

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE CIÊNCIA FLORESTAL
CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL**

GABRIELA COSTA DE ALMEIDA

**INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE USO E
COBERTURA DA TERRA NO SEMIÁRIDO DE PERNAMBUCO**

**RECIFE-PE
2020**

GABRIELA COSTA DE ALMEIDA

**INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE USO E
COBERTURA DA TERRA NO SEMIÁRIDO DE PERNAMBUCO**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Florestal da Universidade Federal Rural de Pernambuco, como parte das exigências para obtenção do título de Bacharel em Engenharia Florestal.

Orientador: Emanuel Araújo Silva
Coorientadora: Giselle Lemos Moreira

**RECIFE-PE
2020**

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- A447i Almeida, Gabriela Costa
Inteligência Artificial na classificação do uso e cobertura da terra no semiárido de Pernambuco / Gabriela Costa Almeida. - 2020.
47 f. : il.
- Orientador: Emanuel Araujo Silva.
Coorientadora: Giselle Lemos Moreira.
Inclui referências e apêndice(s).
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco, Bacharelado em Engenharia Florestal, Recife, 2020.
1. Classificação supervisionada. 2. Caatinga. 3. Random Forest. 4. K-Nearest Neighbors. 5. Artificial Neural Networks. I. Silva, Emanuel Araujo, orient. II. Moreira, Giselle Lemos, coorient. III. Título

CDD 634.9

GABRIELA COSTA DE ALMEIDA

**INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO DE USO E COBERTURA DA
TERRA NO SEMIÁRIDO DE PERNAMBUCO**

Trabalho de Conclusão de Curso apresentado, ao curso de graduação, em Engenharia Florestal da Universidade Federal Rural de Pernambuco, como requisito e título de bacharel em Engenharia Florestal.

Orientador: Professor Doutor Emanuel Araújo Silva

Coorientadora: Giselle Lemos Moreira

COMISSÃO EXAMINADORA

Professor Emanuel Araújo Silva, Dr. (UFRPE)

Professor Everaldo Marques de Lima Neto, Dr. (UFRPE)

Professora Elisiane Alba, Dra. (UFRPE)

Recife, 03 de novembro de 2020

LISTA DE FIGURAS

Figura 01 — Mapa de localização do município de Petrolina, em Pernambuco.	21
Figura 02 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmo Random Forest (RF).....	26
Figura 03 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmos Artificial Neural Networks (ANN).....	27
Figura 04 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmo K-Nearest Neighbors (KNN).	28

LISTA DE TABELAS

Tabela 01 — Bandas do Satélite Landsat 8, sensores OLI e TIRS.....	14
Tabela 02 — Qualidade do mapeamento de acordo com intervalos de índice de Kappa.....	24
Tabela 03 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo Random Forest (RF).....	26
Tabela 04 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo Artificial Neural Networks (ANN).	27
Tabela 05 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo K-Nearest Neighbors (KNN).	28
Tabela 06 — Compilação dos dados de porcentagem cobertura das classes nos diferentes algoritmos utilizados.	29
Tabela 07 — Dados sobre acurácia por Validação Cruzada do tipo K-Fold e índice de Kappa em diversos algoritmos de classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio de vários algoritmos de classificação supervisionada.	30
Tabela 08 — Valores de Acurácia Global nas diferentes repetições da Validação Cruzada do tipo K-Fold nos algoritmos utilizados.	32
Tabela 09 — Ranking de classificação dos dados de Acurácia Global na Validação Cruzada do tipo K-Fold dos algoritmos utilizados por repetição.	33

LISTA DE ABREVIATURAS

ANN	<i>Artificial Neural Networks</i>
BOA	<i>Bottom Of Atmosphere</i>
CPRM	Companhia de Pesquisa de Recursos Minerais
DOS1	<i>Dark Object Subtraction 1</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
IA	Inteligência Artificial
KNN	<i>K-Nearest Neighbors</i>
LANDSAT	<i>Land Remote Sensing Satellite</i>
LCDM	<i>Landsat Data Continuity Mission</i>
MNT	Modelo Numérico do Terreno
NASA	<i>National Aeronautics and Space Administration</i>
NIR	<i>Near Infrared Spectroscopy</i>
OLI	<i>Operational Land Imager</i>
PDI	Processamento Digital de Imagem
QGIS	<i>Quantum Geographic Information System</i>
RF	<i>Random Forest</i>
SCP	<i>Semi-Automatic Classification Plugin</i>
SIG	Sistemas de Informações Geográficas
SWIR	<i>Short Wave Infrared</i>
TIRS	<i>Thermal Infrared Sensor</i>
USGS	<i>United States Geological Survey</i>

RESUMO

A Floresta Tropical Seca brasileira, conhecida como Caatinga, está presente na região nordeste do Brasil e possui características climáticas severas, com clima seco e chuvas mal distribuídas. Essas características climáticas dificultam a análise por sensoriamento remoto devido às grandes diferenças de vegetação entre os períodos seco e chuvoso. Para auxiliar a análise de sensoriamento remoto neste bioma, este trabalho tem como objetivo testar diferentes algoritmos de Inteligência Artificial por meio de classificação supervisionada e identificar padrões de uso e cobertura da terra na cidade de Petrolina, em Pernambuco. Três algoritmos foram testados: *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors* usando o software QGIS e RStudio baseado em imagens LANDSAT 8 do período seco. Foram selecionadas 20 amostras das classes: Água, Agricultura, Área Urbana, Floresta e Solo Exposto, e essas amostras serviram de base para o treinamento dos algoritmos de classificação das imagens. Dados de ocupação e avaliação de qualidade de precisão foram obtidos usando acurácia do mapeamento e índice de Kappa, respectivamente: 0,9878706 e 0,9653555 para *Random Forest*; 0,9199973 e 0,9454833 para *Artificial Neural Networks*, 0,9873741 e 0,9598640 para o *K-Nearest Neighbors*, todos considerados excelentes. Esses valores foram superiores aos encontrados nos algoritmos mais comumente utilizados, como no algoritmo de Máxima Verossimilhança. Observou-se que o uso de algoritmos de inteligência artificial pode gerar melhores resultados na classificação do uso da terra em regiões semiáridas.

Palavras-chave: Classificação supervisionada, Caatinga, *Random Forest*, *K-Nearest Neighbors*, *Artificial Neural Networks*.

ABSTRACT

The Brazilian Tropical Dry Forest, known as Caatinga, is located in Brazil's northeastern region and has severe climatic characteristics, with dry weather and poorly distributed rainfall. Those climatic characteristics make Remote sensing analysis difficult due to its large vegetation differences between the dry and rainy periods. In order to help the remote sensing analysis in this biome, this work aims to test different Artificial Intelligence algorithms through supervised classification and to identify land use and land cover patterns in the city of Petrolina, in Pernambuco. Three algorithms were tested: Random Forest, Artificial Neural Networks, and K-Nearest Neighbors using QGIS and RStudio software based on Landsat 8 images from the dry period. Twenty samples from the classes were selected: Water, Agriculture, Urban Area, Forest, and Exposed Soil, and these samples served as a basis for training the algorithms for the classification of images. Occupancy data and precision quality assessment were obtained using Mapping Accuracy and Kappa Index, respectively: 0.9878706 and 0.9653555 for Random Forest; 0.9199973 and 0.9454833 for Artificial Neural Networks, 0.9873741 and 0.9598640 for K-Nearest Neighbors, all being considered excellent. These values were higher than those found in the most commonly used algorithms, as in the Maximum Likelihood algorithm. It was observed that the use of artificial intelligence algorithms could generate better results in the classification of land use in semiarid regions.

Keywords: *Supervised classification, Caatinga, Random Forest, K-Nearest Neighbors, Artificial Neural Networks.*

SUMÁRIO

1. INTRODUÇÃO.....	9
2. OBJETIVOS	10
2.1. Objetivo Geral	10
2.2. Objetivos Específicos.....	10
3. REVISÃO DE LITERATURA	11
3.1. Caatinga	11
3.2. Sensoriamento Remoto	12
3.2.1. Imagens LANDSAT 8	13
3.3. Geoprocessamento.....	15
3.3.1. Tipos de Classificações	16
3.4. Inteligência Artificial	17
3.4.1. Algoritmos de Classificação	17
3.4.1.1. <i>Random Forest</i>	17
3.4.1.2. <i>Artificial Neural Networks</i>	18
3.4.1.3. <i>K-Nearest Neighbors</i>	20
4. MATERIAL E MÉTODOS	21
4.1. Área de Estudo.....	21
4.2. Processamento de Imagens.....	22
4.3. Seleção de Amostras.....	22
4.4. Classificação.....	23
4.5. Avaliação da Precisão do Mapeamento.....	23
5. RESULTADOS E DISCUSSÃO.....	24
6. CONCLUSÃO.....	34
7. REFERÊNCIAS.....	35
APÊNDICE A — Código utilizado na classificação	44

1. INTRODUÇÃO

A Caatinga é um bioma presente no Nordeste do Brasil e apresenta características específicas, de forma que as tecnologias de classificação utilizadas nos estudos de uso da terra ainda não estão inteiramente adaptadas para seu uso, favorecendo a desinformação e dificultando a preservação e manejo sustentável (LEAL; TABARELLI; SILVA, 2003; PRADO, 2003). Segundo Morais, Francisco e Melo (2014) seu território é um espaço único, pela singularidade de seus atributos climáticos, hidrológicos e ecológicos. Ela possui clima semiárido e a precipitação é mal distribuída, tendo períodos muito secos e outros chuvosos. Possui também vegetação muito adaptada e heterogênea. Por essas características, há dificuldade nos estudos dessa região, principalmente por meio do sensoriamento remoto, que se baseia em fornecer dados da superfície terrestre remotamente.

O sensoriamento remoto trabalha com base em dados captados pelos sensores a certa distância do objeto de interesse, como sensores presentes nos satélites que estão na órbita terrestre e capturam a energia do sol refletida. Com base nos dados extraídos desses sensores, são gerados diversos produtos que podem ser utilizados para monitoramento e planejamento de diversos estudos e ações relacionados a clima, meio ambiente, diversos outros temas.

Alguns dos produtos que o sensoriamento remoto com auxílio do geoprocessamento pode gerar são os mapas de uso e ocupação da terra com base nas imagens de satélite. Na confecção desses produtos são identificados os padrões de uso e ocupação da terra, e as classes podem ser definidas pelo usuário ou computador, a depender do método escolhido, além de possibilitarem inúmeras metodologias. A classificação de uso e cobertura da terra identifica a quantidade e a distribuição das classes no território, sendo muito útil para estudos, diagnósticos e políticas públicas.

Nesse produto, a classificação de imagens é feita com base na identificação de feições por meio dos padrões de energia eletromagnética refletida, transformada em classes de cobertura da terra (SANTOS et al., 2019). Para essa classificação, utilizam-se em seu processo dados provenientes de imagens de sensores como os satélites, por exemplo, e pode ser feita por duas formas: supervisionada e não-supervisionada. A classificação não-supervisionada é, segundo Castillejo-González et al. (2009), a categorização dos dados provenientes de uma imagem por meio do processamento informático, utilizando-se análises estatísticas. Nesse sentido, as classes são geradas a partir dessas análises e sem muita interação com o usuário. Já a classificação supervisionada é, um produto da interação do *software* com o usuário, este reconhecendo locais de classes conhecidas e utilizando-os como amostras, e, a partir deles, o *software* extrapola suas características espectrais transformadas

em classes para as áreas restantes (CASTILLEJO-GONZÁLEZ et al., 2009). Dessa forma, há uma maior quantidade de informações envolvidas no processo, e conseqüentemente, melhores resultados.

Essas tecnologias mencionadas estão dentro do sensoriamento remoto e do geoprocessamento, e com eles é possível aplicar diversos conhecimentos tecnológicos, como por exemplo a Inteligência Artificial (IA). Ela é uma área do conhecimento que tenta simular a inteligência humana em computadores, nela os computadores tomam decisões de forma independente apoiando-se em dados e podendo ser com ou sem supervisão. Existem diversos algoritmos de classificação por meio de Inteligência Artificial, como *Decision Tree*, *Random Forest*, *Support Vector Machine*, *Artificial Neural Networks*, *K-Nearest Neighbors*, *Naive Bayes*,

São escassos os estudos utilizando inteligência artificial para a classificação de uso e cobertura da terra para regiões semiáridas. Desde modo, estudos que comprovem a eficiência da inteligência artificial nessas áreas são necessários, principalmente para estudos ambientais mais robustos utilizando uma larga escala de banco de dados.

2. OBJETIVOS

2.1. Objetivo Geral

Classificar o uso e cobertura da terra em área semiárida de Pernambuco, por meio de utilização de técnicas de inteligência artificial a partir de imagens OLI do Landsat 8.

2.2. Objetivos Específicos

- Realizar o mapeamento do uso e cobertura da terra no município de Petrolina;
- Identificar os percentuais de uso e cobertura da terra por meio de imagens de satélite do Landsat 8 no semiárido de Pernambuco;
- Avaliar a eficiência dos algoritmos *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors*, baseada na classificação de imagens.

3. REVISÃO DE LITERATURA

3.1. Caatinga

A vegetação continental é o tipo de recobrimento mais abrangente e frequente e que expressa interações relevantes entre atividades antrópicas e fatores abióticos naturais (ROSENDO, 2005; LOPES et al., 2010). No centro da região Nordeste do Brasil, predomina o clima semiárido, classificado como BSh (quente e seco) segundo a classificação de Köppen, caracterizado por apresentar apenas duas estações, sendo uma curta e com chuvas em intervalos irregulares e outra de longa estiagem (seca), além da irregularidade temporal, também há irregularidade espacial na distribuição dessas chuvas. Na área há também a presença de solos rasos e altas temperaturas. Seu solo também possui características que impedem o acúmulo de água a longo prazo, além de ser raso, pedregoso e com fertilidade relativamente baixa (SOUTO et al., 2009; ALVES et al., 2017).

Devido a essas condições, há um bioma completamente adaptado, denominado Caatinga (PONZONI, 2001; AB'SABER, 2003; MARENGO, 2008; LACERDA et al., 2010; PAREYN et al., 2013), no qual a vegetação é extremamente sazonal e dinâmica, sendo comum a senescência das folhas nos períodos onde há estiagem, e o aspecto verdejante nos períodos de chuva. Em razão de sua sazonalidade climática intensa em concordância com sua morfologia e mecanismos fisiológicos adaptados para ambientes secos o IBGE (2012) dá a nomenclatura de Savana Estépica à vegetação das regiões semiáridas brasileiras e também existe a nomenclatura Floresta Tropical Seca para esse bioma (CECCON; HUANTE; RINCÓN, 2006).

A ocupação vegetal dessa área é feita de forma rarefeita, com maior densidade nas proximidades dos vales de rios (ANDRADE-LIMA, 1981; AB'SABER, 2003). Mesmo com toda dificuldade de acesso aos recursos hídricos, a Caatinga ocupa a área semiárida com maior percentual populacional e maior biodiversidade do mundo (AB'SABER, 2003; PRADO, 2003). A Caatinga é uma formação de grande evidência na Região Nordeste, e possui extensão de 844.453 km², sendo essa área equivalente a 55% da região, além de totalizar 11% do território nacional (BRASIL, 2020). Seus níveis de radiação solar são muito altos, por causa de sua posição próxima à zona Equatorial, apresentando em consequência disso as baixas precipitações e alta evapotranspiração. (SAMPAIO, 1995)

O bioma é caracterizado por apresentar nas suas paisagens grande diversidade, tendo comportamentos diferentes nos meses de chuva, onde a paisagem muda rapidamente, sua vegetação renasce e cobre-se de folhas, bem como o seu aspecto verde na vegetação que parecia morta, com

isso, o local fica belo e coberto de vida. A vegetação também varia muito sua fisionomia, apresentando diversos portes e densidades (SOUZA, 2008).

Essa vegetação que cobre a região possui muitos padrões morfológicos, esses padrões dependem da fisionomia e das condições climáticas que ela é submetida, e pode apresentar os tipos de formação de caatinga arbustiva, densa ou aberta e perdendo a sua folhagem no período de estiagem até o início nas chuvas, onde retornará ao aspecto verdejante (ALVES, 2017; SOUSA et al., 2008)

A Caatinga apresenta, segundo Alves (2017), sublime riqueza biológica e múltiplas espécies. Além da Caatinga, existem outros biomas similares fora do Brasil, encontrados no México, Índia, Austrália, sudeste da África, Bolívia, Equador e Peru (BLACKIE et al., 2014).

3.2. Sensoriamento Remoto

O Sensoriamento Remoto constitui-se numa importante técnica para análises ambientais, monitoramentos da dinâmica da vegetação e planejamentos na utilização dos recursos naturais, e para dar suporte às pesquisas sendo recorrente a utilização de imagens de satélites com diferentes resoluções espaciais, temporais e radiométricas (CARVALHO JUNIOR, et al., 2008; FERREIRA et al., 2013). Entretanto, devido à ausência de soluções para determinadas utilizações, são necessárias novas técnicas no tratamento de dados de satélite para trabalhos com alta heterogeneidade na vegetação (MALDONADO e CARVALHO, 2003), e fatores como o a heterogeneidade das estruturas, relevo do terreno, ângulo de visada e tipo de solo são os principais obstáculos, pois interferem a resposta espectral da vegetação (OLIVEIRA; CHAVES; LIMA, 2009).

O sensoriamento remoto é, segundo Martins (2015), um conjunto de técnicas para obtenção de informações sobre objetos sem contato físico. Iniciou na invenção da câmera fotográfica que acabou sendo utilizada como instrumento para a realização de fotos aéreas, que eram usadas principalmente no meio militar, onde eram adaptadas para serem mais leves e automáticas. Ele obtém informações sobre objetos, áreas ou fenômenos por meio de dados espectrais sem contato direto com o objeto de estudo por meio de sensores (LILLESAND; KIEFER; CHIPMAN, 2015). Muitos desses sensores utilizam a energia eletromagnética emitida pelo sol e que interage com os alvos nas superfícies.

Sua evolução teve um maior destaque por volta de 1970, onde foram lançados satélites na órbita terrestre provenientes de grandes investimentos e energia. Foram projetos interdisciplinares envolvendo principalmente ramos da biociência, física, geociência, físico-química, química,

computação, mecânica entre outros. Esse tipo de tecnologia é utilizada até hoje, eles trabalham orbitando e capturando imagens continuamente durante toda sua vida útil (FIGUEIREDO, 2005).

Hoje o sensoriamento remoto é baseado principalmente nos satélites orbitais e suas imagens, sendo hoje os mais famosos: LANDSAT, SPOT, CBERS, IKONOS e QUICKBIRD que fazem monitoramento e levantamento dos recursos naturais e o satélite NOAA que faz o levantamento de dados meteorológicos para estudos climáticos e atmosféricos, mas também podendo ser usado no sensoriamento remoto (FIGUEIREDO, 2005).

A energia refletida pela atmosfera pode representar uma fonte de erro em razão de distorções radiométricas, ou seja, alterações nos valores obtidos pelo sensor. Isso acontece pela presença de vapor d'água, aerossóis e demais gases atmosféricos, que provocam distúrbios nos sinais captados, sendo as principais alterações: diminuição da faixa de valores digitais registrados pelo sensor, diminuição do contraste entre superfícies e alteração do brilho de cada ponto da imagem (GANEM, 2017)

A reflectância é uma variável utilizada constantemente no sensoriamento remoto. É uma razão entre a radiância (energia emitida) pela irradiância (energia recebida), ou seja, é o balanço de energia em superfície, é adimensional e expressa em porcentagem (MENESES e ALMEIDA, 2013). Cada objeto possui um comportamento espectral diferente por suas características físicas, químicas e biológicas, portando uma assinatura espectral característica. (ROCHA, 2018)

O sensoriamento remoto tem como um de seus principais objetivos identificar e distinguir a composição de diferentes materiais superficiais, podendo ser tipos de vegetação, padrões de uso da terra, rochas e outros, sendo possível por meio das suas assinaturas espectrais (CRÓSTA, 2002)

3.2.1. Imagens LANDSAT 8

O programa LANDSAT surgiu em meados dos anos 60, como parte do Programa de Levantamento de Recursos Terrestres da pela *National Aeronautics and Space Administration* (NASA). Seu primeiro satélite se chamava *Earth Resources Technology Satellite – 1* (ERTS-1), sendo posteriormente renomeado para LANDSAT 1 (*Land Remote Sensing Satellite 1*), e opera desde 1972, disponibilizando imagens gratuitas desde 1973 (NASA, 2020 e USGS, 2016).

Os satélites do programa LANDSAT registram, continuamente, dados multiespectrais, que capturam imagens da superfície terrestre por completo e seus produtos podem ser utilizados em diversas pesquisas e segmentos por sua boa resolução espacial. Por apresentarem essa boa resolução

espacial, fácil acesso e gratuidade, é amplamente utilizado no mapeamento e monitoramento terrestre, sendo importante fonte de dados (NOVO, 2010).

O LANDSAT 8, inicialmente denominado *Landsat Data Continuity Mission*, (LDCM), foi lançado em 11 de fevereiro de 2013 e foi desenvolvido como uma colaboração entre a NASA e o *United States Geological Survey* (USGS) possuindo os sensores: o *Operational Land Imager* (OLI) e o *Thermal Infrared Sensor* (TIRS), que fornecem cobertura sazonal da massa de terra global com uma resolução espacial de 30 metros (visível, NIR, SWIR); 100 metros (térmico); e 15 metros (pancromático). (NASA, 2020)

O sensor OLI coleta dados para bandas espectrais visíveis, infravermelho próximo e infravermelho de ondas curtas, bem como uma banda pancromática, já o sensor TIRS coleta dados para duas bandas espectrais mais estreitas na região térmica e a precisão cartográfica de 12 m ou melhor (incluindo compensação para os efeitos do terreno) como pode ser visto na Tabela 01. As imagens do Landsat 8 possuem resolução temporal de 16 dias, resolução radiométrica de 16 bits e o tamanho da cena é de 170 Km por 183 Km (NASA, 2020).

Tabela 01 — Bandas do Satélite Landsat 8, sensores OLI e TIRS

Número	Nome	Comprimento de Onda (µm)	Resolução Espacial(m)
Banda 1	Azul Costeiro/Aerosol	0,435 - 0,451	30
Banda 2	Azul	0,452 - 0,512	30
Banda 3	Verde	0,533 - 0,590	30
Banda 4	Vermelho	0,636 - 0,673	30
Banda 5	Infravermelho Próximo (NIR)	0,851 - 0,879	30
Banda 6	Infravermelho Médio (SWIR-1)	1,566 - 1,651	30
Banda 7	Infravermelho Médio (SWIR-2)	2,107 - 2,294	30
Banda 8	Pancromática	0,503 - 0,676	15
Banda 9	Cirrus	1,363 - 1,384	30
Banda 10	Infravermelho Termal (TIRS-1)	10,60 - 11,19	100
Banda 11	Infravermelho Termal (TIRS-2)	11,50 - 12,51	100

Fonte: Adaptado de NASA (2020).

Crósta (2002) fala que a resolução é um termo muito usado em sensoriamento remoto, podendo se referir à vários parâmetros:

- Resolução Espacial – É a resolução que trata da capacidade do sistema sensor em "enxergar" alvos na superfície terrestre, o que significa que quanto menor o objeto identificado, maior a resolução espacial;
- Resolução Radiométrica – É o número de níveis de cinza que o sensor consegue expressar em dados, e quanto mais níveis, maior é a resolução radiométrica;

- Resolução Espectral – É o número de bandas espectrais que o sensor possui, além da largura do intervalo de comprimento de onda que as bandas cobrem. Quanto maior quantidade de bandas e menores larguras nos intervalos, melhor a resolução espectral;
- Resolução Temporal – É o tempo em que o sensor demora para coletar dados do mesmo ponto, ou seja, fazer uma revisita ao local.

3.3. Geoprocessamento

Atualmente na gestão do território as ações de planejamento, ordenação ou monitoramento do espaço devem ser complementadas por análises dos componentes do meio físico-biótico, da ocupação humana e inter-relacionamentos. O geoprocessamento é uma ferramenta útil nesse processo, pois fornece a integração de dados que ele pode requerer (CÂMARA e MEDEIROS, 1998). O geoprocessamento é definido por Varella (2004) como um conjunto de tecnologias que estão voltadas para coleta e tratamento de informações espaciais e servem para um objetivo específico, além de também serem executadas por sistemas específicos, conhecidos como como Sistemas de Informações Geográficas (SIG).

A Cartografia clássica está em constante mudança para a digital pelo constante avanço de tecnologias, permitindo análises que combinem ferramentas de análise espacial, processamento de imagens, geoestatística e modelagem numérica de terreno, permitindo a disponibilização mais eficaz de geoinformações, principalmente por meio de Sistemas de Informações Geográficas (CÂMARA e MEDEIROS, 1998; MEDEIROS; ARAGÃO; GOMES, 2010). Varella (2004) fala que o objetivo dos SIGs é processar informações espaciais, desta forma deve ser capaz de criar com base no real abstrações digitais, armazenar e manejar dados de forma eficiente, identificando bem o relacionamento entre as variáveis espaciais, e facilitando a criação de relatórios e mapas para a compreensão dessas relações.

Nos SIGs a integração de informações é proveniente de diferentes bases de dados como cartografia, imagens de satélite e de MNT (Modelo Numérico de Terreno). Com matérias desse tipo, é possível mapear e classificar terras para otimização de uso, sendo assim importante para aplicação em trabalhos como os de mapeamentos temáticos, análises de risco, classificações técnicas e gerenciamento ambiental (FRANCISCO et al., 2012). Com os SIGs, pode-se gerar mapas, que são importantes ferramentas na gestão do território, dão suporte para as decisões e estão diretamente ligados à etapas de planejamento, ordenamento e uso eficaz dos recursos da terra, podendo servir como instrumentos e ferramentas para registro e a análise das paisagens (LIMA et al., 2007).

O uso de geotecnologias, como o SIG, cresce cada vez mais, possibilitando inovações em processos e modelos que ajudam no tratamento de grande quantidade de dados, além de facilitarem a manipulação de informações indispensáveis aos mapeamentos (CARVALHO et al., 2009). O conceito de geotecnologia é explicado por Silva (2003), que diz que as geotecnologias são as artes e técnicas de estudo da superfície da terra, adaptadas quanto às informações às necessidades dos meios físicos, químicos e biológicos. Algumas ferramentas das geotecnologias são Sensoriamento Remoto, o Processamento Digital de Imagens (PDI), a Geoestatística e os SIGs.

3.3.1. Tipos de Classificações

Crósta (2002) explica que pelo sensoriamento remoto existem duas abordagens diferentes na classificação de imagens multiespectrais, sendo uma denominada de classificação supervisionada, pois ela baseia-se em uma identificação prévia de amostras das classes desejadas que foram selecionadas pelo usuário e as demais áreas serão categorizadas pelo computador, baseando a escolha dos pixels na regra estatística pré-estabelecida. O autor também explica que a outra abordagem, chamada de classificação não-supervisionada, usa como base regras estatísticas pré-estabelecidas, fazendo com que o computador escolha as classes que serão identificadas e quais os pixels pertencentes a cada uma.

Thenkabail e Wu (2012) e Zhong et al. (2016) também falam sobre os algoritmos de classificação supervisionada, os quais relacionam um conjunto de dados de amostras (refletâncias espectrais, índices, etc.) com classes de uso/cobertura da terra, e um treinamento é feito a partir dessa relação, e esses dados geralmente consistem em referências da terra obtidas durante a aquisição de imagens de satélite. Segundo Rosendo (2005) é cada vez mais comum o estudo das modificações do uso da terra e do estado da vegetação com a utilização de imagens de satélite de resolução moderada a global. O uso de classificadores diferentes podem gerar diferentes resultados, pois cada classificador pode ser melhor na identificação de um conjunto de classes que outro (REIS et al., 2017).

No processo de classificação supervisionada é necessário um conhecimento prévio da área de estudo, de preferência locais específicos com observações de campo durante a passagem do satélite para maior acurácia, esses locais são conhecidos como verdade terrestre e podem ser usadas como padrão de comparação para distinção das classes por meio dos pixels (CRÓSTA, 2002).

3.4. Inteligência Artificial

Existem algumas técnicas que podem ser utilizadas para auxiliar na classificação, entre elas está a Inteligência Artificial (IA), que baseia seus estudos na simulação computacional de aspectos da inteligência humana (NISHIDA, 1998). O uso dessas tecnologias tem se mostrado uma alternativa promissora por seu caráter inovador na análise dos dados (NISHIDA e BASTOS, 1998) e oferece modelos de apoio a decisões baseado em dados reais, mesmo que incompletos (SELLITTO, 2002). Essa tecnologia procura resolver problemas do mesmo modo que um ser humano, imitando mecanismos naturais por meio de mecanismos tecnológicos, multidisciplinar, originado da computação, matemática, cibernética, engenharia e psicologia (ARARIBOIA, 1998; NIKOLOPOULOS, 1997), e tem por objetivo gerar sistemas com comportamento inteligente e desempenho equivalente ou superior de um especialista humano (NIKOLOPOULOS, 1997).

O principal atrativo é sua capacidade de aprender por meio de exemplos e generalizar as informações aprendidas, mesmo em conjunto reduzido de exemplos, gerando resposta coerente, demonstra que vai além do mapeamento de relações de entrada e saída (BUCENE e RODRIGUES, 2004). Dentro da Inteligência artificial, há o aprendizado de máquinas, que segundo Scott (2018) são os programas computacionais capazes de processar conjuntos de dados enormes, de muitas fontes e com padrões discernidos entre várias variáveis não selecionadas, possuindo então seus algoritmos, que podem ser vários, como por exemplo: *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors*.

3.4.1. Algoritmos de Classificação

3.4.1.1. *Random Forest*

Dentro da inteligência artificial existe o *Random Forest* (RF), um algoritmo de classificação de conjuntos não paramétrico, ou seja, a distribuição estatística dos dados a serem classificados para as diferentes classes utilizadas não é conhecido. Ele baseia-se em muitas árvores de regressão (BREIMAN, 2001; PETROPOULOS; ARVANITIS; SIGRIMIS, 2012). Seu algoritmo vem sendo utilizado em estudos de classificação de cobertura da terra e apresentando melhor precisão em relação a outros classificadores como: Máxima Verossimilhança, *Artificial Neural Networks* e Árvores de Classificação Única (CASTILLEJO-GONZÁLEZ et al., 2009; LÖW et al., 2015; LEBOURGEOIS et al., 2017; PARENTE et al., 2017; PEÑA-BARRAGÁN et al., 2011; ZHONG; GONG; BIGING, 2014). Suas vantagens são: implementação eficiente em grandes conjuntos de dados usando variáveis

categóricas como preditores; estimação da importância das variáveis individuais na classificação produzindo a probabilidade de classe e eficiência contra sobreajuste (*overfitting*) mesmo para conjuntos de dados ligeiramente desbalanceados (GISLASON; BENEDIKTSSON; SVEINSSON, 2006; KHOSHGOFTAAR; GOLAWALA; VAN HULSE, 2007).

As técnicas mais amplamente utilizadas nesse classificador são o *boosting* e *bagging*. Segundo Breiman (1996), em *boosting*, as árvores sucessivas não dependem de árvores anteriores, cada uma é construída de forma independente utilizando uma amostra do conjunto de dados pré-definido, repetitivamente, e para a previsão é utilizada uma simples maioria de votos, ou seja, a decisão a árvore criada é um modelo individual que aprende com o modelo anterior no reprocessamento. Breiman (1996) também fala que a ideia de *bagging* é a criação de árvores de regressão que se repetem sucessivamente, e a partir desse reprocessamento são obtidos pesos diferenciados para cada variável, gerando uma estrutura de árvore de decisão com pesos. No algoritmo, são utilizadas árvores calculadas independentemente que somadas, geram uma única com pesos que dão a importância de cada variável analisada para a classificação. À medida que o número de preditores aumenta, a precisão também aumenta, mas em certo ponto há um limite que, se ultrapassado, pode gerar um sobreajuste, pois o algoritmo é tão específico à situação que deixa de ser adequado em outros casos. Encontrar o número ideal de preditores irá gerar maior precisão.

3.4.1.2. *Artificial Neural Networks*

O algoritmo *Artificial Neural Networks* (ANN) é baseado no funcionamento do cérebro humano, segundo Queiroz, Rodrigues e Gómez (2004), as Redes Neurais passam por processos de aprendizado por meio das conexões entre neurônios.

Existem diversas classificações dentro das ANNs (por forma de aprendizado, pelas suas características, ou pela sua estrutura), na classificação por forma de aprendizado, o aprendizado supervisionado apresenta-se sucessivamente à conjunto de dados referentes à entradas e saídas e nesse processo a rede ajusta os pesos das conexões entre os elementos do processamento, até que o erro atinja o valor mínimo desejado. No aprendizado por reforço ao invés de receber as saídas corretas para a rede para cada treinamento individual, se fornece um valor informando se a saída está correta ou não (MEDEIROS, 1999).

No aprendizado não-supervisionado o conjunto de dados é analisado pela rede, e em seguida ela determina as propriedades do conjunto de dados, aprendendo a refletir em sua saída utilizando padrões, regularidades e correlações para agrupar esses dados em classes (MEDEIROS, 1999).

Muitas configurações são variações de números de neurônios por camada, ou de tipo de funções de ativação dos mesmos, até mesmo número de camadas ocultas ou das conexões entre eles, onde cada modificação afeta o tempo e eficiência do treinamento, e também há outro fator que interfere: os algoritmos utilizados, parâmetros e coeficientes (BINOTI; BINOTI; LEITE et al., 2014).

No processo de escolha do número de neurônios a serem utilizados um número excessivo pode gerar memorização dos dados de treinamento, chamado de *overfitting*, que prejudica sua aplicação em outro conjunto de dados similar. Também há chances de acontecer o efeito reverso quando há escolha de um número menor de neurônios a serem utilizados, que podem não ser suficientes para a realização da tarefa causando um ajuste pouco eficiente dos dados, chamado de *underfitting* (BINOTI; BINOTI; LEITE et al., 2014).

O processamento dos dados se dá por obtenção e armazenamento de conhecimento. Segundo Mazhar et al. (2013), são apresentados padrões à rede em uma camada de entrada denominada X_i , que passam por camadas intermediárias responsáveis pelo processamento de grande parte dos dados. Nesse processamento, os dados inseridos na camada de entrada são multiplicados pelo peso W_{ji} , e também passam por processo de polarização denominada Θ_j que ajusta o erro residual. O valor ainda passa por uma função de ativação ou transferência e seu valor é normalizado em um intervalo fechado, onde os dados processados saem por uma camada Y_j . Quanto mais neurônios são inseridos nas camadas intermediárias, maior se torna a precisão do algoritmo.

Seu cálculo se dá pelas equações:

$$I_j = W_{ji}X_i + \Theta_j \quad (1)$$

$$Y_i = f(I_j) = f(W_{ji}X_i + \Theta_j) \quad (2)$$

Onde:

X_i = Dados de entrada;

W_{ij} = Peso ajustado;

Θ_j = Erro residual;

Y_i = Dados de Saída.

3.4.1.3. *K-Nearest Neighbors*

O algoritmo de classificação denominado *K-Nearest Neighbors* (K-NN) ou *K-Nearest Neighbors* (KNN) define uma função de distância entre um conjunto de dados e é construída uma previsão considerando todas as possíveis distâncias entre todos os pontos do conjunto, sendo todos classificados. O classificador observa em qual classe a o objeto mais se encaixa e atribui um valor. Utiliza principalmente a função euclidiana como base (XU et al., 2013).

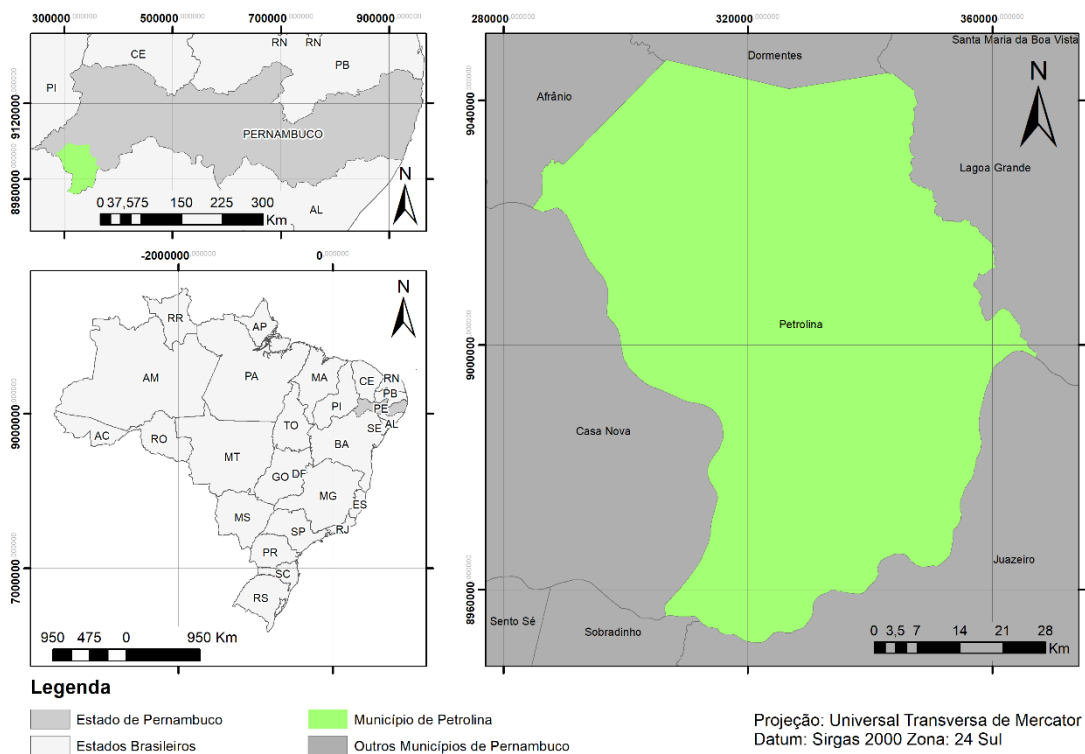
Ele funciona com base na analogia com o conjunto de dados de calibração fornecido pelo usuário, encontrando um grupo de k amostras que estão com valores mais próximos das amostras desconhecidas. A partir dessas k amostras, o algoritmo classifica as amostras desconhecidas por meio de um cálculo que identifica a classe que está com valor de pixel mais próximo do valor da amostra desconhecida. (THANH NOI e KAPPAS, 2018). O cálculo para identificar os valores de amostras mais próximas pode ser feito por diversos parâmetros, como as distâncias Euclidiana, Manhattan, Mahalanobis, Minkowski, e outros (ÖZSAKABAŞI, 2008).

4. MATERIAL E MÉTODOS

4.1. Área de Estudo

O estudo foi desenvolvido no município de Petrolina, na região de desenvolvimento do Sertão do São Francisco, em Pernambuco (PERNAMBUCO, 2013) como pode ser visto na Figura 01. O município se localiza nas coordenadas centrais $09^{\circ} 23' 20''$ S e $040^{\circ} 30' 11''$ W. A região apresenta vegetação típica do bioma Caatinga, predominantemente Hiperxerófila com alguns trechos de Floresta Caducifólia, clima predominantemente Semiárido quente e seco, com variações de temperatura entre 24°C e 28°C , sendo o mês mais frio do ano julho e o mais quente o de novembro (CPRM, 2019). A precipitação média anual é de 560mm, mais concentrada nos meses de janeiro a abril. O município possui uma área de 4518,29 km², o que representa 4,81% do território do estado. Ele está inserido na Depressão Sertaneja, com relevo suave-ondulado e cortado por vales estreitos; faz parte da Macrobacia do Rio São Francisco, do Grupo de Bacias de Pequenos Rios Interiores e da Bacia Hidrográfica do Riacho do Pontal (IBGE, 2018; CPRM, 2019; INPE, 2019).

Figura 01 — Mapa de localização do município de Petrolina, em Pernambuco.



Fonte: A autora.

4.2. Processamento de Imagens

Para o estudo, foram utilizadas imagens do satélite Landsat 8 oriundas do sensor OLI (*Operational Land Image*), com resolução média espacial de 30 metros, adquiridas no site do *United States Geological Survey*, nas órbitas/pontos: 217/66 e 217/67 referentes à data 29 de outubro de 2016 com recobrimento de nuvens menor que 10%.

Antes da análise, foram feitas correções atmosféricas nas imagens por meio do *plugin Semi-Automatic Classification*, no *software QGIS 3.4* (QGIS Development Team, 2019), estas foram reprojatadas para o Sistema de Referência SIRGAS 2000, em projeção UTM na zona 24S.

Com o auxílio do *Semi-Automatic Classification Plugin* (SCP) do QGIS, desenvolvido por Congedo (2014), obteve-se a reflectância na base da superfície (BOA, sigla do inglês *Bottom Of Atmosphere*) por meio de um processo onde foi realizada a conversão dos valores em níveis digitais (ND) pelo método DOS1 (*Dark Object Subtraction 1*), eliminando a interferência da atmosfera.

Em seguida foi criado um mosaico de bandas similares das duas imagens para criação de uma única imagem que possibilite o recobrimento total da área de estudo. Na imagem gerada foi realizado o recorte da área do município, por meio de arquivos em formato *shapefile* obtidos gratuitamente no site do IBGE (2017).

4.3. Seleção de Amostras

No processo de classificação de uso e ocupação da terra, foram feitas seleções de amostras em vetor para a classificação, sendo no mínimo 20 amostras por classe analisada. Sendo elas:

- Água - Área com presença de corpos hídricos;
- Floresta - Áreas de cobertura arbustiva-arbórea;
- Agropecuária - Áreas de cultivos de monoculturas e com vegetação rala e espaçada ou criação de animais;
- Solo Exposto - Área sem vegetação e afloramentos rochosos;
- Área Urbana - Áreas com presença de construções civis.

Essa seleção de amostras foi realizada com delimitação aleatória de áreas de interesse das classes, visualizadas por meio do *Plugin Quick Map Services*, por meio de imagens de satélite atualizadas do Google, onde é possível visualizar em alta resolução imagens em baixa escala, além de comparar com outros satélites.

4.4. Classificação

Esses conjuntos de amostras das categorias são a base para a estruturação da árvore de decisão dos algoritmos utilizados: *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors*. As amostras homogêneas escolhidas de todas as categorias e os dados de reflectância das bandas de 2 a 7 foram utilizados para o treinamento dos classificadores. Os algoritmos *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors* foram processados no *software* RStudio (2019), foram utilizados os pacotes: *raster* (HIJMANS, 2019), *caret* (KUHN, 2019), *rgdal* (BIVAND, 2019), *randomForest* (LIAW, 2018), *kermlab* (KARATZOGLOU, 2019), *tsutils* (KOURENTZES, 2019) e *nnet* (RIPLEY, 2019).

4.5. Avaliação da Precisão do Mapeamento

A precisão dos classificadores foi analisada com base no índice Kappa e na Acurácia Global do mapeamento, ambos sendo calculados por meio do *software* RStudio (RStudio Team, 2019), que gerou em seus resultados por meio dos algoritmos utilizados. O Índice de Kappa (Equação 3) avaliou a concordância entre a verdade terrestre e o mapa temático. Nesse cálculo incluiu-se todos os elementos da matriz de erro, com seu resultado variando de 0 a 1, sendo valores mais próximos de 1 mais precisos (MOREIRA, 2005), como descrito na Tabela 02.

Esse cálculo se dá pela equação:

$$K = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} * X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} * X_{+i})} \quad (3)$$

Em que:

K= índice de exatidão Kappa;

r = número de linhas da matriz;

X_{ii} = número de observações na linha i e coluna i;

X_{i+} = totais marginais da linha i;

X_{+i} = totais marginais da coluna i;

N = número total de observações.

Tabela 02 — Qualidade do mapeamento de acordo com intervalos de índice de Kappa.

Valor de Kappa	Qualidade do mapeamento
< 0,00	Péssima
0,01 a 0,20	Ruim
0,21 a 0,40	Razoável
0,41 a 0,60	Boa
0,61 a 0,80	Muito Boa
0,81 a 1,00	Excelente

Fonte: Landis e Koch 1977, citados por Moreira 2005.

A forma de validação utilizada foi a Validação Cruzada, do tipo K-Fold, onde os dados foram separados em grupos de teste e de treinamento, com 30 repetições (k), gerando em cada repetição um modelo novo pela variação dos conjuntos de treinamento e teste. No fim, é calculada uma média dos resultados (MONTAÑO et al., 2017). Como descrito por Thiersch (1997) conforme citado por Alba (2019), também foi montado um ranking das acurácias globais com pesos, sendo 1 o algoritmo mais eficiente e 3 o menos eficiente. Após esse ranking, os valores passaram pelo teste de Friedman e Nemenyi com nível de significância a 5% para avaliar as diferenças estatísticas.

5. RESULTADOS E DISCUSSÃO

Após as análises em *Software RStudio*, foram geradas imagens *raster* em formato GeoTIFF, e delas foram criados mapas para facilitar a análise visual. Após isso, foram extraídas informações referentes às áreas de cada classe nos algoritmos de classificação.

Foi observado que nas classificações por *Artificial Neural Networks* e *K-Nearest Neighbors* houve uma maior identificação da classe Floresta, com 53,52% e 52,14% respectivamente, do que na classificação por *Random Forest*, com 47,52%. Embora seja a situação oposta na classe Agropecuária, onde o algoritmo *Random Forest* obteve maior cobertura da classe com 41,26%, e *Artificial Neural Networks* e *K-Nearest Neighbors* com respectivamente 34,46% e 35,75%.

A classe Água apresentou pouca mudança, sendo identificada a cobertura de 1,06% no algoritmo *Random Forest*, 1,09% no algoritmo *Artificial Neural Networks* e 1,10% em *K-Nearest Neighbors*. A classe Área Urbana também variou pouco, sendo a cobertura mínima de 2,14% em *Artificial Neural Networks*, 2,22% em *K-Nearest Neighbors* e máxima de 2,43%, em *Random Forest*.

A classe de Solo Exposto obteve resultados muito parecidos em *K-Nearest Neighbors* e *Artificial Neural Networks*, com 8,79% e 8,77% respectivamente, já o algoritmo *Random Forest* obteve uma leve diminuição para 7,71%. As classes de cobertura não identificadas foram mínimas e de mesma porcentagem para todos os algoritmos, sendo esse valor 0,02%.

Alguns problemas foram identificados no mapeamento, como a ocorrência de alguns pontos da classe Área Urbana espalhados ao longo da área de estudo, mesmo sem evidências de ocupação real. Isso ocorreu em todas as classificações, e quando verificou-se a ocupação atual, foi descoberto Solo Exposto. Apesar da confusão dos classificadores, a área de classificação incorreta foi mínima.

Galletti e Myint (2014) efetuaram mapeamentos em regiões áridas no Arizona, e informaram que é inevitável obter confusões no mapeamento devido as similaridades espectrais, pois as regiões apresentam áreas agrícolas exploradas que podem ser confundidas com regiões expostas, além de áreas com pouca ocupação urbana que podem ser confundidas com áreas expostas. Sugerem também que a classificação multi-temporal pode ser uma alternativa para a identificação mais precisa das classes.

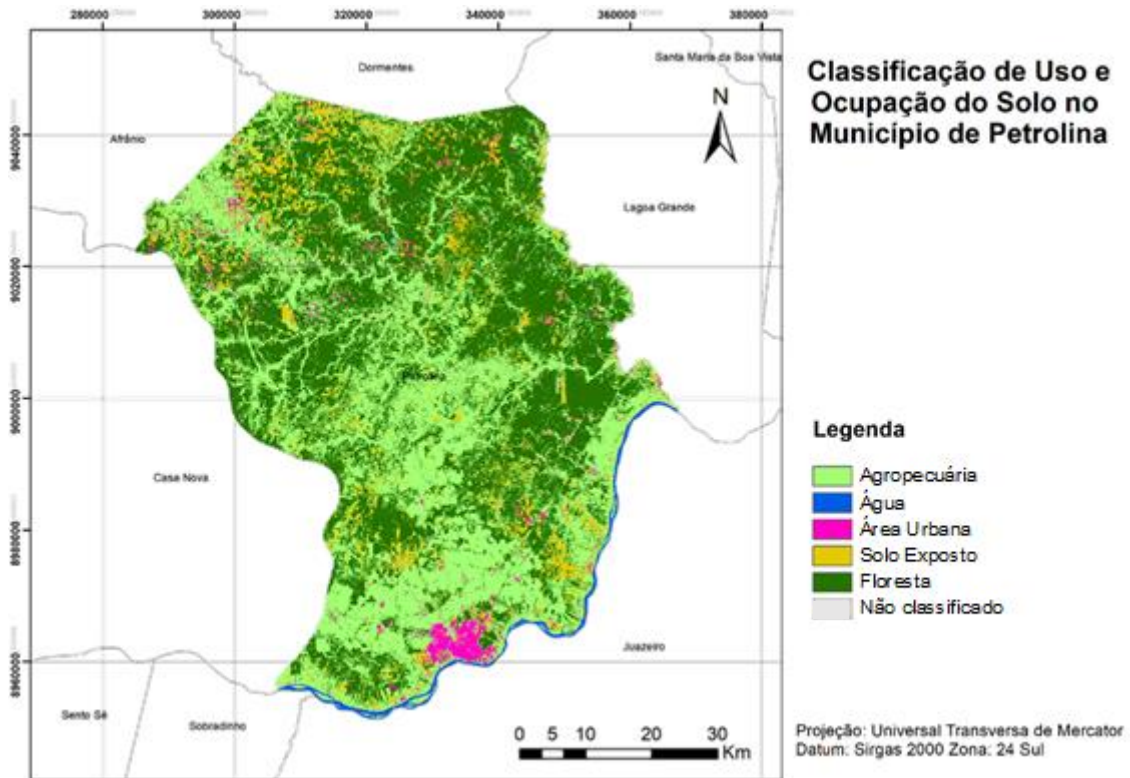
Sano et al. (2007) classificaram o bioma Cerrado, e tiveram dificuldades por causa de confusões de classes pela similaridade espectral dos alvos. Indicam também que a sazonalidade do bioma é um fator de grande influência no mapeamento do bioma, pois dificulta a identificação de vários aspectos.

Essa confusão foi identificada no trabalho de Rossini-Penteado (2007), que avaliou a área urbana de Ubatuba, em SP. Nesse estudo a classe Solo Exposto foi confundida com diversas outras classes como: Asfalto, por meio da areia úmida presente na praia que possui espectro similar. E concreto/amianto e cerâmica também pelo mesmo motivo, causando classificações de solo exposto nos telhados e de telhados em solos expostos.

O trabalho de Ribeiro, Fonseca e Kux (2011) que avaliou a área próxima ao Rodoanel Mário Covas, na Região Metropolitana de São Paulo também apresentou erro similar entre as categorias Solo Exposto e Telha Cerâmica, dificultando a separação da real área de solo exposto na pesquisa. Uma justificativa utilizada foi de que as telhas cerâmicas possuem composições semelhantes, e por isso espectros similares.

Pode-se observar na Figura 02 e na Tabela 03 que em comparação com os outros algoritmos, há a presença de mais áreas na cor verde-clara, correspondente à classe Agropecuária, o que confirma visualmente os dados obtidos acima.

Figura 02 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmo *Random Forest* (RF).



Fonte: A autora.

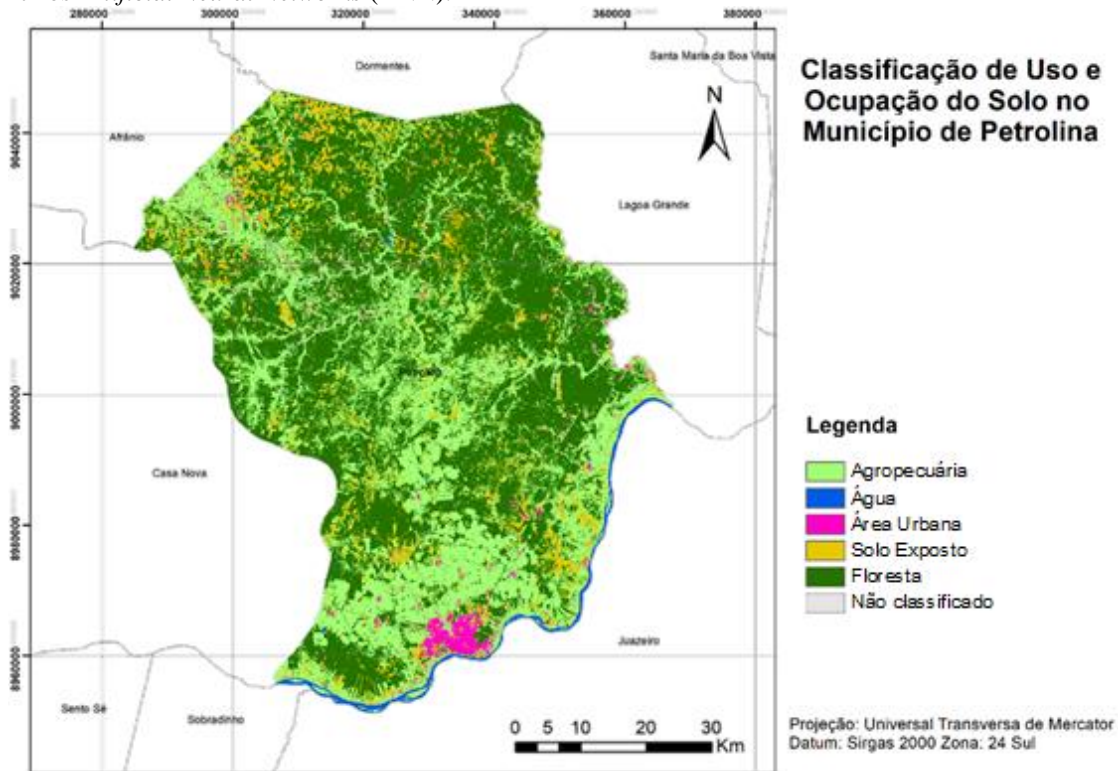
Tabela 03 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo *Random Forest* (RF).

Classe	Recobrimento (%)	Área (Km ²)
Agropecuária	41,26%	1864,37
Água	1,06%	47,80
Área Urbana	2,43%	109,81
Solo Exposto	7,71%	348,40
Floresta	47,52%	2147,22
Não Classificado	0,02%	0,69
TOTAL:		4518,29

Fonte: A autora.

Já no algoritmo *Artificial Neural Networks*, houve uma maior quantidade da classe Floresta identificada, bem como menor quantidade da classe Agropecuária, podendo ser notado na área central do mapa na Figura 03, bem como nos dados da Tabela 04.

Figura 03 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmos *Artificial Neural Networks* (ANN).



Fonte: A autora.

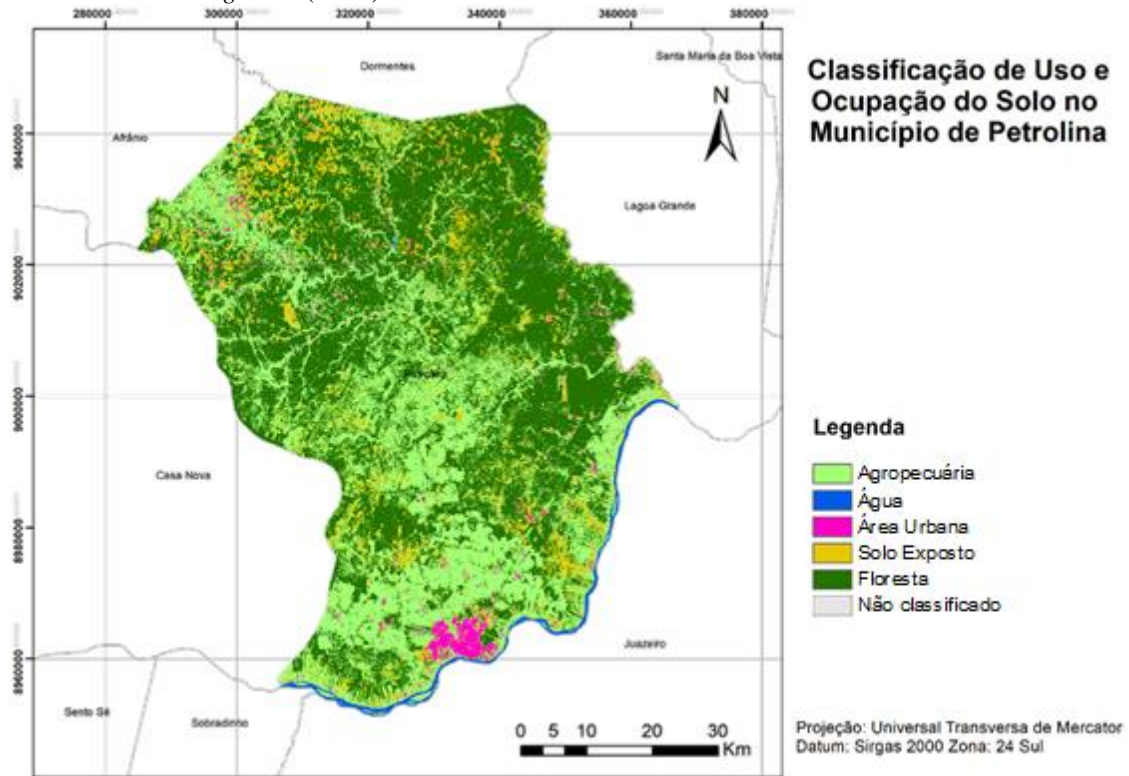
Tabela 04 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo *Artificial Neural Networks* (ANN).

Classe	Recobrimento (%)	Área (Km ²)
Agropecuária	34,46%	1557,10
Água	1,09%	49,44
Área Urbana	2,14%	96,49
Solo Exposto	8,77%	396,43
Floresta	53,52%	2418,15
Não Classificado	0,02%	0,69
TOTAL:		4518,29

Fonte: A autora.

Já o algoritmo *Artificial Neural Networks* apresentou valores intermediários para as duas classes, possuindo apenas uma identificação da classe Solo Exposto um pouco maior que os dois métodos anteriores, como pode ser visto na Figura 04 e na Tabela 05.

Figura 04 — Mapa de Classificação de Uso e Ocupação da terra no Município de Petrolina, em Pernambuco, por meio do algoritmo *K-Nearest Neighbors* (KNN).



Fonte: A autora.

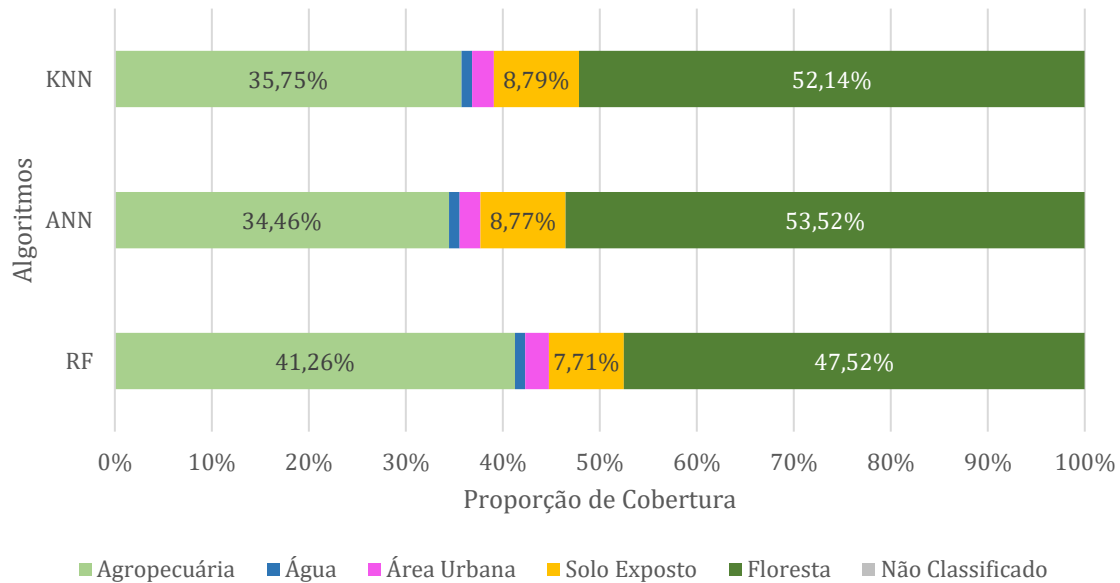
Tabela 05 — Dados da classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio do algoritmo *K-Nearest Neighbors* (KNN).

Classe	Recobrimento (%)	Área (Km ²)
Agropecuária	35,75%	1615,06
Água	1,10%	49,72
Área Urbana	2,22%	100,10
Solo Exposto	8,79%	396,96
Floresta	52,14%	2355,76
Não Classificado	0,02%	0,69
TOTAL:		4518,29

Fonte: A autora.

Os dados da classificação foram reunidos, conforme observado na Tabela 06 e Gráfico 01, onde é possível comparar os valores de cobertura em cada classe de cada algoritmo testado.

Gráfico 01 — Dados de cobertura das classificações de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio de diferentes algoritmos de classificação supervisionada.



Fonte: A autora.

Tabela 06 — Compilação dos dados de porcentagem cobertura das classes nos diferentes algoritmos utilizados.

Classe	RF	RNA	ANN
Agropecuária	41,26%	34,46%	35,75%
Água	1,06%	1,09%	1,10%
Área Urbana	2,43%	2,14%	2,22%
Solo Exposto	7,71%	8,77%	8,79%
Floresta	47,52%	53,52%	52,14%
Não Classificado	0,02%	0,02%	0,02%

Fonte: A autora.

Após a classificação, foram analisados a acurácia de cada classificação e o índice de Kappa, que utiliza uma matriz de confusão para processamento dos dados e resultado do índice, sendo os resultados mais próximos de 1 mais precisos, os valores obtidos podem ser visualizados na Tabela 07.

Tabela 07 — Dados sobre acurácia por Validação Cruzada do tipo K-Fold e índice de Kappa em diversos algoritmos de classificação de uso e ocupação da terra do município de Petrolina, em Pernambuco, por meio de vários algoritmos de classificação supervisionada.

Algoritmo	Acurácia	Kappa	Qualidade do Mapeamento
<i>Random Forest</i>	0,9878706	0,9653555	Excelente
<i>Artificial Neural Networks</i>	0,9199973	0,9454833	Excelente
<i>K-Nearest Neighbors</i>	0,9873741	0,9598640	Excelente

Fonte: A Autora.

Conforme observado, os valores de Acurácia e Kappa para todos os algoritmos de classificação foram considerados excelentes em razão de sua qualidade, sendo os valores mais altos no algoritmo *Random Forest*.

No trabalho de Zanetti, Braga e Duarte (2017), que comparou os algoritmos de classificação supervisionada: Paralelepípedo, Distância Euclidiana, Máxima Verossimilhança e *Artificial Neural Networks* em Viçosa, MG, o índice de Kappa encontrado foi de 0,894 em *Artificial Neural Networks*, um resultado um pouco inferior ao encontrado no presente estudo. Outro trabalho que avaliou a eficiência dos classificadores *Artificial Neural Networks* e Máxima Verossimilhança em Viçosa, MG, foi desenvolvido por Chagas et al. (2008) que encontraram um valor ligeiramente inferior, com índice de Kappa de 0,80 em *Artificial Neural Networks*.

Demarchi, Sartori e Zimback (2011) que avaliaram algoritmos de classificação na Sub-Bacia do Córrego das Três Barras, obtiveram um índice de Kappa de 0,4652 no algoritmo *K-Nearest Neighbors*. Apesar do valor mais baixo, a classificação ainda é considerada boa.

Fernandes et al. (2015) também classificou uma região de Caatinga em Sergipe nos anos 1992, 2003 e 2013 por meio do algoritmo Máxima Verossimilhança, e obteve o índice de Kappa em 0,60 e exatidão global em 82,5% para a classificação do ano de 2003. Os índices dos outros anos não foram mencionados.

Coelho et al. (2014) classificou uma região semiárida em Pernambuco inserida na microrregião do Pajeú, também utilizando a Máxima Verossimilhança em três anos: 1991, 2000, 2010. Os índices Kappa obtidos foram acima de 0,90 para as imagens de 1991 e 2000. Para a imagem de 2010 o índice Kappa ficou próximo de 0,80.

Nery et al. (2013) classificou uma região na bacia do Rio Vieira em Minas Gerais por meio de três algoritmos: Máxima Verossimilhança, Máxima Verossimilhança (*Iterated Conditional Modes*) e Distância Mínima Euclidiana, e obteve os Índices de Kappa respectivamente 0,74, 0,75 e 0,61, sendo considerados muito bons.

Souza, Teixeira, Silva (2009) utilizaram o método Máquina de Vetores Suporte para a classificação de imagens da região de Caatinga em Iguatu, no Ceará, e obtiveram como resultado o Índice de Kappa de 0,65 que é classificado como bom.

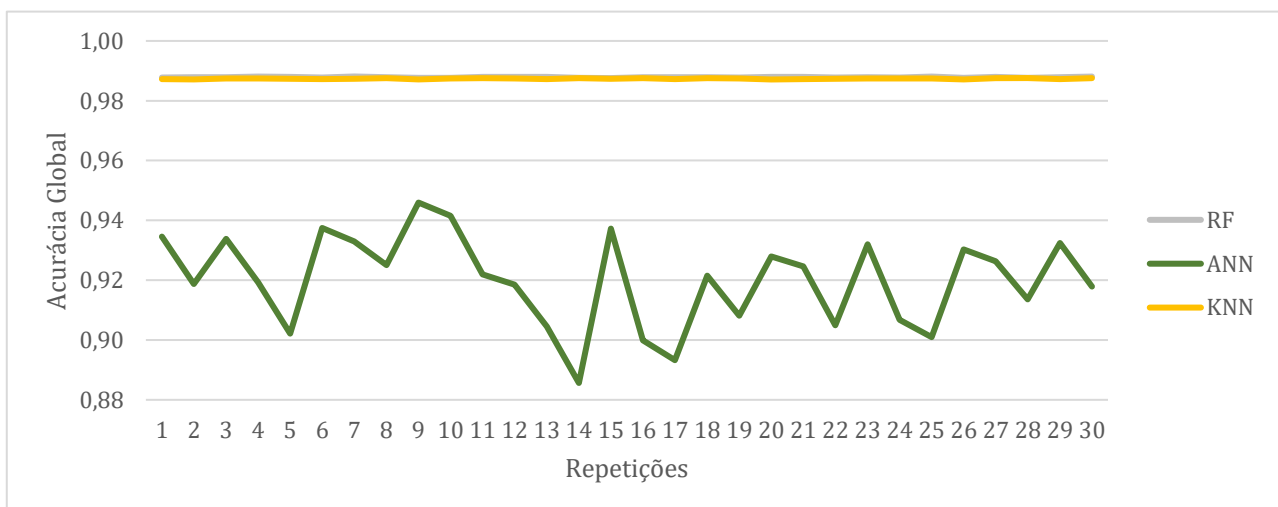
Reis et al. (2017) estudaram diversos métodos, como Máxima Verossimilhança, Máquina de Vetores Suporte, Árvore de Decisão e *K-Nearest Neighbors* no município de Belterra, PA. Obtiveram os índices de Kappa respectivamente de 0,699, 0,701, 0,702 e 0,702, esses valores são considerados muito bons no índice.

Alguns estudos desenvolvidos com diferentes classificadores obtiveram resultados inferiores ao encontrados neste estudo. O mais utilizado na maioria dos trabalhos é o algoritmo Máxima Verossimilhança, apresenta em sua maioria valores considerados bons ou muito bons nos Índices de Kappa obtidos, porém os algoritmos testados nesse trabalho por meio do *software* RStudio atingiram níveis considerados excelentes.

Para a Acurácia Global, foram obtidos os dados da verificação cruzada, com $k = 30$. A partir dessas repetições, foram criadas tabelas com os dados (Tabela 08) e um ranking de classificações dos algoritmos em cada repetição, com base nos conjuntos de treinamento e de teste.

Observou-se que os algoritmos *Random Forest* e *K-Nearest Neighbors* obtiveram resultados muito semelhantes, com seus valores oscilando muito pouco e bem próximos, já o algoritmo *Artificial Neural Networks* obteve menor Acurácia Global em todas as repetições, além de apresentar uma oscilação no valor, conforme Gráfico 02. Pode-se notar que no Ranking da Tabela 09 não há variações nas posições dos classificadores.

Gráfico 02 — Valores de Acurácia Global nas diferentes repetições da verificação cruzada tipo K-Fold dos diferentes algoritmos utilizados.



Fonte: A Autora.

Tabela 08 — Valores de Acurácia Global nas diferentes repetições da Validação Cruzada do tipo K-Fold nos algoritmos utilizados.

Repetição	RF	ANN	KNN
1	0,98777240	0,93456780	0,98724250
2	0,98791890	0,91877330	0,98713290
3	0,98786400	0,93381860	0,98744360
4	0,98804680	0,91924960	0,98740710
5	0,98799180	0,90212220	0,98735210
6	0,98780900	0,93750990	0,98727910
7	0,98811970	0,93292300	0,98735210
8	0,98784570	0,92508310	0,98751670
9	0,98766280	0,94595450	0,98713290
10	0,98762630	0,94151260	0,98740710
11	0,98793710	0,92195650	0,98758970
12	0,98795530	0,91857140	0,98744370
13	0,98799180	0,90449580	0,98726080
14	0,98766290	0,88563810	0,98753490
15	0,98758970	0,93723590	0,98738860
16	0,98782730	0,89986480	0,98753500
17	0,98791880	0,89321640	0,98727900
18	0,98788230	0,92151650	0,98755320
19	0,98780900	0,90811700	0,98740710
20	0,98795530	0,92795210	0,98711460
21	0,98797370	0,92465840	0,98726090
22	0,98779090	0,90493810	0,98729740
23	0,98791880	0,93200790	0,98744350
24	0,98777260	0,90671510	0,98744360
25	0,98808340	0,90097160	0,98740700
26	0,98764450	0,93029100	0,98715090
27	0,98801030	0,92634430	0,98751670
28	0,98771770	0,91358600	0,98758980
29	0,98791880	0,93240950	0,98720600
30	0,98810140	0,91791830	0,98753480
MÉDIA	0,98787063	0,91999731	0,98737411

Fonte: A Autora.

Tabela 09 — Ranking de classificação dos dados de Acurácia Global na Validação Cruzada do tipo K-Fold dos algoritmos utilizados por repetição.

Repetição	RF	ANN	KNN
1	1	3	2
2	1	3	2
3	1	3	2
4	1	3	2
5	1	3	2
6	1	3	2
7	1	3	2
8	1	3	2
9	1	3	2
10	1	3	2
11	1	3	2
12	1	3	2
13	1	3	2
14	1	3	2
15	1	3	2
16	1	3	2
17	1	3	2
18	1	3	2
19	1	3	2
20	1	3	2
21	1	3	2
22	1	3	2
23	1	3	2
24	1	3	2
25	1	3	2
26	1	3	2
27	1	3	2
28	1	3	2
29	1	3	2
30	1	3	2
MÉDIA	1	3	2

Fonte: A Autora.

Por meio do teste de Friedman e Nemenyi, os algoritmos Random Forest e *Nearest Neighbors* não diferem estatisticamente, já o algoritmo *Artificial Neural Networks* difere, apresentando um resultado inferior.

6. CONCLUSÃO

O uso e ocupação da terra em regiões semiáridas como a Caatinga é de difícil classificação por sua sazonalidade, causando diversos erros principalmente por similaridades espectrais dos alvos. Classes como Floresta e Solo Exposto podem ser confundidas em épocas mais secas, subestimando sua área de cobertura, bem como confusões em outras categorias como Área Urbana, Solo Exposto e outros, por seus materiais de origem similar que causam confusão espectral. Essa confusão é mais comum em imagens de período seco, como as utilizadas neste trabalho.

Mesmo com os problemas observados, a classificação obteve grande similaridade entre os valores da porcentagem de cobertura das classes para os diferentes métodos, e também foram obtidos ótimos resultados no uso desses algoritmos de Inteligência Artificial, como foi visto por meio do Índice de Kappa de 0,9653555, 0,9454833 e 0,9598640 respectivamente para os algoritmos *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors*, sendo considerados excelentes, bem como os valores de Acurácia Global de 0,9878706, 0,9199973 e 0,9873741 respectivamente para os algoritmos *Random Forest*, *Artificial Neural Networks* e *K-Nearest Neighbors*, também considerados excelentes.

Não houve diferença estatística para os algoritmos *Random Forest* e *K-Nearest Neighbors*, tendo eles melhores resultados em comparação com o algoritmo *Artificial Neural Networks*. Também apresentaram desempenho superior a outros classificadores utilizados mais amplamente, como por exemplo o Máxima Verossimilhança, podendo ser uma alternativa fácil e eficiente para futuros estudos da região.

7. REFERÊNCIAS

- AB'SABER, A. N. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. São Paulo. Ateliê Editorial. 2003.
- ALBA, E. **Influência e análise da cobertura florestal na modificação do albedo com o uso de Inteligência Artificial e Sensoriamento Remoto**. 2019. 121 f. Tese (Doutorado em Engenharia Florestal). Universidade Federal de Santa Maria, Santa Maria, 2019.
- ALVES, A. R. et al. Nutrientes na biomassa aérea e na serapilheira em áreas de caatinga em Floresta, PE. **Pesquisa Florestal Brasileira**, v. 37, n. 92, p.413-420, 29 dez. 2017. Embrapa Florestas. Disponível em: <https://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/1060>. Acesso em: 17 ago. 2020
- ANDRADE-LIMA, D. The Caatinga Dominion. **Revista Brasileira de Botânica**. v.4: p.149-153. 1981
- ARARIBOIA, G. **Inteligência Artificial**. Rio de Janeiro: Ed. LTC, 1988.
- BINOTI, D. H. B.; BINOTI, M. L. M. S; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Revista Ciência da Madeira (Brazilian Journal of Wood Science)**, v. 5, n. 1, p. 10-12, 2014. Disponível em: <https://periodicos.ufpel.edu.br/ojs2/index.php/cienciadamadeira/article/view/4067>. Acesso em: 13 ago. 2020
- BIVAND, R. S. **RGDAL: Bindings for the 'Geospatial' Data Abstraction Library**. R package version 1.5-18. Disponível em: <https://cran.r-project.org/web/packages/rgdal/index.html>. Acesso em: 09 set. 2019.
- BLACKIE, R. et al. **As florestas tropicais secas: O estado de conhecimento global e recomendações para futuras pesquisas**. 2014. Disponível em: https://www.cifor.org/publications/pdf_files/WPapers/DPBlackie1401P.pdf Acesso em: 20 ago. 2020
- BRASIL. **Ministério do Meio Ambiente. Biomas: Caatinga**. Brasília, DF, 2020. Disponível em: <https://www.mma.gov.br/biomas/caatinga>. Acesso em: 21 set. 2020
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- BREIMAN, L. Bagging predictors. **Machine learning**, v. 24, n. 2, p. 123-140, 1996.
- BUCENE, L. C.; RODRIGUES, L. H. A. Utilização de redes neurais artificiais para avaliação de produtividade do solo, visando classificação de terras para irrigação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 8, n. 2-3, p. 326-329, 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662004000200025&lng=en&enrm=iso. Acesso em 20 set. 2020.

CÂMARA, G.; MEDEIROS, J. S. **Geoprocessamento para projetos ambientais**. 2.ed. INPE. São José dos Campos, 194p, 1998. Disponível em: http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/. Acesso em: 15 jun. 2020.

CARVALHO JÚNIOR, O. A. et al. Mapeamento da vegetação na floresta atlântica usando o classificador de árvore de decisão para integrar dados de sensoriamento remoto e modelo digital de terreno. **Revista Brasileira de Geofísica**, v. 26, n. 3, p. 331-345, 2008.

CARVALHO, C. C. N; ROCHA, W. F; UCHA, J. M. Mapa digital de solos: Uma proposta metodológica usando inferência fuzzy. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.1, p.46–55, 2009.

CASTILLEJO-GONZÁLEZ, I. L. et al. Object-and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. **Computers and Electronics in Agriculture**, v. 68, n. 2, p. 207-215, 2009.

CECCON, E.; HUANTE, P.; RINCÓN, E. Abiotic factors influencing tropical dry forests regeneration. **Brazilian archives of Biology and Technology**, v. 49, n. 2, p. 305-312, 2006.

CHAGAS, C. S. et al. Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 3, p. 319-327, 2009.

COELHO, V. H. R. et al. Dinâmica do uso e ocupação do solo em uma bacia hidrográfica do semiárido brasileiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 1, p. 64-72, 2014.

CONGEDO, L. Land cover classification of cropland: A tutorial using the semi-automatic classification plugin for QGIS. **Rome [ITL]: Sapienza University**, p. 1-25, 2014. Disponível em: <https://www.directionsmag.com/article/1428>. Acesso em: 15 set. 2020.

CPRM. **Projeto cadastro de fontes de abastecimento por água subterrânea - Estado de Pernambuco**: Diagnóstico do município de Petrolina, 2019. Disponível em: http://rigeo.cprm.gov.br/xmlui/bitstream/handle/doc/16562/Rel_Petrolina.pdf?sequence=1. Acesso em: 05 jan. 2019.

CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Instituto de Geociências. Departamento de Metalogênese e Geoquímica. Campinas – SP. 2002.

DEMARCHI, J. C.; SARTORI, A. A C.; ZIMBACK, C. R. L. **Métodos de classificação de imagens orbitais para o mapeamento do uso do solo**: estudo de caso na Sub-Bacia do Córrego das Três Barras. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 15, p. 2644-2651, 2011.

FERNANDES, M. R. M. et al. Mudanças do uso e de cobertura da terra na região semiárida de Sergipe. **Floresta e Ambiente**, v. 22, n. 4, p. 472-482, 2015.

FERREIRA, H. S. et al. Aplicação de técnicas de sensoriamento remoto para monitorar o comportamento da vegetação de Caatinga frente às condições climáticas no município de Petrolina - PE. Foz do Iguaçu. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto**, 2013. Disponível em: <http://marte2.sid.inpe.br/rep/dpi.inpe.br/marte2/2013/05.29.00.52.39>. Acesso em: 27 dez. 2018.

FIGUEIREDO, D. **Conceitos básicos de sensoriamento remoto**. Companhia Nacional de Abastecimento-CONAB. Brasília-DF, 2005.

FRANCISCO, P. R. M. Classificação e mapeamento das terras para mecanização do Estado da Paraíba utilizando sistemas de informações geográficas. 122f. 2010. **Dissertação** (Mestrado em Manejo de Solo e Água). Centro de Ciências Agrárias. Universidade Federal da Paraíba, Areia.

FRANCISCO, P. R. M. et al. Mapeamento da Caatinga com Uso de Geotecnologia e Análise da Umidade Antecedente em Bacia Hidrográfica. **Revista Brasileira de Geografia Física**. v. 3, p. 676-693. 2012. Disponível em: https://www.researchgate.net/publication/303920280_Mapeamento_da_Caatinga_com_Uso_de_Geotecnologia_e_Analise_da_Umidade_Antecedente_em_Bacia_Hidrografica. Acesso em: 10 set. 2020.

GALLETTI, C. S.; MYINT, S. W. Land-use mapping in a mixed urban-agricultural arid landscape using object-based image analysis: A case study from Maricopa, Arizona. **Remote Sensing**, v. 6, n. 7, p. 6089-6110, 2014.

GANEM, K. A. **Classificação da cobertura do solo na Caatinga a partir de imagens do Landsat-8 e da ferramenta Google Earth Engine: uma comparação entre dados com e sem correção atmosférica**. 2017. Disponível em: <https://repositorio.unb.br/handle/10482/23501>. Acesso em: 03 set. 2020.

GISLASON, P. O. BENEDIKTSSON, J. A. SVEINSSON, J. R. Random forests for land cover classification. **Pattern Recognition Letters**, v. 27, n. 4, p. 294-300, 2006.

HIJMANS, R. J. **RASTER**: Geographic data analysis and modeling. R package version 3.3-13. Disponível em: <https://cran.r-project.org/web/packages/raster/index.html>. Acesso em 09 set 2019.

IBGE. **Downloads: Geociências**. Disponível em: http://servicodados.ibge.gov.br/Download/Download.ashx?u=geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2017/Brasil/BR/br_municipios.zip. Acesso em: 03 jan. 2019.

IBGE. **Cidades: Petrolina**. Disponível em: <https://cidades.ibge.gov.br/brasil/pe/petrolina/panorama>. Acesso em: 21 dez. 2018.

IBGE. **Manual técnico da vegetação brasileira**. 2.ed. Rio de Janeiro: IBGE, 2012. 271p. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv63011.pdf>. Acesso em: 16 set. 2020.

- INPE. **Estação de Petrolina - Climatologia Local**. Disponível em: http://sonda.ccst.inpe.br/estacoes/petrolina_clima.html. Acesso em: 03 jan. 2019.
- KARATZOGLU, A. **KERNLAB**: Kernel-Based Machine Learning Lab. R package version 0.9-29. Disponível em: <https://cran.r-project.org/web/packages/kernlab/index.html>. Acesso em: 09 set. 2020.
- KHOSHGOFTAAR, T. M.; GOLAWALA, M.; VAN HULSE, J; et al. **An empirical study of learning from imbalanced data using random forest**: Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE international conference on. IEEE, 2007. p. 310-317. IEEE. <http://dx.doi.org/10.1109/ictai.2007.46>.
- KOURENTZES, N. **TSUTILS**: Time Series Exploration, Modelling and Forecasting. R package version 0.9.2. Disponível em: <https://cran.r-project.org/web/packages/tsutils/index.html>. Acesso em: 09 set. 2020
- KUHN, M. **CARET**: Classification and Regression Training. R package version 6.0-81. Disponível em: <https://cran.r-project.org/web/packages/caret/index.html>. Acesso em 09 set. 2019.
- LACERDA, F. F. et al. Extremos e variabilidade climática no Nordeste brasileiro e em Pernambuco. **Mudanças climáticas e impactos ambientais**. Ed. Universitária da UFPE, Recife, p. 1-23, 2010. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/894335/1/MudancasClimaticaspag123.pdf>. Acesso em: 02 jan. 2019.
- LEAL, I. R.; TABARELLI, M.; SILVA, J. M. C. Ecologia e conservação da Caatinga: uma introdução ao desafio, p. XII-XVI. **Ecologia e conservação da Caatinga**. Recife, Editora Universitária da UFPE, 2ª Edição, 822p, 2005. Disponível em: https://s3.amazonaws.com/academia.edu.documents/40548222/5_livro_ecologia_e_conservao_da_caatinga_203.pdf. Acesso em: 26 dez. 2018.
- LEBOURGEOIS, V. et al. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). **Remote Sensing**, v. 9, n. 3, p. 259, 2017. <https://www.mdpi.com/2072-4292/9/3/259/pdf>. Acesso em: 01 jan. 2019.
- LIAW, A. **RANDOMFOREST**: Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6-14. Disponível em: <https://cran.r-project.org/web/packages/randomForest/index.html>. Acesso em: 09 set. 2019.
- LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote Sensing and Image Interpretation**. 7a ed. Hoboken: Wiley, 2015.
- LIMA, D. F. B.; REMPEL, C.; ECKHARDT, R. R. Análise Ambiental da Bacia Hidrográfica do Rio Taquari - Proposta de Zoneamento Ambiental. **Revista Geografia**, v. 16, n. 1, 2007.

LOPES, H. L. et al. Análise de índices de vegetação na bacia do rio Brígida, sertão do estado de Pernambuco. **Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, v. 3, p. 01-08, 2010. Disponível em: https://www3.ufpe.br/cgtg/SIMGEOIII/IIISIMGEO_CD/artigos/FotoeSR/SR_e_PDI/A_144.pdf Acesso em: 28 dez. 2018.

LÖW, F. et al. Mapping abandoned agricultural land in Kyzyl-Orda, Kazakhstan using satellite remote sensing. **Applied Geography**, v. 62, p. 377-390, 2015.

MARTINS, G. **Sensoriamento Remoto Aplicado ao Meio Ambiente**. Notas de Aula, 2015.

MARENGO, J. A. Vulnerabilidade, impactos e adaptação à mudança do clima no semi-árido do Brasil. **Parcerias estratégicas**, v. 13, n. 27, p. 149-176, 2010. Disponível em: http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/viewFile/329/323. Acesso em: 23 dez. 2018.

MALDONADO, F. D. CARVALHO, V. C. Indicador de degradação a escala sub-regional para detecção de mudanças na cobertura das terras com sensoriamento remoto. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 11, p. 1353-1356, 2003.

MAZHAR, F. et al. On using neural networks in UAV structural design for CFD data fitting and classification. **Aerospace Science and Technology**, v. 30, n. 1, p. 210-225, 2013. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1270963813001454> Acesso em: 10 nov. 2019.

MEDEIROS, J. S. **Bancos de dados geográficos e redes neurais artificiais: tecnologias de apoio à gestão do território**. 1999. Tese de Doutorado. Universidade de São Paulo. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8135/tde-25072001-095526/publico/TeseJSM.pdf>. Acesso em:

MEDEIROS, C.N.; ARAGÃO, M. C. A.; GOMES, D. D. M. Ceará em Mapas Interativo – sistema de informações geográficas na Internet utilizando software livre. **Anais XXIV Congresso Brasileiro de Cartografia**, 2010, Aracaju – SE.

MENESES, P. R.; ALMEIDA, T. de. **Introdução ao processamento de imagens de sensoriamento remoto**. Universidade de Brasília, Brasília, 2012.

MONTAÑO, R. A. N. R. et al. Artificial Intelligence Models to Estimate Biomass of Tropical Forest Trees. **Polytech. Open Libr. Int. Bull. Inf. Technol. Sci.**, v. 56, p. 29-37, 2017.

MORAIS, L. G. B. L; FRANCISCO, P. R. M; MELO, J. A. B. Análise da cobertura vegetal das terras de região semiárida com o uso de geotecnologias. **Polêm!ca**, v. 13, n. 3, p. 1345-1363, 2014.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. UFV, 2005.

NASA. **National Aeronautics and Space Administration: History**. 2020. Disponível em: <https://landsat.gsfc.nasa.gov/about/history/>. Acesso em: 10 set. 2020.

NERY, C. V. M. et al. Avaliação das técnicas de classificação Maxver, Maxver-ICM e Distância Mínima Euclidiana de acordo com Índice Kappa. **Revista brasileira de geografia física**, v. 6, n. 2, p. 320-328, 2013.

NIKOLOPOULOS, C. **Expert systems: introduction to first and second generation and hybrid knowledge based systems**. Marcel Dekker, Inc., 1997.

NISHIDA, W. **Uma rede neural artificial para classificação de imagens multiespectrais de sensoriamento remoto**. 101 p. (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 1998. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/77542/138042.pdf>. Acesso em: 02 jan. 2019.

NISHIDA, W.; BASTOS, L. C. Classificação de imagens de sensoriamento remoto utilizando uma rede neural artificial com função de base radial. **Anais IX Simpósio Brasileiro de Sensoriamento Remoto**, Santos, Brasil, p. 11-18, 1998. Disponível em: http://marte.sid.inpe.br/col/sid.inpe.br/deise/1999/02.11.11.58/doc/8_122p.pdf. Acesso em 03 já. 2019.

NOVO, E. L. M. **Sensoriamento remoto: princípios e aplicações**. 4. ed. São Paulo: Blucher, 2010. 388 p.

OLIVEIRA, W. M.; CHAVES, I. B.; LIMA, E. R. V. Índices espectrais de vegetação de caatinga em um neossolo litólico do semiárido paraibano. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 14, p. 2103-2110, 2009. Disponível em: <http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.17.00.32/doc/2103-2110.pdf>. Acesso em: 24 dez. 2018.

ÖZSAKABAŞI, F. **Classification of forest areas by k nearest neighbor method**. 2008. Dissertação (Mestrado em Tecnologias de Informação Geodética e Geográfica) - Technologies Department, Middle East Technical University. Antalya, Turquia. 2008.

PARENTE, L. et al. Monitoring the brazilian pasturelands: A new mapping approach based on the landsat 8 spectral and temporal domains. **International Journal of Applied Earth Observation and Geoinformation**, v. 62, p. 135-143, 2017.

PAREYN, F. et al. Cuidando da Caatinga. **Associação de Plantas do Nordeste (APNE) e Royal Botanic Gardens, KEW**, 2013. Disponível em: https://www.researchgate.net/profile/William_Milliken3/publication/266317232_Cuidando_da_Caatinga/links/542c383e0cf277d58e8c231e/Cuidando-da-Caatinga.pdf. Acesso em: 02 jan. 2019.

PEÑA-BARRAGÁN, J. M. et al. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. **Remote Sensing of Environment**, v. 115, n. 6, p. 1301-1316, 2011.

PERNAMBUCO. **Perfil Municipal de Petrolina**. CONDEPE/FIDEM. Recife, 2013

PETROPOULOS, G. P.; ARVANITIS, K.; SIGRIMIS, N. Hyperion Hyperspectral Imagery Analysis Combined with Machine Learning Classifiers for Land Use/Cover Mapping. **Expert Systems with Applications**, 39, p. 3800-3809, 2012.

PONZONI, F. J. Comportamento espectral da vegetação. **Sensoriamento remoto: reflectância de alvos naturais**, v. 1, p. 157-199, 2001.

PRADO, D. E. **As caatingas da América do Sul**. **Ecologia e conservação da Caatinga**, v. 2, p. 3-74, 2003.

QGIS Development Team, (2018). **QGIS Geographic Information System**. Open Source Geospatial Foundation Project. Disponível em: <http://www.qgis.org/>. Acesso em: 12 dez. 2018.

QUEIROZ, R. B; RODRIGUES, A. G.; GÓMEZ, A. T. Estudo comparativo entre as técnicas máxima verossimilhança gaussiana e redes neurais na classificação de imagens IR-MSS CBERS 1. I **WorkComp Sul**, v. 1, 2004. Disponível em: <http://www.academia.edu/download/6744262/2307.pdf>. Acesso em: 24 out. 2019.

R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2019. Disponível em: <http://www.r-project.org/>. Acesso em: 02 set. 2019.

REIS, M. S. et al. Análise das incertezas envolvidas em classificação multi-legendas da cobertura da terra com suporte de simulação Monte Carlo. **Revista Brasileira de Cartografia**, v. 69, n. 9, 2017. Disponível em: <http://www.seer.ufu.br/index.php/revistabrasileiracartografia/article/download/44090/23315/0>. Acesso em: 12 out. 2019.

RIBEIRO, B. M. G.; FONSECA, L. M. G.; KUX, H. J. H. Mapeamento da cobertura do solo urbano utilizando imagens Worldview-Ii e o sistema Interimage. **Revista Brasileira de Cartografia**, v. 63, 11.

RIPLEY, B. **NNET: Feed-Forward Neural Networks and Multinomial Log-Linear Models**. R package version 7.3-14. Disponível em: <https://cran.r-project.org/web/packages/nnet/index.html>. Acesso em: 09 set. 2019.

ROSENDO, J. S. Índices de vegetação e monitoramento do uso do solo e cobertura vegetal na bacia do rio Araguari-MG-utilizando dados do sensor Modis. **Uberlândia: Universidade Federal de Uberlândia**, 2005. Disponível em: http://ecologia.ib.usp.br/lepac/bie5759/jussara_santos.pdf. Acesso em: 28 dez. 2018.

ROSSINI-PENTEADO, D. et al. Classificação orientada por regiões em imagem IKONOS para a identificação e análise da cobertura do solo urbano de Ubatuba (SP). **Anais XII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, SC, 2007.

- RStudio Team. **RStudio: Integrated Development for R**, 2019. RStudio, Boston, Massachusetts. Disponível em: <http://www.rstudio.com/>. Acesso em: 10 set. 2019.
- SAMPAIO, E. V. S. B. Overview of the Brazilian caatinga. **Seasonally dry tropical forests**. (Eds SH Bullock, HA Mooney, E Medina) pp. 35–63. 1995.
- SANTOS, A. S. R. M. et al. Métodos de Classificação Supervisionada Aplicados no Uso e Ocupação do Solo no Município de Presidente Médice–RO. **Biodiversidade**, v. 18, n. 1, 2019. Disponível em: <http://www.periodicoscientificos.ufmt.br/ojs/index.php/biodiversidade/article/view/8242>. Acesso em: 24 set. 2019.
- SANO, E. E. et al. Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical savanna. **International Journal of Remote Sensing**, v.28, p.2739-2752, 2007.
- SCOTT, I. A. **Machine learning and evidence-based medicine**. 2018.
- SELLITTO, M. A. **Inteligência artificial: uma aplicação em uma indústria de processo contínuo**. Gestão e Produção, v. 9, n. 3, p. 363-376, 2002.
- SILVA, A. B. **Sistemas de informações georeferenciadas: conceitos e fundamentos**. Campinas: Editora da UNICAMP, 2003.
- SOUSA, R. F. et al. Estudo da evolução espaço-temporal da cobertura vegetal do município de Boa Vista-PB, utilizando geoprocessamento. **Revista Caatinga**, v. 21, n. 3, 2008.
- SOUTO, P. C. et al. Características químicas da serapilheira depositada em área de Caatinga. **Revista Caatinga**, Mossoró, v. 22, n. 1, p. 264-272, 2009. Disponível em: <https://periodicos.ufersa.edu.br/index.php/caatinga/article/view/1195> Acesso em: 19 jul. 2020.
- SOUZA, B. F. S.; TEIXEIRA, A. S.; SILVA, A. T. F. Classificação de bioma caatinga usando Support Vector Machines (SVM). **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto**, p. 7917-7924, 2009.
- THANH NOI, P.; KAPPAS, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. **Sensors**, v. 18, n. 1, p. 18, 2018.
- THENKABAIL, P. S.; WU, Z. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data. **Remote Sensing**, v. 4, n. 10, p. 2890-2918, 2012. MDPI AG. <http://dx.doi.org/10.3390/rs4102890>.
- USGS. **LANDSAT Satellit Missions**. Disponível em: <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-satellite-missions>. Acesso em: 06. Set. 2020.
- VARELLA, C. A. A. **Geoprocessamento na Agricultura de Precisão**. Apostila. 1a Semana Acadêmica de Engenharia de Agrimensura. UFRRJ, 2004.

XU, Y. et al. Coarse to fine K nearest neighbor classifier. **Pattern recognition letters**, v. 34, n. 9, p. 980-986, 2013. <http://doi:10.1016/j.patrec.2013.01.028>.

ZANETTI, J., BRAGA, S. L. F. e DUARTE, O. C. D. **Comparação dos métodos de classificação supervisionada de imagem máxima verossimilhança, distância euclidiana, paralelepípedo e redes neurais em imagens Vant, utilizando o método de exatidão global, índice Kappa e o TAU.** IV Simpósio Brasileiro de Geomática. Presidente Prudente – SP. 2017.

ZHONG, L.; GONG, P.; BIGING, G. S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. **Remote Sensing of Environment**, v. 140, p. 1-13, 2014.

ZHONG, L. et al. Automated mapping of soybean and corn using phenology. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 119, p. 151-164, 2016.

APÊNDICE A — Código utilizado na classificação

```

#Pacotes
install.packages("rgdal")
install.packages("raster")
install.packages("caret")
install.packages("randomForest")
install.packages("kernlab")
install.packages("tsutils")
install.packages("nnet")

#Bibliotecas
library(caret)
library(raster)
library(caret)
library(randomForest)
library(kernlab)
library(tsutils)
library(nnet)

require(caret)
require(raster)
require(caret)
require(randomForest)
require(kernlab)
require(tsutils)
require(nnet)

#Imagem
B2<-raster("b2.tif")
B3<-raster("b3.tif")
B4<-raster("b4.tif")
B5<-raster("b5.tif")
B6<-raster("b6.tif")
B7<-raster("b7.tif")
img<-brick(B2,B3,B4,B5,B6,B7)
plot(img)
names(img)<- c(paste0("B",2:7, coll=""))
names(img)
plot(img)
trainData <- shapefile("classificacao.shp")
plot(trainData)

#Classificação
responseCol <- "class"
plot(trainData)
dfAll = data.frame(matrix(vector(), nrow = 0, ncol = length(names(img)) + 1))
dfAll = data.frame(matrix(vector(), nrow = 0, ncol = length(names(img)) + 1))
for (i in 1:length(unique(trainData[[responseCol]]))){

```

```

category <- unique(trainData[[responseCol]][i])
categorymap <- trainData[trainData[[responseCol]] == category,]
dataSet <- extract(img, categorymap)
if(is(trainData, "SpatialPointsDataFrame")){
  dataSet <- cbind(dataSet, class = as.numeric(rep(category, nrow(dataSet))))
  dfAll <- rbind(dfAll, dataSet[complete.cases(dataSet),])
}
if(is(trainData, "SpatialPolygonsDataFrame")){
  dataSet <- dataSet[!unlist(lapply(dataSet, is.null))]
  dataSet <- lapply(dataSet, function(x){cbind(x, class = as.numeric(rep(category, nrow(x))))})
  df <- do.call("rbind", dataSet)
  dfAll <- rbind(dfAll, df)
}
}

```

#Modelagem por Random Forest (RF)

```

modFit_rf <- train(as.factor(class) ~., method = "rf", data = dfAll) ## classif por Random Forest
modFit_rf
plot(modFit_rf)
beginCluster(2)
preds_rf <- clusterR(img, raster::predict, args = list(model = modFit_rf))
preds_rf
endCluster()
plot(preds_rf)
writeRaster(preds_rf, filename = "rfclass2.tif")
plot(rfclass2.tif)

```

#Modelagem por K-Nearest Neighbors (KNN)

```

set.seed(1)
modFit_knn <- train(as.factor(class) ~., method = 'knn', data = dfAll)
modFit_knn
beginCluster(2)
preds_knn <- clusterR(img, raster::predict, args = list(model = modFit_knn))
preds_knn
endCluster()
plot(modFit_knn)

```

#Modelagem por Artificial Neural Networks (ANN)

```

modFit_nnet <- train(as.factor(class) ~., method = 'nnet', data = dfAll)
beginCluster(2)
preds_nnet <- clusterR(img, raster::predict, args = list(model = modFit_nnet))
endCluster()
plot(modFit_nnet)

```