modelo

Por meio do *script* de desenvolvimento do modelo, realizou-se a etapa de treinamento do algoritmo. Posteriormente, utilizando os dados disponíveis para teste, efetuou-se a predição da infestação de ferrugem em 44 pontos. Com o resultado do valor predito e o dado da infestação real registrada em campo, analisou-se a performance do modelo. O resumo dessa análise está apresentado na Figura 5, no qual é indicado para os 44 pontos testados a classificação predita e a classificação real.

Figura 5 – Performance do modelo na predição da infestação da ferrugem do cafeeiro nos dados de teste

		Centros Experimentais			
		CEMA	CEPC	CESP	CETP
2021	Jan				—
	Fev	—			
	Mar				
	Abr				
	Mai				
	Jun				
	Jul				
	Ago				
	Set				
	Out				—
	Nov				
	Dez				—

Legenda			
Real		Saudável	
Predito		Infestado	
		Sem imagem disponível	—

Fonte: Autor.

Utilizando da informação obtida dessa análise, foi gerada a Matriz de Confusão do modelo, apresentada na Figura 6. A diagonal principal da matriz, que representa os acertos, está destacada em cinza escuro e a secundária, que representa os erros, em cinza claro.

Figura 6 – Matriz de Confusão do modelo desenvolvido

Classes		Predito	
		Saudável	Infestado
Real	Saudável	29	8
	Infestado	2	5

Fonte: Autor.

Conforme é possível observar, em geral o desempenho do modelo foi razoável, predizendo corretamente em 34 dos 44 pontos testados. É importante realçar que na maioria dos casos infestados o modelo conseguiu prever corretamente, errando em apenas 2 dos 7 casos registrados em campo.

A maior quantidade de erros ocorreu em casos falso positivos, em que o modelo apontou a região como infestada, porém, ela se enquadrava na classe saudável. Entende-se no presente trabalho que esses casos são menos problemáticos em comparação com os falsos negativos, uma vez que, o apontamento de infestação em áreas saudáveis ainda pode indicar um possível risco eminente de contágio da doença. Casos como estes ainda vem a servir como um alerta ao usuário, havendo a necessidade de avaliação em campo para verificação do apontamento.

Considera-se que o apontamento mais problemático é o de casos falso negativos, em que o modelo aponta uma região como saudável, porém, ela se enquadra na classe infestada. É observado no Quadro 1 que essa situação ocorreu em apenas duas vezes das sete possíveis, no registro do mês de agosto de 2021 em CEMA e CESP. Tal situação pode ser justificada por diferentes fatores, a hipótese principal é de que o modelo possa ter dificuldade em detectar a ferrugem com precisão em determinados períodos do ano, visto que os dois erros ocorreram no mesmo mês. Verifica-se também que a maior quantidade de erros na predição do modelo aconteceu nos meses de maio a agosto, indicando assim a necessidade de maior atenção nos apontamentos realizados nesse período.

Além disso, outros fatores que não foram considerados no modelo podem também interferir na classificação, como a presença de outro tipo de pragas ou doenças no talhão; manejos culturais como podas, limpeza das linhas e adubação; entre outras etapas da

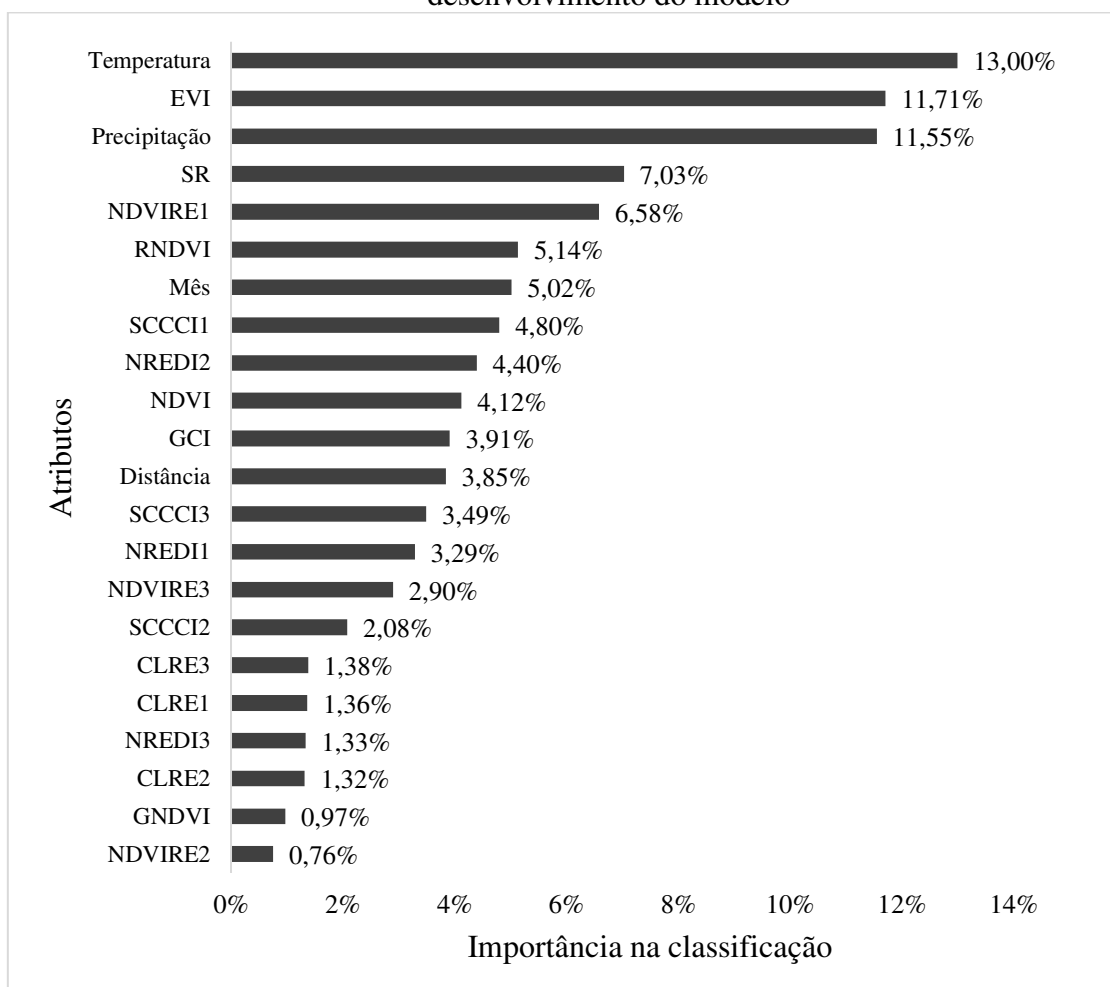
produção agrícola como replantio e colheita. Tais parâmetros fogem da alçada do presente trabalho por serem mais complexos de quantificar e exigir registros a campo com maior frequência, o que provavelmente interferiria na viabilidade do estudo.

De modo geral a performance do modelo cumpriu com o principal objetivo de gerar um aviso para todos os possíveis focos de incidência da ferrugem, utilizando de informações de fácil obtenção (imagens de satélite e dados meteorológicos). Assim sendo, salienta-se que todo o apontamento do modelo deve ser analisado com respaldo técnico e não dispensa a necessidade da avaliação dos casos em campo.

3.3 Importância dos Atributos

Conforme comentando, um dos grandes diferenciais do classificador *Random Forest* é a geração do *ranking* de importância dos atributos. Esse recurso possibilita a análise de quais variáveis o modelo está considerando como mais relevante para realizar a classificação. No presente modelo de detecção de ferrugem do cafeeiro foram utilizados o total de 22 atributos, sendo eles: o cálculo dos 18 índices vegetativos, a informação do mês de registro, a precipitação mensal, a temperatura mensal e a distância da imagem de satélite selecionada para o dia do registro a campo. O resultado do ranking de importância dos atributos gerado para o modelo é apresentado na Figura 7.

Figura 7 – *Ranking* de importância para a classificação dos atributos utilizados para desenvolvimento do modelo



Fonte: Autor.

Conforme é possível observar, três atributos se destacaram como mais importantes na classificação, sendo eles: temperatura mensal; o índice vegetativo, *Enhanced Vegetation Index* (EVI); e a precipitação mensal, com valores superiores a 10%. Os atributos que apresentaram menor importância na classificação foram os índices vegetativos CLRE3, CLRE1, NREDI3, CLRE2, GNDVI e NDVIRE2, com valores inferiores a 2%. O restante dos atributos apresentou valores de importância na classificação intermediários entre 2% a 7%.

Segundo Pires (2018), os fatores ambientais, como temperatura e precipitação, assumem papéis importantes no crescimento e desenvolvimento da cultura do cafeeiro, devido à influência nos estádios fenológicos críticos e à interferência na produtividade da cultura. Segundo os trabalhos de Silva-Acunã (1985), Carvalho (1991), Matiello (1991) e Talamini (1999) a maior ou menor incidência da ferrugem do cafeeiro é influenciada, principalmente,

por fatores climáticos, como a temperatura e precipitação. Desse modo, justifica-se a notável importância que esses atributos apresentaram na classificação.

A respeito do índice EVI, segundo Moraes, Pereira e Egidio, 2009 (2009), foi identificado que o EVI é bastante sensível a variações de aspectos da vegetação sendo indicado para avaliação do estado das culturas. Souza et al. (2015), constatou através da técnica de mineração de dados que o EVI pode ser utilizado nos modelos de ocorrência da ferrugem do cafeeiro, inclusive substituindo a necessidade das variáveis meteorológicas como temperatura e precipitação. A boa colocação (2º) do índice no *ranking* de importância na classificação é respaldada pelo também bom desempenho do índice em outros estudos de vegetação similares (ZHAO et al., 2021; TORGBOR et al., 2021)

Relacionado aos outros atributos analisados, é importante ressaltar que, apesar dos menores valores de importância apresentados, cada um teve sua contribuição na classificação. A exclusão de qualquer um desses atributos compromete o desempenho do modelo. É importante destacar também que todos os atributos utilizados na concepção do modelo são de fácil obtenção, exigindo apenas de dados meteorológicas da região e uma imagem de satélite com boa visualização para cálculo dos índices vegetativos. Esse fato torna o presente modelo de predição de ferrugem do cafeeiro simples e de fácil aplicação, sendo um diferencial de trabalhos com propósito similar já existente na literatura científica.

3.4 Métricas de Desempenho

A avaliação da performance do modelo de monitoramento da ferrugem do cafeeiro foi realizada por meio das métricas de desempenho calculadas com base na etapa de teste. O resultado do cálculo de cada uma das métricas está apresentado na tabela 7.

Tabela 7 – Resultado das métricas de desempenho do modelo desenvolvido

Métricas de desempenho	
Acurácia Global	77,3%
Precisão	38,5%
<i>Recall</i>	71,4%
Especificidade	78,4%
AUC	0,749

Fonte: Autor.

A respeito da acurácia global obtida, a interpretação é de que dentre todas as predições que o modelo fez, abrangendo casos saudáveis e infestados, 77,3% foram corretas. De acordo com Rodrigues (2019), a acurácia é uma boa indicação geral de como foi a performance do modelo, porém, é necessário a análise em conjunto de outras métricas de desempenho para justificar o bom desempenho de uma classificação.

A precisão foi a única métrica de desempenho do modelo notavelmente baixa. A interpretação do valor obtido é que dentre todos as predições que o modelo fez para a classe infestada, 38,5% estão corretas, ou seja, em alguns casos o modelo está classificando regiões saudáveis como infestadas (falso-positivo). Entende-se que essas situações ocorreram devido ao grau de sensibilidade do modelo, isto é, para que se tenha uma precisa identificação dos casos infestados, é induzido no desenvolvimento do modelo um maior rigor dessa classe, com objetivo de evitar os problemáticos casos de falso-negativos.

Segundo Pinheiro e Ascoli (2021), a métrica de desempenho precisão é mais significativa quando os casos falsos positivos são mais prejudiciais que os falsos negativos, situação que não representa a realidade do modelo desenvolvido. Diante disso, compreende-se no presente trabalho que dentre as métricas de desempenho analisadas a precisão assume o parâmetro menos importante na performance do modelo.

Por outro lado, considera-se que a métrica *recall* é a mais significativa na representação do desempenho do modelo, uma vez que, avalia dentre todas as situações de classe positiva observadas, quantas são preditas corretamente, ou seja, interpreta-se para o modelo desenvolvido, a capacidade de detectar com sucesso os casos infestados. Segundo Pinheiro e Ascoli (2021), a métrica de desempenho *recall* é mais significativa em situações que os casos falsos negativos são considerados mais prejudiciais que os falsos positivos, situação que ocorre no contexto do modelo.

A interpretação do resultado obtido para a métrica *recall* é que dentre todos os casos infestados observados a predição do modelo foi correta em 71,4%. Esse valor é considerado razoável, tendo em vista a simplicidade dos dados utilizados para desenvolvimento do modelo. Além disso, entende-se que esse valor de *recall* alcançado é suficiente para auxiliar na identificação dos possíveis focos de incidência da ferrugem no cafeeiro, principalmente em grandes lavouras, onde o monitoramento a campo da doença é complexo e trabalhoso.

Conforme discutido, o motivo de erros na predição do modelo pode estar associado a diversos fatores, uma vez que, características da lavoura cafeeira, como por exemplo, densidade de plantio, idade da lavoura, cultivar, face do relevo e diferentes tratos culturais, podem ter influência relevante no comportamento espectral das plantas de café. Tudo isso

contribuiu para o aumento da dificuldade no processo de monitoramento remoto (MOREIRA; ADAMI; RURDORFF, 2004).

A respeito da métrica especificidade, o raciocínio é similar ao da precisão, porém avalia-se a capacidade do modelo de detectar resultados negativos, ou seja, interpreta-se para o modelo desenvolvido, dentre todos os casos classificados pelo modelo como saudáveis, quantos estão corretos. O valor obtido de 78,4% é considerado bom diante o total de casos saudáveis (negativos) presente nos dados de teste (Tabela 4).

Por último, a métrica de desempenho AUC baseado na curva ROC, representa o equilíbrio do modelo entre a taxa de falso-positivo (especificidade) e a taxa de verdadeiro-positivo (*recall*). Esse parâmetro é importante pois é avaliado como o desempenho geral da classificação, permitindo a comparação com uma classificação aleatória. Temos que quanto mais próximo de 1, melhor é considerado a performance do modelo (MEURER; TOLLES, 2017). Diante disso, o valor obtido de 0,749 é considerado bom e indica que o modelo está prevendo na maioria dos casos corretamente as classes propostas. O valor de AUC obtido é reflexo direto dos também considerados bons valores das métricas *recall* e especificidade.

De um modo geral, entende-se cada métrica tem suas peculiaridades que são levadas em consideração na escolha de como o modelo de classificação é avaliado (RODRIGUES, 2019). Ressalta-se que a análise e discussão realizada não buscou identificar uma melhor ou pior que a outra, e sim analisar o problema e identificar as mais e menos significativas para o trabalho.

3.5 Considerações Finais

Diante do exposto, constata-se que, de uma maneira geral, o modelo desenvolvido pode auxiliar na tarefa de detecção remota da ferrugem do cafeeiro em escala regional. Segundo Meira, Rodrigues e Moraes (2009), a existência de medidas de controle economicamente viáveis, a variação na intensidade da doença a cada ano agrícola e sua importância para a produção de café justificam o desenvolvimento de modelos de alerta da ferrugem do cafeeiro. Kushalappa et al. (1984) desenvolveram dois sistemas de predição para ferrugem, que visavam recomendar o momento oportuno para uso de agroquímicos, os sistemas foram avaliados durante dois anos agrícolas (KUSHALAPPA et al., 1986). Nesse período, o número de aplicações de fungicidas, em comparação a uma lavoura tratada com número fixo de aplicações, diminuiu. A aplicação dos modelos como o desenvolvido no

presente trabalho proporciona maior retorno financeiro ao produtor de café e otimização de seus insumos agrícolas e do tempo operacional.

Apesar do esforço de inúmeros pesquisadores, infelizmente o uso de modelos de detecção remota de pragas e doenças não fazem parte da cultura do produtor ao redor do mundo, ainda que existam trabalhos científicos renomados na área (FLETCHER; REDDY, 2016; CHEMURA; MUTANGA; DUBE, 2017; ZHANG 2019; LEI, 2021). Entende-se que essa situação ocorra no geral por falta de conhecimento da sociedade e pelo intuitivo distanciamento entre os institutos científicos e o produtor rural. Todavia, o desenvolvimento das pesquisas na área e a tecnologia do sensoriamento remoto segue avançando continuamente, originando modelos cada vez mais precisos e aplicáveis. Acredita-se que esse avanço constante do conhecimento na área é o que pode proporcionar, gradativamente, o necessário interesse da sociedade nessa temática.

4 CONCLUSÃO

Com a elaboração desse estudo desenvolveu-se um modelo para monitoramento remoto da infestação da ferrugem no cafeeiro em escala regional por meio da utilização de índices vegetativos, dados de temperatura e precipitação da região e o algoritmo de *Machine Learning, Random Forest*. Entende-se que o uso dos 18 diferentes índices vegetativos foi capaz de detectar as alterações no comportamento espectral das plantas de café provocado pela infestação da doença. Os dados de temperatura e precipitação foram considerados essenciais para a boa performance do modelo, tendo em vista a relação direta dessas variáveis meteorológicas com a presença da ferrugem do cafeeiro.

Diante dos resultados apresentados, o desempenho do modelo desenvolvido foi satisfatório considerando a proposta principal de emissão de alertas para possíveis focos de incidência da ferrugem do cafeeiro. A adoção de atributos de fácil obtenção para construção do modelo proporcionou maior aplicabilidade do trabalho, possibilitando a replicação da metodologia para outras cultivares, regiões e doenças. O desenvolvimento de modelos como o do presente trabalho, contribui substancialmente para o avanço do conhecimento sobre o controle e monitoramento de pragas e doenças de grandes culturas no mundo.

As métricas de desempenho analisadas apresentaram consistência com o propósito do trabalho e nortearam a avaliação da performance do modelo. Contudo, reitera-se das limitações do modelo desenvolvido, tendo em vista toda a complexidade envolvida na detecção de pragas e doenças via sensoriamento remoto orbital. Entende-se que o aperfeiçoamento do modelo se daria mediante a inserção de maior quantidade de dados de registro de infestação e atributos para classificação como, umidade do solo, densidade de plantio e o registro de outros tipos de pragas e doenças. Assim sendo, mais estudos a respeito do monitoramento remoto da ferrugem do cafeeiro devem ser realizados para a evolução do controle dessa doença no mundo.

REFERÊNCIAS

- ALVES, M. C. et al. Modeling spatial variability and pattern of rust and brown eye spot in coffee agroecosystem. **Journal of pest science**, v. 82, n. 2, p. 137-148, 2009.
- BARNES, E. M. et al. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. **Proceedings of the 5th international conference on precision agriculture**. Madison, USA: American Society of Agronomy. 2000.
- BARROS, A. S.; FARIAS, L. M.; MARINHO, J. L. A. Aplicação do Índice de Vegetação por Diferença Normalizada (NDVI) na Caracterização da Cobertura Vegetativa de Juazeiro Do Norte-CE. **Revista Brasileira de Geografia Física**, v. 13, n. 6, p. 2885-2895, 2020.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- CARVALHO, A. C. et al. Panorama e importância econômica do café no mercado internacional de commodities agrícolas: uma análise espectral. **Revista Agroecossistemas**, v. 9, n. 2, p. 223-249, 2018.
- CARVALHO, V. L. **Influência de níveis de produção sobre a evolução da ferrugem e a composição química das folhas do cafeeiro**. Dissertação (Mestrado em Fitossanidade) - Universidade Federal de Lavras, Lavras, MG, 1991.
- CHEMURA, A.; MUTANGA, O.; DUBE, T. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. **Precision Agriculture**, v. 18, n. 5, p. 859-881, 2017.
- CNCAFE, Conselho Nacional do Café. Brasil: a nação do café, 2022 – Blog. Disponível em: <<https://cncafe.com.br/brasil-a-nacao-do-cafe/#:~:text=O%20caf%C3%A9%20segue%20compondo%20um,R%24%20756%20bilh%C3%B5es%20em%202021>>. Acesso em: 12 out. de 2022.
- CONAB. Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de café**, v. 9, n.3, Brasília, DF, 2022. Disponível em: <https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe/item/download/44310_265a13c31f2d1a2d604c2a6c1a542b6f>. Acesso em: 05 nov. de 2022.
- COPERNICUS CLIMATE CHANGE SERVICE (C3S). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. **Copernicus Climate Change Service Climate Data Store (CDS)**, v. 15, n. 2, p. 2020, 2017.
- COUTO, P. Assessing the accuracy of spatial simulation models. **Ecological Modelling**, v. 167, n. 1-2, p. 181-198, 2003.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). **Controle da ferrugem o cafeeiro (Hemileia vastatrix) no estado de Rondônia**. EMBRAPA, 1999. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/76700/1/FOL-7730-0001.pdf>>. Acesso em: 10 ago. de 2022.
- EUROPEAN SPACE AGENCY. **Sentinel-2 user handbook** . ESA. 2019. Disponível em: <https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook>. Acesso em: 10 set. de 2022.

- FERREIRA, L. G.; FERREIRA, N. Cl.; FERREIRA, M. E. Sensoriamento remoto da vegetação: evolução e estado da arte. **Acta Scientiarum. Biological Sciences**, v. 30, n. 4, p. 379-390, 2008.
- FERRI, C.; HERNÁNDEZ-ORALLO, J.; FLACH, P. A. A coherent interpretation of AUC as a measure of aggregated classification performance. **Proceedings of the 28th International Conference on Machine Learning (ICML-11)**, p. 657-664, 2011.
- FLETCHER, R. S.; REDDY, K. N. Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. **Computers and Electronics in Agriculture**, v. 128, p. 199-206, 2016.
- FUNK, C. et al. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. **Scientific data**, v. 2, n. 1, p. 1-21, 2015.
- GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. VSURF: an R package for variable selection using random forests. **The R Journal**, v. 7, n. 2, p. 19-33, 2015.
- GITELSON, A.; MERZLYAK, M. N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. **Journal of Photochemistry and Photobiology B: Biology**, v. 22, n. 3, p. 247-252, 1994.
- GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996.
- GITELSON, A. A. et al. Remote estimation of canopy chlorophyll content in crops. **Geophysical Research Letters**, v. 32, n. 8, 2005.
- HUETE, A. et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195-213, 2002.
- JORDAN, C. F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663-666, 1969.
- JUSTICE, C. O. et al. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE transactions on geoscience and remote sensing**, v. 36, n. 4, p. 1228-1249, 1998.
- KHAN, M. Y. et al. Automated prediction of Good Dictionary EXamples (GDEX): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. **Complexity**, 2021.
- KUHN, M. et al. **Applied predictive modeling**. New York: Springer, v. 26, 2013.
- KUSHALAPPA, A. C. et al. Equations for predicting the rate of coffee rust development based on net survival ratio for monocyclic process of *Hemileia vastatrix*. **Fitopatologia Brasileira**, v.9, p.255-271, 1984.
- KUSHALAPPA, A. C.; HERNÁNDEZ, T. A.; LEMOS, H. G. Evaluation of simple and complex coffee rust forecasts to time fungicide application. **Fitopatologia Brasileira (Brasil)** v. 11, n. 3, p. 515-526, 1986.
- KUSHALAPPA A. C.; ESKEA A. B. Resistance. Coffee rust: epidemiology, resistance, and management. **CRC Press, Boca Raton**. p. 171-292, 1989.

LEI, S. et al. Remote Sensing Detecting of Yellow Leaf Disease of Arecanut Based on UAV Multisource Sensors. **Remote Sensing**, v. 13, n. 22, p. 4562, 2021.

LIU, L. et al. A disease index for efficiently detecting wheat fusarium head blight using sentinel-2 multispectral imagery. **IEEE Access**, v. 8, p. 52181-52191, 2020.

MARIN, D. B. et al. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. **Computers and Electronics in Agriculture**, v. 190, p. 106476, 2021.

MATIELLO, J. B. **O café: do cultivo ao consumo**. São Paulo: Editora Globo, 1991.

MATIELLO, J. B. et al. **Cultura de café no Brasil: manual de recomendações**. Rio de Janeiro: MAPA/Procafé, 2010.

MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. **Pesquisa Agropecuária Brasileira**, v. 44, p. 233-242, 2009.

MEURER, W. J.; TOLLES, J. Logistic regression diagnostics: understanding how well a model predicts outcomes. **Jama**, v. 317, n. 10, p. 1068-1069, 2017.

MISRA, S.; LI, H.; HE, J. **Machine learning for subsurface characterization**. Gulf Professional Publishing, 2019.

MORAES, S. A. **A ferrugem do cafeeiro: importância, condições predisponentes, evolução e situação no Brasil**. Campinas: Instituto Agrônômico, 1983.

MORAES, E. C.; PEREIRA, G.; EGIDIO, A. R. A. I. Uso dos produtos EVI do sensor MODIS para a estimativa de áreas de alta variabilidade intra e interanual no bioma Pantanal. **Geografia**, v. 34, p. 757-767, 2009.

MOREIRA, M. A.; ADAMI, M.; RURDORFF, B. F. T. Spectral and temporal behavior analysis of coffee crop in Landsat images Abstract. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 3, p. 223–231, 2004.

PEREIRA, A. R.; CAMARGO, A. P.; CAMARGO, M.P. **Agrometeorologia de cafezais no Brasil**. Campinas: INSTITUTO AGRONÔMICO, 2008.

PINHEIRO, R. M.; ASCOLI C. P. Análise de Técnicas de Aprendizado de Máquina para Classificação de Lotes de Sementes de Soja. **Anais da 7ª Semana Integrada UFPel 2021 e XXI Encontro de Pós-Graduação: UFPel**, 2021.

PIRES, M. S. O. **Caracterização radiométrica orbital da ferrugem e da fenologia do cafeeiro em diferentes sistemas de irrigação**. Dissertação (Mestrado em Engenharia Agrícola) - Programa de Pós-Graduação em Engenharia Agrícola, Universidade Federal de Lavras, Lavras, MG, 2018.

POZZA, E. A.; CARVALHO, V. L.; CHALFOUN, S. M. Sintomas de injúrias causadas por doenças em cafeeiro. **Semiologia do cafeeiro: sintomas de desordens nutricionais, fitossanitárias e fisiológicas**, p. 68-106, 2010.

PRATES, W. R. **Curva ROC e AUC em Machine Learning**. 2020. Disponível em <<https://cienciaenegocios.com/curva-roc-e-auc-em-machine-learning/>>. Acesso em: 29 nov. de 2022.

QUEIROZ, D. M. et al. (Ed.) **Agricultura digital**. Oficina de Textos, 2021.

RODRIGUES, V. Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças? 2019. Disponível em: <<https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>>. Acesso em: 10 set. de 2022.

ROUSE, J. W. et al. Monitoring vegetation systems in the great plains with ERTS proceeding. **Third Earth Reserves Technology Satellite Symposium**, Greenbelt: NASA SP-351. 1974.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning**: From theory to algorithms. Cambridge university press, 2014.

SILVA-ACUNÃ, R. **Fatores que influenciam o processo da ferrugem do cafeeiro** (Hemileia vastatrix Berk. e Br.). Universidade Federal de Viçosa, Viçosa, 1985.

SILVA, R. A. **Repositório da Empresa de Pesquisa Agropecuária de Minas Gerais**. EPAMIG, Viçosa-MG, Brasil, 2021.

SOUZA, V. C. O et al. Aplicação de algoritmos de mineração de dados no reconhecimento de padrões influentes na ocorrência da ferrugem (Hemileia vastatrix berk. &br) em cafeeiros na região sul de Minas Gerais. **Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, João Pessoa-PB, Brasil, 25 a 29 de abril de 2015, INPE.

STEHMAN, S. V.; FOODY, G. M. Key issues in rigorous accuracy assessment of land cover products. **Remote Sensing of Environment**, v. 231, p. 1-23, 2019.

STORY, M.; CONGALTON, R. G. Accuracy assessment: a user's perspective. **Photogrammetric Engineering and remote sensing**, v. 52, n. 3, p. 397-399, 1986.

TALAMINI, V. **Progresso de ferrugem e da cercosporiose do cafeeiro (Coffea arabica L.) irrigado e fertirrigado por gotejamento**. Dissertação (Mestrado em Fitopatologia) - Universidade Federal de Lavras, Lavras, 1999.

THAMADA, T. T.; GIROLAMO NETO, C. D.; MEIRA, C. A. A. Sistema de alerta da ferrugem do cafeeiro. **VIII Simpósio de Pesquisa dos Cafés do Brasil**, Salvador - BA, 2013.

TORGBOR, B. A. et al. Assessing the Potential of Sentinel-2 Derived Vegetation Indices to Retrieve Phenological Stages of Mango in Ghana. **Horticulturae**, v. 8, n. 1, p. 11, 2021.

/, L. Current status and management of coffee leaf rust in Brazil. **Tropical Plant Pathology**, v. 41, n. 1, p. 1-8, 2016.

ZHANG, J. et al. Monitoring plant diseases and pests through remote sensing technology: A review. **Computers and Electronics in Agriculture**, v. 165, p. 104943, 2019.

ZHAO, H. et al. Automatic estimation of crop disease severity levels based on vegetation index normalization. **Remote Sensing**, v. 12, n. 12, p. 1930, 2020.

ZHENG, Q.; HUANG, W.; CUI, X. New Spectral Index for Detecting Wheat Yellow Rust Using Sentinel-2 Multispectral Imagery. **Sensors**, v. 18, n. 3, p. 868, 2018.