

EVOLUÇÃO DAS TÉCNICAS DE CLASSIFICAÇÃO DE IMAGENS DO SENSORIAMENTO REMOTO UTILIZADAS NA PRODUÇÃO CIENTÍFICA BRASILEIRA

EVOLUTION OF REMOTE SENSING IMAGE CLASSIFICATION TECHNIQUES USED IN BRAZILIAN SCIENTIFIC PRODUCTION

Wendell Santana FAGUNDES, Mauro José ALIXANDRINI JÚNIOR

Universidade Federal da Bahia. Escola Politécnica. Rua Professor Aristides Novis, 2. Federação. Salvador – BA.

E-mails: wendell-16@hotmail.com; mauro.alixandrini@ufba.br

Introdução
Conhecimento chave
Materiais e métodos
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)
Pesquisa de dados
Mineração de dados
Resultados e discussões
Conclusões
Referências

RESUMO - Atualmente, a interpretação de imagens é baseada na classificação automática. Esta se subdivide em diversos métodos, cada um baseado em um tipo de algoritmo. Deste modo, convém identificar as técnicas de classificação mais utilizadas na produção científica brasileira, bem como seus respectivos resultados e sua evolução, ao longo dos anos, a fim de orientar futuros trabalhos que pretendem utilizar estes métodos, bem como identificar tendências de suas aplicações. Este trabalho apresenta uma análise bibliométrica da produção de teses e dissertações brasileiras acerca dos classificadores automáticos de imagens, quantificando e analisando os métodos de classificação mais mencionados. A pesquisa foi efetuada na base BDTD (Biblioteca Digital Brasileira de Teses e Dissertações), para obter as principais características dos métodos empregados. Os resultados apontaram que os classificadores automáticos mais mencionados em números absolutos em ordem decrescente são: Máxima Verossimilhança, Redes Neurais, *Support Vector Machine* (SVM), Árvores de decisão, *Random Forest*, Mínima Distância Euclidiana, Mapeador de Ângulo Espectral (SAM), *K-Nearest Neighbours* (KNN) e ISODATA. Concluiu-se que os algoritmos que se atrelam ao conceito de Machine Learning estão em ascensão em números de publicações mais recentes, e que os classificadores Árvores de Decisão e Redes Neurais apresentam os valores mais elevados de acurácia nos trabalhos analisados.

Palavras-chave: Classificação de imagens. Técnicas de classificação. Sensoriamento remoto. Classificação automática de imagens

ABSTRACT - Currently, the images interpretation is based on automatic classification. This is subdivided into several methods, each one based on a type of algorithm. Thus, it is important to identify the classification techniques most used in Brazilian scientific production, as well as their respective results and their evolution over the years. In order to guide future works that intend to use these methods as well as identify trends in their applications. This work presents a bibliometric analysis of the production of Brazilian theses and dissertations about automatic image classifiers as well quantifying as analyzing the most mentioned classification methods. The research was carried out in the BDTD (Brazilian Digital Library of Theses and Dissertations) database, to obtain the main characteristics of the methods employed. The results showed that the most mentioned automatic classifiers, in absolute numbers in descending order are: Maximum Likelihood, Neural Networks, Support Vector Machine (SVM), Decision Trees, Random Forest, Minimum Euclidean Distance, Spectral Angle Mapper (SAM), K-Nearest Neighbors (KNN) and ISODATA. It was concluded that the algorithms are linked to the Machine Learning concept are on the rise in numbers of more recent publications, and that the Decision Trees and Neural Networks classifiers present the highest accuracy values in the analyzed works.

Keywords: Images classification. Classification techniques. Remote Sensing. Images automatic classification.

INTRODUÇÃO

Desde o marco inicial do sensoriamento remoto orbital, no lançamento do satélite soviético Sputnik-1 no ano de 1957, até a atualidade, vem sendo produzida uma variedade de dados orbitais e, ordenados e classificados, permitem análises espaço-temporais de períodos e áreas cada vez maiores. Se por um lado este fato contribui consideravelmente no desenvolvimento de pesquisas científicas e no planejamento e gestão acerca do espaço geográfico, por outro, o desafiante trabalho de transformar estes dados em

conhecimentos evocam métodos e conceitos das ciências computacionais para viabilização do processo.

Em meio a isso, a área de interpretação e processamento de dados do sensoriamento remoto (SR) tem absorvido fundamentações conceituais dentro da Ciência de Dados, com o intuito de tornar mais eficiente o processamento de imagens orbitais. Assim, termos como *Big-Data* e *Machine Learning* (Aprendizado de Máquina) estão cada dia mais recorrentes no meio das

geotecnologias, e sua compreensão é basilar para uma correta aplicação das técnicas computacionais de interpretação de imagens. O Big Data está atrelado ao tamanho ou volume do conjunto de dados em si, ou seja, ao grande espaço ocupado nos dispositivos de armazenamento computacional, mas também à variedade dos dados e à velocidade com que eles são processados. Já o aprendizado de máquina se relaciona à automatização dos processos de interpretação de imagens, onde um algoritmo é criado e treinado para executar tarefas em função de dados de treinamento. Neste contexto, autores como Oliveira (2019) e Liu et al. (2018), apontam que estamos vivendo a era do *Remote Sensing Big Data*, onde, segundo Wang et al. (2018), reinventam-se as formas de processamento, análise e interpretação dos dados de sensoriamento remoto para obtenção de conhecimento.

Segundo Saqui (2020) as imagens de SR são importantes fontes de informação acerca das áreas analisadas, e são constituídas por bandas espectrais, sendo que cada banda representa uma região diferente do espectro Eletromagnético. De acordo com Araújo (2017) uma das técnicas mais utilizadas para a extração de informações das imagens de satélite é a classificação, a qual permite ainda a incorporação destas informações em programas de geoprocessamento. Este último autor ilustra também que a classificação pode ser entendida como a partição do espaço de atributos segundo alguns critérios, e que os métodos de classificação podem ser divididos em dois grandes grupos: os classificadores por pixel e por região. Convém salientar que o termo “classificação automática” é genérico, dado que existem vários tipos de algoritmos para rotular as feições presentes num conjunto de imagens, ou seja, pode-se dizer que existem vários métodos de classificação automática. Atualmente, muitos destes algoritmos de classificação se atrelam aos conceitos de Aprendizado de Máquina, pois, de acordo a Bruzzone & Persello (2010) as técnicas de *Machine learning* têm sido amplamente utilizadas nas últimas décadas para a análise e classificação de dados de sensoriamento remoto, e hoje em dia estão se tornando as técnicas mais modernas, substituindo os classificadores neurais mais tradicionais.

Dentro desta variabilidade de métodos de classificação sempre fica o questionamento sobre qual algoritmo traz os melhores resultados, qual é o mais eficiente, levando-se em conta a relação entre o custo computacional e acurácia dos resul-

tados, ou ainda qual o de mais fácil aplicação. Existem diversos trabalhos que comparam classificadores, como os mais recentes de Souza et al. (2016); Oliveira (2019); Li et al. (2020); Abdollahi & Pradhan (2021). Porém, normalmente, estes trabalhos se limitam a comparar de dois a cinco tipos de algoritmos. Outra característica destas publicações são as condicionantes metodológicas, pois o mais comum é a utilização de dados de apenas um ou dois sensores, áreas de estudos específicas, adoção ou não de amostras de classe, entre outras. Entretanto, contrastando com os fatos supracitados, há de se destacar aqui as contribuições do trabalho de Ghamisi et al. (2017), pois este artigo analisou a maioria dos métodos de classificação espectral existentes na literatura, comparando criticamente as abordagens de classificação hiperespectral mais poderosas de diferentes pontos de vista, incluindo sua precisão de classificação e complexidade computacional. O artigo também sugere uma lógica para a escolha de um classificador apropriado com base no aplicativo em questão.

De acordo com Ghamisi et al. (2017) os diversos classificadores fornecem desempenhos diferentes para imagens distintas, indicando que não há nenhum classificador que forneça consistentemente melhores resultados de classificação para diferentes cenas. Estes autores afirmam também que, em última análise, na escolha de um determinado classificador pesa mais a simplicidade de ajuste dos parâmetros e configurações do que a qualidade dos resultados obtidos pelo classificador. Esta é uma consideração importante, pois se observa que a comunidade científica produziu classificadores capazes de fornecer alta acurácia nas classificações. Entretanto, as diferenças competitivas entre classificadores existentes estão mais relacionadas à sua simplicidade e configurações de ajuste do que a seus resultados. Neste mesmo sentido, Li et al. (2020) concluíram em seus trabalhos que a maioria dos métodos atuais tem limitações de classificação, e que nenhum classificador é o mais adequado para todas as situações de estudo. O método de classificação deve ser selecionado de acordo com as características espectrais da imagem de sensoriamento remoto e os requisitos reais de trabalho. O método certo é aquele que pode melhorar a precisão da classificação.

Independentemente do método de classificação, não se pode discutir a relevância desta técnica para transformar os dados brutos de SR em mapas

temáticos e informações acerca do uso e cobertura terrestre, sendo que a classificação de imagens pode ser aplicada em diferentes áreas do conhecimento, como na agricultura, no planejamento territorial urbano e rural, gerenciamento de recursos hídricos, monitoramento ambiental e ecológico, geologia, aplicações militares, entre outros. Na área acadêmica é vasta a bibliografia que emprega as técnicas de classificação de dados do SR. Para ilustrar melhor esta afirmativa, destaca-se aqui aplicações recentes feitas na agricultura de precisão por Chunjing et al. (2017); Chen et al. (2016); Hasituya et al. (2017). Em análises ambientais e ecológicas, os trabalhos de Flores et al. (2020); Mahdianpari et al. (2020); Han et al. (2021). No escopo de recursos hídricos: Choubin et al. (2017); Sekertekin (2021). Já Turgay & Yaşar (2021); Duan et al. (2020) usaram a classificação para mapeamentos voltados à geologia. Dentre outras inúmeras publicações e aplicações acerca da classificação de imagens que poderiam ter sido citadas.

Assim, tendo em vista a variedade de métodos de classificação e o estágio Big data do sensoriamento remoto, convém identificar a evolução dos métodos de classificação utilizados na produção científica brasileira, a fim de orientar futuros trabalhos que pretendem usar estes métodos, bem como identificar as atuais tendências de suas aplicações dentro da produção científica brasileira.

Deste modo, o presente trabalho propôs uma análise bibliométrica das teses e dissertações brasileiras direcionadas à classificação de imagens, levantando-se quais os classificadores mais mencionados por estes trabalhos de conclusão de curso. Este estudo analisou ainda a quantidade de menções de cada classificador por teses, por dissertações, por instituição de ensino, e principalmente a evolução da aplicação de cada classificador ao longo dos anos, possibilitando identificar quais os métodos mais atuais. Uma outra contribuição deste trabalho foi a catalogação, por meio de tabela síntese, dos resultados de algumas teses e dissertações acerca dos classificadores, onde elencou-se a acurácia, a área de aplicação e o sensor, entre outras informações relevantes, que permitirão compreender melhor a aplicabilidade dos métodos de classificação e seus respectivos resultados.

Conhecimento Chave

Segundo Florenzano (2011), a metodologia para gerar conhecimento a partir de dados do SR compreende as etapas de pré-processamento,

extração de informação e apresentação da informação. No pré-processamento, geralmente, se faz a melhoria geométrica, posicional e radiométrica da imagem, além de alguns processos que facilitem a obtenção dos resultados, como o recorte da área de estudo, por exemplo. A extração da informação é a etapa precípua da metodologia, onde o processo de transformar os dados orbitais em conhecimento acerca do espaço é intermediado pela aplicação de diferentes técnicas, desde as mais simples como a interpretação visual, até as mais complexas que empregam princípios de redes neurais artificiais, *Machine Learning* ou *Deep Learning*. A etapa final de apresentação da informação se volta à representação cartográfica, assunto não menos importante, mas que não é foco deste estudo.

Atualmente, o modo mais difundido para extrair informações de dados do SR é a classificação automática de imagens, a qual pode ser compreendida como o termo genérico dado “ao processo de atribuir significado a um pixel em função de suas propriedades numéricas” (Novo, 2008). De modo mais aprofundado, Richards & Jia (1999) conceituam a classificação como a técnica de rotular pixels de uma imagem em função das suas particularidades espectrais, dentro de um programa computacional capaz de reconhecer a similaridade espectral dos *pixels*.

Os métodos de classificação são agrupados segundo várias abordagens. No presente estudo, mencionaremos três dessas abordagens por serem as mais recorrentes na literatura. As abordagens que agrupam tipologia de classificadores de imagens são: quanto ao modo de treinamento do algoritmo; quanto a adoção ou não de uma função de probabilidade conhecida para cada classe; e quanto ao tipo de informação do *pixel* aplicada como entrada. A figura 1 traz um resumo dos tipos de classificadores segundo estas três abordagens.

Com relação a abordagem sobre o modo de treinamento do classificador, a classificação pode ser dividida em duas categorias: classificadores supervisionados e não supervisionados. O Supervisionado classifica os dados de entrada para cada classe utilizando, a priori, um conjunto de amostras representativas conhecidas como parâmetros de treinamento. Em contrapartida, a classificação não supervisionada desconsidera o treinamento por amostras, classificando os dados com base apenas em um número arbitrário de centros de cluster iniciais que podem ser especi-

ABORDAGEM	TIPO DE CLASSIFICADOR	CARACTERÍSTICAS	EXEMPLOS DE CLASSIFICADORES
Se as amostras de treinamento são usadas ou não.	Supervisionado	classificam os dados de entrada usando um conjunto de amostras para cada classe, conhecidas como amostras de treinamento.	MAXVER, MinDis, PLLP, SVM, SAM, AD, RF
	Não supervisionado	também conhecidas como clustering, não consideram amostras de treinamento, rotulam posteriormente em função de centros de clusters iniciais que pode ser especificado pelo usuário ou selecionado arbitrariamente	Isodata, Isoseg, KNN
Se alguma suposição sobre o distribuição dos dados de entrada é considerado ou não	Paramétrico	Os classificadores paramétricos são baseados na suposição de que a função de densidade de probabilidade é conhecida para cada classe.	MAXVER, MinDis, PLLP, Isoseg, SAM, Isodata
	Não Paramétrico	Classificadores não paramétricos não se limitam por quaisquer suposições sobre a distribuição de dados de entrada.	SVM, RNA, KNN, RF, AD
Que tipo de informação de pixel é usado.	Por Pixel	Vetores de pixel amostrais alimentam os classificadores com o entradas.	MAXVER, MinDis, PLLP, Isodata, SVM, SAM, RF
	Orientada a objeto	Nesta abordagem, uma técnica de segmentação aloca um rótulo para cada pixel na imagem de forma que pixels com o mesmo rótulo compartilhem certas características visuais. Neste caso, os objetos são conhecidos como unidades subjacentes após a aplicação da segmentação. Classificação é conduzido com base nos objetos em vez de um único pixel	Isoseg, SVM, AD, RF

Legenda: MAXVER = Máxima verossimilhança; MinDis = Mínima Distância; PLLP = Paralelepípedo; SVM = Support Vector Machine; AD = Árvores de decisão; RF = Random Forest; RNA = Redes Neurais; KNN = K-Nearest Neighbours

Figura 1 - Tipos de classificadores de imagem segundo três abordagens (Adaptado de Ghamisi et al., 2017).

ficados pelo usuário ou arbitrados (Ghamisi et al., 2017). Uma terceira categoria desta abordagem, que tem aparecido na literatura, é a semi-supervisionada ou híbrida, a qual, segundo Theodoridis & Koutroumbas (2003, apud Oliveira, 2020) é aplicada quando apenas uma parte dos dados tem rótulo conhecido e se necessita buscar informações sobre os dados não rotulados. Também no caso da “clusterização”, o modo semi-supervisionado pode ser usado de maneira que a pequena parcela de dados rotulados seja utilizada como parâmetro para incluir ou não um ponto em um determinado cluster.

A abordagem acerca da adoção ou não de uma função de probabilidade conhecida para cada classe bifurca-se em duas categorias: Paramétrica e não-paramétrica. Nos algoritmos paramétricos, conforme Andreola (2009), a forma geral da função de densidade de probabilidade, que representa o comportamento dos dados, é supostamente conhecida, além disso estima-se os parâmetros existentes através das amostras de treinamento para cada classe, levantadas pelo interpretador. Já no caso dos algoritmos não-paramétricos, segundo Sartori (2006), assume-se o não conhecimento da forma da função de probabilidade, sendo que as funções de decisão são produzidas diretamente a partir das amostras de treinamento.

Já na abordagem que leva em consideração o tipo de informação do pixel aplicada como entrada, a classificação pode ser por *pixel* ou por região (ou orientada a objeto). Na classificação por pixel se utiliza somente a informação espectral de cada pixel para classificar regiões homogêneas, a partir de medidas de distâncias ou de proba-

bilidades de um pixel pertencer a uma classe em especial (Meneses & Almeida, 2012). Já na classificação orientada a objeto o elemento de classificação deixa de ser o pixel, e passa a ser um conglomerado de *pixels*, os quais apresentam homogeneidade em suas propriedades espectrais. (Branco et al., 2018).

A última etapa do trabalho classificatório é a avaliação da acurácia das representações temáticas do mapa gerado. Segundo Meneses & Almeida (2012) a acurácia temática de uma classificação está vinculada a diversos fatores, destacando-se a complexidade do terreno, as resoluções espaciais e espectrais do sistema sensor, o método de classificação utilizado, os dados amostrais da verdade terrestre, entre outros. Novo (2008) adverte que a alta correlação entre os canais espectrais, ou seja, a semelhança visual e numérica de bandas do sensor, pode ser prejudicial na distinção das classes. Segundo Farias et al. (2012), esta correlação tem origem no sombreamento resultante da topografia, da proximidade ou sobreposição das faixas espectrais adjacentes e do próprio comportamento espectral dos alvos terrestres.

A acurácia de um mapa classificado é estimada por indicadores estatísticos, sendo a matriz de erro ou confusão a forma mais utilizada. De acordo com Congalton (1991) a matriz de erro é uma matriz quadrada de números que indicam unidades de amostragem (ou seja, pixels, aglomerados de pixels, ou polígonos) atribuídas a uma determinada classe, correlacionada à classe verificada em solo. A matriz de confusão pode ser usada como um ponto de partida para uma série de indicadores estatísticos descritivos e analíticos

(Congalton, 1991), dentre eles destacam-se a exatidão global e o índice Kappa.

De acordo com Meneses & Almeida (2012) a exatidão global ou acurácia geral (*overall accuracy*) é computada dividindo-se o número total de pontos corretamente classificados (soma dos elementos ao longo da diagonal principal da matriz de erros) pelo número total de pontos de validação.

Deste modo, um resultado com 100% de acurácia indica que todos os pixels da imagem foram corretamente classificados, baseando-se num conjunto de dados que compõe a verdade terrestre. Um resultado com 50% de acurácia

significa, teoricamente, que metade dos pixels da imagem foi classificada de modo correto.

O índice Kappa pode ser conceituado como uma medida de associação que descreve e testa o grau de concordância (confiabilidade e precisão) de uma classificação (Kotz & Johnson, 1983 Apud Perroca & Gaidzinski, 2003). Segundo Figueiredo & Vieira (2007) a estimação do Kappa é amplamente usada na avaliação de acurácia temática, caracterizando-se como uma técnica multivariada discreta que usa todos os elementos da matriz de erro em seu cálculo. Esse índice varia de 0 (concordância nula) a 1 (concordância perfeita).

MATERIAIS E MÉTODOS

Os dados principais do presente estudo foram obtidos por meio de pesquisas bibliométricas em uma base acadêmica de teses e dissertações brasileiras, a BDBT, onde foram efetuadas buscas avançadas com palavras-chave específicas, cujas análises se materializaram nos principais resultados desta pesquisa. Efetuou-se também pesquisas bibliográficas em outras bases como a Scopus, *Google Scholar* e Scielo para embasamento teórico deste trabalho. A seguir serão apresentadas as etapas e demais informações relevantes acerca da produção deste trabalho.

Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)

A Biblioteca Digital Brasileira de Teses e Dissertações é uma base de pesquisas acadêmicas focada nos trabalhos de conclusão de cursos brasileiros em nível de pós-graduação stricto sensu. Esta biblioteca foi desenvolvida e é coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). A BDTD integra os sistemas de informação de teses e dissertações produzidas nas instituições de ensino e pesquisa do Brasil, estimulando o registro e a publicação destes trabalhos em meio eletrônico. A BDTD, em parceria com as instituições brasileiras de ensino e pesquisa, possibilita que a comunidade brasileira de ciência e tecnologia possa publicar e difundir suas teses e dissertações produzidas no País e no exterior, dando maior visibilidade à produção científica nacional (IBICT, 2021).

No sítio eletrônico da BDTD constam, nesta data, 684.779 publicações acadêmicas, sendo 500.797 dissertações e 183.983 teses, referentes a 125 instituições de ensino superior diferentes. A base BDTD possui várias ferramentas de buscas avançadas, utilizando algoritmo de estatística

Fuzzy, busca por proximidade, busca por intervalo, incremento de um termo e por operadores booleanos.

Pesquisa dos dados

Antes de se aplicar a pesquisa avançada na BDTD, foi necessário um estudo prévio em artigos, periódicos e bases acadêmicas sobre os diversos tipos de classificadores e seus respectivos termos referentes, como forma de tornar mais efetiva a busca na BDTD.

Uma ação que contribuiu bastante, além da leitura de artigos, foi uma busca avançada na base Scopus, utilizando-se os termos “*remote sensing*”, “*classification*”. Deste modo, o retorno da pesquisa foram milhares de trabalhos acadêmicos acerca da classificação de imagens de sensoriamento remoto, que poderiam ser minerados de diferentes formas em função das ferramentas de refinamento de busca presente na Scopus. Uma destas ferramentas, não disponível na BDTD, é o refinamento por outras palavras-chave que se vinculam aos termos da busca inicial. Ao verificar tais palavras-chave do refinamento, foi possível identificar vários termos que se referem aos muitos métodos de classificação, bem como a quantificação de trabalhos que mencionam estes métodos. Salienta-se que muitas destas palavras-chave tinham alguma variação por conta de plural e singular, como por exemplo “*decision tree*” e “*decision trees*”; outras tinham variações por abreviação, como “*Support Vector Machine*” e “*SVM*”; em outros casos, as variantes se davam pelos diferentes termos referentes a um mesmo tipo de algoritmo ou por subdivisões dentro de um mesmo método, como por exemplo “*Artificial Neural Network*”, “*Neural Network*” e “*Deep Neural Network*”. Assim, tal pesquisa

prévia possibilitou elencar os métodos de classificação mais mencionados na base Scopus e seus principais termos referentes, como forma de nortear a etapa de buscas na BDTD. Ressalta-se ainda que foi necessário um conhecimento prévio sobre os tipos de classificadores, adquirido por meio de estudos em artigos e outros trabalhos científicos.

Com a lista das possíveis terminologias que se referem aos vários métodos de classificação automática de imagens do SR, iniciaram-se as buscas na base BDTD, onde estes termos se tornaram as palavras-chave para as buscas.

Deste modo, efetuou-se uma busca para cada tipo de classificador. Definiu-se, por autoevidência, que o termo “sensoriamento remoto” iria ser aplicado em todas as buscas, e, atrelados a ele por meio do operador booleano “AND” os signos que se referem ao método pesquisado, estes últimos separados pelo operador “OR”. Os resultados eram analisados por amostragem aleatória para verificar se realmente a palavra-chave estava contida no trabalho.

Ao longo das buscas, alguns classificadores primariamente listados, não foram colocados em pauta por conta da sua baixa quantidade de menções, a exemplo do *Isoseg* e *Paralelepípedo*. Determinou-se arbitrariamente que classificadores com menos de 10 menções não seriam pautados neste estudo. Outro classificador que não foi listado nos resultados foi o *K-means*. Os termos “*k-means*” ou “*k-medias*” estavam causando inconsistências na busca, pois trabalhos que continham os termos “*means*” ou “*médias*” estavam sendo vinculados

nos resultados da pesquisa, não havendo, na base BDTD uma ferramenta que pudesse distinguir os termos.

Após cada análise prévia de resultados, fazia-se o *download* em arquivo CSV contendo os detalhes das bibliografias resultantes da pesquisa. No total foram 382 trabalhos pesquisados. Com este arquivo em disco local, procedeu-se o refinamento dos dados em software de manipulação de planilhas.

Mineração dos dados

Utilizando programa de planilhas eletrônicas procedeu-se o refinamento e análise dos resultados de cada classificador. Foram quantificadas as menções totais, mas também por nível de trabalho (Tese ou dissertação), por quinquênio e pela instituição acadêmica. Em seguida, as informações foram materializadas em gráficos e quadros para melhor ilustrarem os resultados desta pesquisa.

Por fim, foram selecionados 15 trabalhos, dentre os encontrados em todas as pesquisas da BDTD, e com base neles foi construída uma tabela síntese que relaciona as acurácias de cada classificador, encontradas nos trabalhos, mas também os satélites utilizados e a temática central (característica) da classificação. Diante da inviabilidade de se analisar 382 trabalhos na construção da tabela síntese, os 15 trabalhos que a compunha foram selecionados segundo os seguintes critérios: Mais relevantes segundo ordenamento da BDTD; Publicações dos últimos 10 anos; Textos que contemplem algum tipo de comparação, seja entre classificadores ou entre sensores. Além destes critérios, estabeleceu-se também que cada classificador tivesse pelo menos três autores elencados.

RESULTADOS E DISCUSSÕES

Os principais resultados deste trabalho consistem em tabelas e gráficos que elencam os classificadores automáticos de imagem de SR mais mencionados em trabalhos de conclusão de curso *stricto sensu* no Brasil. A seguir serão apresentados tais resultados com as respectivas discussões acerca de cada um.

Em números absolutos, constatou-se que o classificador de Máxima Verossimilhança (MAXVER) é o mais mencionado, dentre todos os trabalhos pesquisados. Em seguida tem-se, em ordem decrescente no número de menções, os classificadores de redes neurais (RNA), *Support Vector Machine* (SVM) e Árvores de decisão (AD), todos com mais de 50 menções. Na sequência, também em ordem decrescente e com menos de 50 menções temos o *Random Forest*

(RF), Mínima Distância (MINDIS), *Spectral Angle Mapper* (SAM), *K-Nearest Neighbours* (KNN) e ISODATA.

Para ilustrar melhor esta análise, foi plotado o gráfico da figura 2, que elenca os classificadores mais mencionados em teses e dissertações no Brasil.

O classificador MAXVER, o mais utilizado, se baseia no teorema de Bayes (Santos et al., 2003), sendo uma das relações mais relevantes em se tratando de probabilidade condicional (Bussab & Morettin, 2010). Segundo Santos et al. (2010), o princípio básico de utilização do MAXVER é o conhecimento a priori da região estudada (classificador supervisionado), no intuito de indicar áreas de treinamento mais verossímeis possíveis à classe atribuída, para

então, por meio do Teorema de Bayes, classificar os *pixels* da imagem. O conjunto de treinamento permite ao algoritmo determinar o diagrama de

dispersão das classes e suas distribuições de probabilidade, supondo estas distribuições como gaussianas.

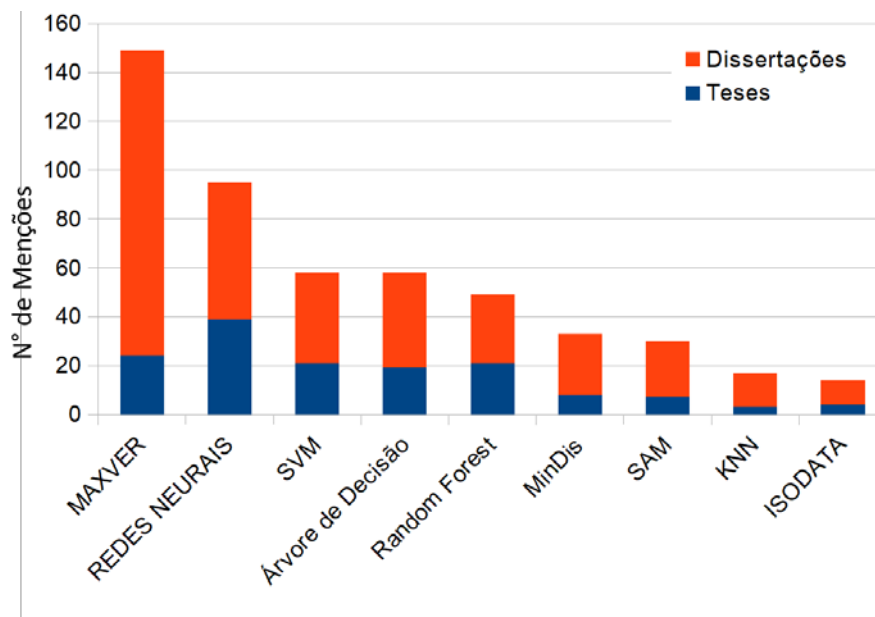


Figura 2 - Classificadores mais mencionados em teses e dissertações no Brasil.

Nem sempre números absolutos são suficientes para descrever um fenômeno. Diante disso, a análise de menções por quinquênios possibilitou uma melhor compreensão de como a aplicação de cada classificador evoluiu ao longo dos anos.

Os gráficos plotados na figura 3 mostram quanto cada classificador foi mencionado ao longo dos quinquênios (3a) e as menções proporcionais de cada classificador por quinquênio (3b), na produção acadêmica de pós-graduação *stricto sensu* no Brasil.

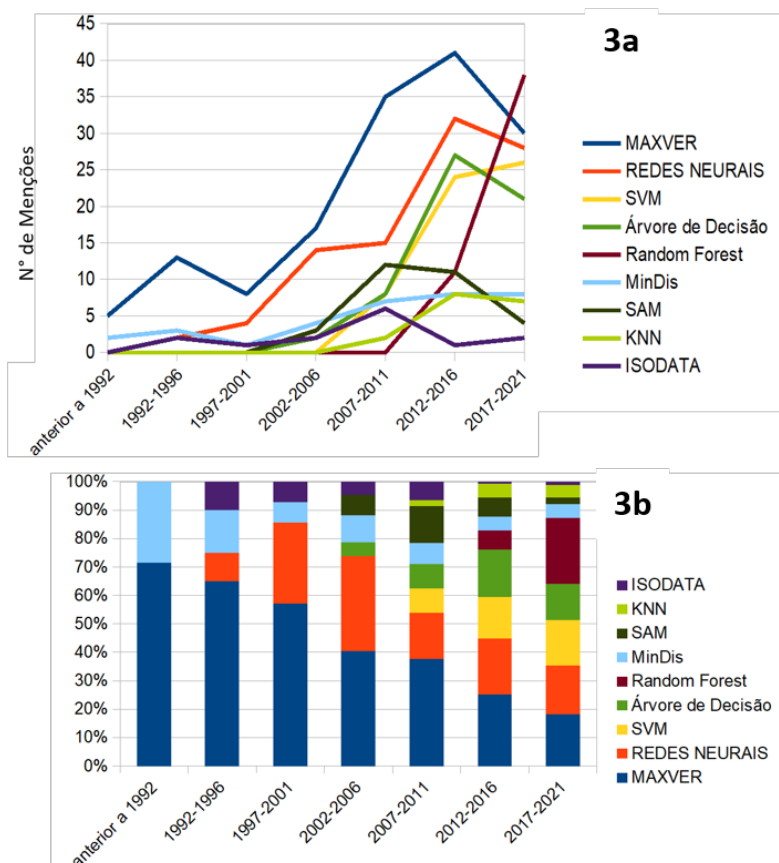


Figura 3. (“a” e “b”): Evolução das menções a classificadores em teses e dissertações do Brasil ao longo dos últimos seis quinquênios.

Analisando o gráfico acima, é possível identificar que o classificador MAXVER foi o mais mencionado em quase todos os quinquênios, menos no último, que teve predominância do *Random Forest* como o mais aplicado em teses e dissertações brasileiras. O *Random Forest* pode ser sintetizado como um conjunto de árvores de decisão geradas dentro de um mesmo objeto, em que cada conjunto (objeto) passa por um mecanismo de votação (*bagging*) elegendo a classe mais votada. Tal método é uma combinação de preditores de árvores, de maneira que cada árvore necessita dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores na floresta. O método RF é um algoritmo supervisionado de *machine learning* que usa aprendizado por conjuntos para classificação e regressão (Rambo, 2020). Outra informação que se pode tirar do gráfico a respeito do RF é que ele passou a ser mencionado apenas a partir do quinquênio 2007-2011, em contrapartida, ele foi o classificador com ascensão de menções mais expressiva entre os classificadores listados, observada entre os quinquênios 2012-2016 e 2017-2021.

Junto ao RF, tem-se o SVM e ISODATA em fase de ascensão em suas curvas. O MINDIS manteve número igual de menções nos dois últimos quinquênios, enquanto os demais classificadores estão num período de decréscimo no número de aplicações nos trabalhos acadêmicos. O SVM é uma técnica de *machine learning* que se fundamenta em princípios de Minimização do Risco Estrutural, buscando minimizar erros com relação ao conjunto de treinamento (risco empírico), e com relação ao conjunto de teste, ou seja, com o conjunto de amostras não empregadas no treinamento do classificador (risco na generalização). O SVM objetiva a obtenção de equilíbrio entre esses erros, minorando o excesso de ajustes relacionados às amostras de treinamento e aumentando a capacidade de generalização (Araújo, 2017). Esta recente ascensão do RF e do SVM denotam a modernidade destes classificadores, e se deve, provavelmente, pelo fato de serem técnicas de aprendizado de máquina, as mais modernas neste âmbito.

Quanto ao ISODATA, presume-se que sua ascensão não é significativa por conta das costumeiras variações em sua curva ao longo dos quinquênios, mas também pela proporção desta ascensão, tendo uma menção do quinquênio 2012-2016 e duas menções em 2017-2021. O

ISODATA é um algoritmo de agrupamento não supervisionado que permite ajuste do número de classes automaticamente durante as interações por fusão de agrupamentos semelhantes e divisão de classes com elevado desvio padrão (Jensen, 2009).

O MINDIS, que manteve linearidade na curva de menções nos últimos dois quinquênios, classifica um novo *pixel* a partir da determinação da menor distância espectral entre o *pixel* em questão e os centroides das classes, que, por sua vez, são formadas por suas respectivas amostras de treinamento (Araújo, 2017).

Percebe-se também que o quinquênio 2012-2016 foi o período de ápice da curva de três importantes classificadores: O já mencionado MAXVER, Redes Neurais e Árvores de decisão. As Redes Neurais são um conjunto de neurônios artificiais que tentam imitar a estrutura neural biológica, e, conforme Gao (2009), os neurônios artificiais são nós que recebem entradas a partir de um conjunto de nós conectados, e todas estas conexões são similares aos dendritos de um neurônio biológico. Os sinais de entrada do nó artificial são combinados linearmente ou não-linearmente, gerando uma saída através de uma função de transferência (limiar). Estas saídas repassadas ao neurônio artificial se baseiam no axônio biológico. Cada unidade atua como um processador de reconhecimento de padrões simples, sendo que as entradas são avaliadas de acordo à força sináptica e assim se produz uma saída. Na implementação computacional, o neurônio artificial é como uma soma ponderada dos sinais numéricos de entrada.

Já as Árvores de Decisão ou *Decision tree* são técnicas de aprendizado de máquina que realizam a classificação efetuando a criação de padrões com os dados de treinamento e, a partir da árvore obtida, classificar novas amostras desconhecidas. Estas árvores consistem numa hierarquia de nós internos e externos os quais são conectados por ramos. Cada nó representa uma decisão sobre uma variável por meio de testes lógicos, que em seguida se ramifica para o próximo nó descendente ou resultado final, que neste último caso se chama folha (Rambo, 2020).

O SAM teve curva com um comportamento mais particular, começando a ser mencionado no quinquênio 1997-2001, teve seu ápice entre 2007-2011, manteve certa linearidade até 2012-2016, entrando num período de desuso no último quinquênio.

O SAM, de acordo com Souza (2013 apud

Kruse et al., 1993), é um classificador espectral que determina de similaridade entre dois espectros através do cálculo do ângulo formado entre os vetores que os representam matematicamente, em um espaço cuja dimensionalidade é igual a quantidade de bandas. Este ângulo entre os dois vetores é chamado de ângulo espectral e seu valor é inversamente proporcional à similaridade espectral entre os espectros. O algoritmo SAM generaliza a interpretação geométrica dos vetores para o espaço n dimensional ($n = n^\circ$ de bandas).

O classificador KNN é um classificador não supervisionado, pois dispensa amostras de treinamento para construir um modelo. Este algoritmo realiza comparativo com os dados iniciais, em seguida os *pixels* desconhecidos são alocados na classe mais comum das “K” amostras de treinamento que estão mais próximas no espaço. Assim, um “K” baixo produz um limite de decisão mais complexo, enquanto que um “K” de maior valor produzirá uma generalização maior na classificação (Maxwell et al., 2018). Analisando sua curva no gráfico é possível identificar que teve suas primeiras aplicações nos anos 2002-2006, e seu ápice de menções no quinquênio 2012-2016, tendo uma pequena redução em suas aplicações no último quinquênio.

Em se tratando das instituições de pesquisa e ensino, o resultado da bibliometria indicou que o Instituto Nacional de Pesquisas Espaciais (INPE) foi o que produziu o maior número de teses e dissertações com a temática de classificação de imagens de sensoriamento remoto, representando 18,64% de todos os trabalhos brasileiros nesta área. Este número é muito expressivo, dado que foram catalogadas 45 instituições de ensino e pesquisa com produção neste âmbito. No gráfico de pizza da figura 4, estão listadas as instituições de ensino e pesquisa com maior representatividade na produção de teses e dissertações acerca de classificadores de imagem, destacando-se as nove primeiras instituições: INPE, UFRGS (Universidade Federal do Rio Grande do Sul), UNB (Universidade de Brasília), UFV (Universidade Federal de Viçosa), UFPR (Universidade Federal do Paraná), UNICAMP (Universidade Estadual de Campinas), USP (Universidade de São Paulo), UNESP (Universidade Estadual Paulista) e UFRRJ (Universidade Federal Rural do Rio de Janeiro). A categoria “outros” do gráfico representa 27 outras instituições de ensino superior do Brasil, não mencionadas por sua baixa representatividade individual em relação aos números totais.

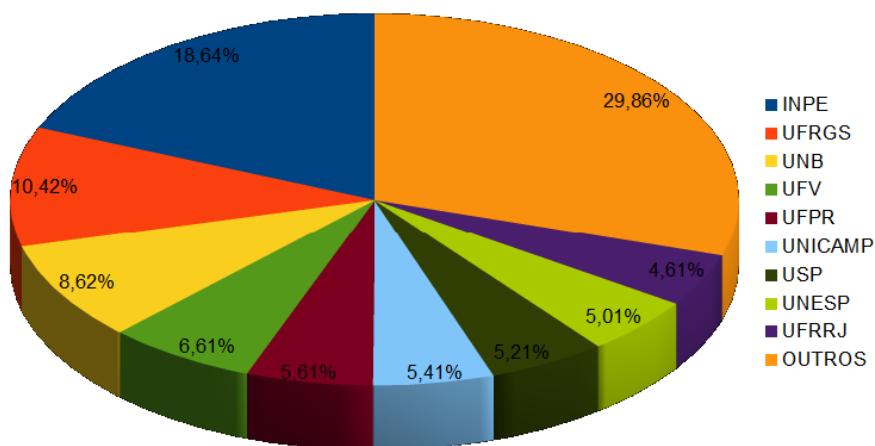


Figura 4 - Distribuição das teses e dissertações a respeito de classificação de imagens do SR por instituição de ensino e pesquisa do Brasil.

Quanto ao estudo bibliográfico dos 15 trabalhos selecionados, obteve-se como principal resultado a tabela síntese expressa na figura 5, a seguir. Convém salientar que o fato de se priorizar trabalhos recentes (dos últimos 10 anos), pode não evidenciar tanto classificadores mais antigos, diante disso, se estabeleceu que no mínimo cada classificador tivesse pelo menos três autores elencados.

Foi possível identificar nestes trabalhos que, em geral, os classificadores têm bons e ótimos

resultados de acurácia, com destaque para as Redes Neurais, que obtiveram o maior valor mínimo de acurácia dentre os classificadores catalogados, mas também às Árvores de Decisão, das quais foi possível obter o maior valor de acurácia dentre as teses e dissertações estudadas. O SVM e o MAXVER são os classificadores mais regulares, tanto em termos de aplicação, de baixa variância na acurácia, ou seja, mantém sempre um nível satisfatório quando aplicados, bem como aplicações em diferentes temáticas. O

AUTORES	ACURÁCIA GERAL (OVERALL ACCURACY) DO CLASSIFICADOR										SATELITE										CARACTERÍSTICA DA CLASSIFICAÇÃO				
	MAXVER	RNA	SVM	AD	RF	SVM	MINDIS	KNN	ISODATA	LANDSAT-8	SENTINEL-1	CECETE	RAPIETE	SUOT	WORLDVIEW	TERRA (MODIS)	EO-1 (HYPERION)	ER-2 (AVIRIS)	ECOSIS	ALCS (PALSAR)		ENVISAT (ASAR)	RAADMAT-2	TERRASAR-X	
Marinho, R. R. (2011)								0,917 ^a 0,950 ^a 0,915 ^a 0,900 ^a											X ^a	X ^a	X ^a	X ^a		Mapeamento de inundações e deslizamentos de terra	
Albuquerque, R. W. (2012)	0,860		0,861								X														Uso e cobertura do solo
Passo, D. PO. (2013)	0,858					0,737								X											Identificação de alvos urbanos
Santos, A. B. (2013)		0,846 ^a 0,957 ^a	0,871 ^a 0,960 ^a					0,612 ^a 0,883 ^a									X ^a	X ^a							Uso e cobertura do solo
Silva Júnior, C. A. (2014)		0,826	0,981											X											Deteção de áreas com plantio de soja
Durán, G. Y. B. (2014)						0,519 ^a 0,535 ^a					X ^a		X ^a												Comportamento espectral de alvos urbanos
Abade, N. A. (2015)						0,538	0,643							X											Mapeamento da cobertura vegetal natural e exótica
Paídua, C. S. (2016)*	0,835	0,891							X																Uso e cobertura do solo
Aranjo, E. M. (2017)*	0,530 ^a 0,780 ^a		0,810 ^a		0,670 ^a		0,490 ^a	0,680 ^a	0,430 ^a 0,400 ^a	X ^a						X ^a									Discriminação dos usos e ocupação do solo no entorno de reservatório
Neves, A. K. (2017)		0,774		0,766	0,783				X																Áreas de Pastagem na Amazônia Legal
Schwartzman, L. F. G. (2017)	0,927						0,721		X																Mapeamento da cobertura do solo em regiões de floresta tropical
Souza, W. W. C. (2019)	0,687		0,755			0,681		0,781				X													Uso e cobertura do solo
Nóbrega, R. A. A. (2019)	0,782			0,929					X			X**													Mapeamento de Vendas em APA
Costa, D. H. (2020)			0,740		0,772			0,827		X															Distinção de culturas agrícolas
Rambo, E. M. (2020)			0,839	0,869	0,816				X																Sinistro em culturas agrícolas

*Nestes casos o parâmetro de acurácia foi o fator Kappa
**O mapa classificado desta imagem serviu apenas como parâmetro de entrada para o classificador baseado em árvore de decisão

Figura 5 - Quadro síntese a respeito das características das classificações de dados do SR realizadas por alguns autores de teses e dissertações do Brasil.

RF também é uma boa alternativa quando se busca elevadas acurácias, pois de acordo com Breiman (2001) a inserção de aleatoriedade do RF torna este classificador preciso. O RF, ao lado do KNN, tiveram boas acurácias e regularidades em seus resultados, só que um pouco abaixo do SVM e MAXVER.

Os classificadores SAM e MINDIS, com valores um pouco menos expressivos que o RF e KNN, se mostraram estáveis e com valores de acurácia regulares, dentro dos seus resultados estudados.

Por fim, o ISODATA com resultados instáveis, atingindo limiares de 0,95 num estudo e valores de 0,40 noutro, não se mostrou confiável para classificações de produtos finais, podendo ser aplicado em classificações prévias que requerem baixo custo computacional.

Entretanto, adverte-se que o limiar médio de

acurácia global não deve ser o único norteador na escolha de um classificador, pois a acurácia varia conforme os tipos de classe do trabalho classificatório. Cada classe se comporta de maneira diferente ao ser analisada pelo algoritmo, pois há classes que possuem conduta espectral e características interpretativas mais ímpares, enquanto que outros tipos de classe apresentam particularidades tais que elevam o nível de confusão durante o processo automático de classificação, impactando nos resultados de acurácia. Isto evidencia a necessidade de estudar cada trabalho de classificação de imagens e sua forma de aplicar o classificador em questão: as abordagens classificatórias, os tipos de classe em análise, o tipo de terreno, o tipo de uso do solo (urbano ou rural), o sensor que gerou as imagens, entre outros aspectos.

CONCLUSÕES

O classificador MAXVER, o mais mencionado em números gerais na produção de teses e dissertações do Brasil, se mostra como o classificador de maior confiança da comunidade acadêmica da área de sensoriamento remoto. Pois, além de ser muito utilizado em aplicações diversas, sempre aparece como uma referência quando um estudo resolve comparar métodos de classificação. Além disso, seus resultados de acurácia o respaldam para sua aplicação geral.

Quanto as Redes Neurais e as Árvores de Decisão, são os métodos com resultados de acurácia mais elevados. Suas aplicabilidades em trabalhos de teses e dissertações são relevantes,

estando ranqueadas entre os métodos mais utilizados. A RNA é uma técnica mais antiga, já as AD's apesar de mais recentes obtiveram ápice de aplicações junto as RNA's no quinquênio 2012-2016, e no último quinquênio tiveram um decréscimo em sua utilização.

O SVM, ao lado do RF, dois classificadores de *Machine Learning*, estão em ascensão, podendo-se dizer que são umas das técnicas mais modernas de classificação de dados do SR, dentre as analisadas neste estudo. Tanto por seus números de aplicações expressivos em trabalhos acadêmicos recentes, bem como pelas datas em que seus usos foram identificados no Brasil. Estes classifica-

dores apresentam boas e ótimas acurácias em suas respectivas aplicações, o SVM com números um pouco superiores que o RF. Estes métodos se mostraram eficientes quando usados na classificação de dados hiperespectrais.

O KNN e o SAM são classificadores com uso menos recorrente, porém seus resultados de acurácia os respaldam para usos gerais no sensoriamento remoto. Junto ao MAXVER, são classificadores por medidas espectrais e de aplicação simples, podendo serem utilizados mais corriqueiramente em mapeamentos com dados multiespectrais.

Sobre o MINDIS e o ISODATA, conclui-se que tiveram resultados de acurácia mais modestos em relação aos demais classificadores elencados. Por outro lado, o MINDIS registrou mais menções em teses e dissertações, bem como suas acurácias

tiveram mais regularidade. O ISODATA, apesar de ter alcançado resultados expressivos nos trabalhos de Marinho (2011), obteve resultados pouco convincentes no trabalho de Araújo (2017), denotando que seu uso requer certos cuidados, principalmente em se tratando de classificação de dados hiperespectrais.

Durante a execução desta pesquisa as maiores dificuldades foram impostas pelas limitações das ferramentas de pesquisa da BDTD, o que, por outro lado, não tira a importância deste portal de indexação de trabalhos acadêmicos de pós-graduação stricto sensu brasileiros. Para trabalhos futuros, recomendar-se-á a utilização de outras bases acadêmicas como fontes de trabalhos para as análises bibliométricas, no âmbito de classificadores de imagens do SR.

REFERÊNCIAS

- ABDOLLAHI, A. & PRADHAN, B. Integrated technique of segmentation and classification methods with connected components analysis for road extraction from orthophoto images. *Expert Systems with Applications*, v. 176, 2021.
- ALBUQUERQUE, R.W. **Monitoramento da cobertura do solo no entorno de hidrelétricas utilizando o classificador SVM (Support Vector Machines)**. São Paulo. 2011. Dissertação (Mestrado em Engenharia de Transportes) - Escola Politécnica, Universidade de São Paulo.
- ANDREOLA, R. **Support vector machines na classificação de imagens hiperespectrais**. Porto Alegre, 2009. Tese (Doutorado), Dissertação (Mestre em Sensoriamento remoto). Universidade Federal do Rio Grande do Sul.
- ARAÚJO, E.M. **Utilização do sensoriamento remoto para detecção de diferentes alvos no entorno de reservatórios no semiárido**. Fortaleza, 2017. 159 p. Tese (Doutorado em Engenharia Agrícola), Universidade Federal do Ceará.
- BRANCO, T.; FIDALGO, E.; FRANCISCO, C.; DART, R.D.O.; PRADO, R.; BRANCO, T.L. Análise orientada a objetos para a classificação do uso e cobertura da terra do assentamento São José da Boa Morte, RJ. *Embrapa Solos-Boletim de Pesquisa e Desenvolvimento (INFOTECA-E)*, 2018.
- BREIMAN, L. Random forests. *Machine learning*, v. 45, n. 1, p. 5–32, 2001
- BRUZZONE, L. & PERSELLO, C. Recent trends in classification of remote sensing data: Active and semi supervised machine learning paradigms. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2010. *Paper...IGARSS*, 2010, p. 3720-3723.
- BUSSAB, W.O. & MORETTIN, P.A. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010. 540 p.
- CHEN, Z.; REN, J.; TANG, H.; SHI, Y.; LENG, P.; LIU, J.; HASIYUYA. Progress and perspectives on agricultural remote sensing research and applications in China. *Yaogan Xuebao/Journal of Remote Sensing*, 20(5): (2016).
- CHUBIN, B.; SOLAIMANI, K.; HABIBNEJAD ROZHAN, M.; MALEKIAN, A. Watershed classification by remote sensing indices: A fuzzy c-means clustering approach. *Journal of Mountain Science*, v. 14, n. 10, 2017.
- CHUNJING, Y.; YUEYAO, Z.; YAXUAN, Z.; LIU, H. Application of convolutional neural network in classification of high resolution agricultural remote sensing images. In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. China. 2017. *Paper...ISPRS*, v. 42, 2017.
- CONGALTON, R.G. **A review of assessing the accuracy of classifications of remotely sensed data**. *Remote Sensing of Environment*, v. 49, n. 12, p. 1671–1678, 1991.
- COSTA, D.H. **Uso de séries temporais Sentinel 1 na identificação de culturas agrícolas utilizando modelos de machine learning**. Brasília, 2020. 48 p. Dissertação (Mestrado em Geografia), Universidade de Brasília.
- DUAN, G.; ZHANG, J.; ZHANG, S. Assessment of landslide susceptibility based on multiresolution image segmentation and geological factor ratings. *International Journal of Environmental Research and Public Health*, v. 17, n. 21, p. 1-10, 2020.
- DURÁN, G.Y.B. **Estudo da resposta espectral de alvos urbanos com espectroscopia de reflectância e imagens de alta resolução espacial**. Santa Maria, 2014. 209 p. Dissertação (Mestrado) – Universidade Federal de Santa Maria, Centro de Ciências Naturais e Exatas.
- FARIA, V.G.; SILVA BARROS, P.P.; CASSIANO, C.C.; STRABELI, T.F.; FORTE, Y.A.; MAIA, C.B.; BOSQUILIA, R.W.D. Avaliação de modelos numéricos do terreno na extração automática de redes de drenagem com SIG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XVII, João Pessoa. 2015. *Anais...João Pessoa: Instituto nacional de pesquisa Espaciais*, 2015.
- FIGUEIREDO, G.C. & VIEIRA, C.A.O. Estudo do Comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XIII, Florianópolis, 2007. *Anais...Florianópolis: INPE*, 2011, p. 5755-5762.
- FLORES, P.C.M.; DAVID, L.T.; SIRINGAN, F.P. Mangrove forest cover change (1947-2018) at the river mouth section of the jaro floodway, Iloilo, Philippines. In: INDIA GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2020. *Paper... In: GARSS: IEEE*, 2020.
- GAO, J. **Digital Analysis of Remotely Sensed Imagery**. United States: McGraw-Hill Professional, 2009. 674 p.
- GHAMISI, P.; PLAZA, J.; CHEN, Y.; LI, J.; PLAZA, A.J. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, v. 5, n. 1, p. 8-32, 2017.
- HAN, S.; MENG, Q.; LIU, H.; PENG, Y.; HAN, J.; JIN, S.; LI, H. Refined land-cover classification mapping using a multi-scale transformation method from remote sensing, unmanned aerial vehicle, and field surveys in sanjiangyuan national park, china. *Journal of Applied Remote Sensing*, v. 15, n. 1, 2021.
- HASITUYA; CHEN, Z.; LI, F.; HONGMEI. Mapping plastic-mulched farmland with C-band full polarization SAR remote

- sensing data. **Remote Sensing**, v. 9, n. 12, 2017.
- IBICT, R.C. Vídeos do V Simpósio Internacional de Preservação Digital. In: SEMINÁRIO INTERNACIONAL DE PRESERVAÇÃO DIGITAL, 2021. Campinas, 5, 2021.
- JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em Recursos Terrestres**. Tradução de J. C. N. Epiphânio. São José dos Campos, SP: Parênteses, 2009. 598 p.
- KOTZ, S. & JOHNSON, N.L. **Encyclopedia of statistical sciences**. New York: John Wiley e Sons, 1983. v. 4. 352 p.
- KRUSE, F.A.; BOARDMAN, J.W.; LEFKOFF, A.B.; HEIDEBRECHT, K.B.; SHAPIRO, A.T.; BARLOON, P.J.; GOETZ, A.F.H. The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data. **Remote Sensing of Environment**, v. 44, p.145-163, 1993.
- LI, X.C.; LIU L.L.; HUANG, L.K. **Comparison of several remote sensing image classification methods based on ENVI**. Paper presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives: 2020.
- LIU, P.; DI, L.; DU, Q.; WANG, L. Remote sensing big data: theory, methods and applications. **Sensoriamento Remoto**, v. 10, n. 5, p. 711, 2018.
- MAHDIANPARI, M.; GRANGER, J.E.; MOHAMMADI-MANESH, F.; SALEHI, B.; BRISCO, B.; HOMAYOUNI, S.; LANG, M. Meta-analysis of wetland classification using remote sensing: A systematic review of a 40-year trend in north America. **Remote Sensing**, 2020.
- MARINHO, R.R. **Uso de imagens SAR orbitais em desastres naturais: mapeamento de inundações e deslizamentos de terra ocorridos em novembro de 2008 no Vale do Itajaí-SC**. São José dos Campos. 2011. 173 p. Tese (Doutorado), Instituto Nacional de Pesquisas Espaciais.
- MAXWELL, A.E.; WARNER, T.A.; FANG, F. **Implementation of machine learning classification in remote sensing: An applied review**. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2784–2817, 2018.
- MENESES, P.R. & ALMEIDA, T.D. **Introdução ao Processamento de Imagens de Sensoriamento Remoto**. Universidade Federal de Brasília, 2012. 276 p.
- NEVES, A.K. **Mineração de dados de sensoriamento remoto para detecção e classificação de áreas de pastagem na Amazônia Legal**. São José dos Campos, 2017. 77 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais.
- NÓBREGA, R.A.A. **Métodos de sensoriamento remoto no mapeamento de veredas na APA Rio Pandeiros**. Belo Horizonte. 2016. 93 p. Dissertação (Mestrado), Universidade Federal de Minas Gerais.
- NOVO, E.M.L.D.M. **Sensoriamento Remoto: Princípios e aplicações**. 3. ed. São Paulo: BLUCHER, 2008.
- OLIVEIRA, P.D.S. **Uso de aprendizagem de máquina e redes neurais convulsionais profundas para a classificação de áreas queimadas em imagens de alta resolução espacial**. 2019. Brasília. 23 p. Dissertação (Mestrado em Geografia), Universidade de Brasília.
- OLIVEIRA, S.S.T. **Explorando paralelismo em big data no processamento de séries temporais de imagens de sensoriamento remoto**. Goiânia. 2020. PhD thesis, Universidade Federal de Goiás. Instituto de Informática.
- PÁDUA, C.S. **Redes neurais artificiais auto-organizáveis na classificação não-supervisionada de imagens multiespectrais de sensoriamento remoto**. Piracicaba. 2016. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.
- PASSO, D.P. **Análise da qualidade de classificadores para identificação de alvos urbanos em imagens de alta resolução espacial: uma aplicação com as imagens do satélite Worldview II**. Brasília, 2013.106 p. Dissertação (Mestrado em Geociências), Universidade de Brasília.
- PERROCA, M.G. & GAIDZINSKI, R.P. **Avaliando a confiabilidade Inter avaliadores de um instrumento para classificação de pacientes: coeficiente kappa**. **Revista da Escola de Enfermagem - USP**, v. 37, n. 1, p. 72–80, 2003.
- RAMBO, E.M. **Avaliação de sinistros agrícolas via sensoriamento remoto orbital e aprendizado de máquina**. Cascavel, 2020. 89 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná.
- RICHARDS, J.A. & JIA, X. **Remote sensing digital image analysis**. [S.l.]: Springer, v. 3, 1999.
- SANTOS, A.B. **Combinação de Múltiplas Abordagens de Classificação para Interpretação de Imagens Hiperspectrais de Sensoriamento Remoto**. Belo Horizonte. 2013, 83 p. Dissertação (Mestrado em Geografia), Universidade Federal de Minas Gerais.
- SANTOS, A.R.; PELUZIO, T.M.O.; SAITO, N.S. **Spring 5.1.2 passo a passo: Aplicações práticas**. Alegre: Caufes, 153 p., 2010.
- SANTOS, R.L.; OHATA, A.T.; QUINTILHA, J.A. **A análise bayesiana na classificação supervisionada de imagens: Aplicação na determinação de classes de uso do solo (Mogi das Cruzes – SP)**. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11, 2003, Belo Horizonte. **Anais**, INPE, 2003, p. 2119 – 2122.
- SAQUI, D. **Um novo método Wrapper multiobjetivo para seleção de bandas de Imagens Hiperspectrais**. São Carlos. 2020. 189 p. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de São Carlos.
- SARTORI, L.R. **Métodos para extração de informações a partir de imagens multiespectrais de escalas grandes**. Presidente Prudente, 2006. Dissertação (Mestrado). Universidade Estadual Paulista.
- SCHWARTZMAN, L.F.G. **Coertura do solo em regiões de floresta tropical: estudo de caso do município de Tefé - AM**. Porto Alegre. 2017. 108 p. Dissertação (Mestrado) - Curso de Sensoriamento Remoto, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia., Universidade Federal do Rio Grande do Sul.
- SEKERTEKIN, A. & ZADBAGHER, E. Simulação da futura distribuição da temperatura da superfície terrestre e avaliação da ilha de calor urbana com base na área de superfície impermeável. **Ecological Indicators**, n. 122, p. 107230, 2021.
- SILVA JÚNIOR, C.A. **Estimativa e discriminação de áreas de soja [Glycine max L.] no estado do Paraná com dados mono e multitemporais do sensor MODIS**. Maringá. 2014. 188 p. Dissertação (Mestrado) - Curso de Agronomia, Universidade Estadual de Maringá.
- SOUZA, C.G.; CARVALHO, L.; AGUIAR, P.; ARANTES, T.B. Aprendizagem de máquina e variáveis de sensoriamento remoto para o mapeamento da cafeicultura. **Boletim de Ciências Geodésicas**, v. 22, n. 4, p. 751–773, 2016.
- SOUZA, W.W.C. **Análise e propostas de procedimentos técnicos para a elaboração de mapas de paisagem aplicados no planejamento ambiental da RMSP**. São Paulo. 2019. Tese (Doutorado em Geografia Física) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.
- THEODORIDIS, S. & KOUTROUMBAS, K. **Pattern Recognition**. Elsevir, 2ª ed. 2003.
- TURGAY, S.O., & YAŞAR, E. Evaluation of hydrocarbon occurrences of hatay with integrated methods, SE turkey. **Geomechanics and Geophysics for Geo-Energy and Geo-Resources**, v. 7, n. 3, 2021.
- WANG, L.; MA, Y.; YAN, J.; CHANG, V.; ZOMAYA, A. Y. pipsCloud: High performance cloud computing for remote sensing big data management and processing. **Future Generation Computer Systems**, v. 78, p. 353-368, 2018.

Submetido em 15 de setembro de 2021

Aceito para publicação em 13 de outubro de 2022