

Aplicação de aprendizado de máquina para análise do crescimento urbano: o caso das regiões administrativas Ceilândia e Pôr do Sol no Distrito Federal

RESUMO

Briza da Silva Aguiar
brizassaguiar@gmail.com
orcid.org/0000-0002-6378-9286
Universidade de Brasília (UnB),
Brasília, Distrito Federal, Brasília.

Gustavo Macedo de Mello Baptista
gmbaptista@gmail.com
orcid.org/0000-0002-1973-2725
Universidade de Brasília (UnB),
Brasília, Distrito Federal, Brasília.

O propósito do presente trabalho foi avaliar o crescimento da mancha urbana nas Regiões Administrativas Ceilândia e Pôr do Sol, no Distrito Federal, e determinar suas tendências de crescimento. A análise foi realizada através da classificação pixel a pixel, em imagens TM/Landsat-5 e OLI/Landsat-8, adquiridas nos anos de 1984, 1995, 2006, 2016 e 2021. As cenas foram submetidas a conversão em reflectância de superfície e correção geométrica. Foi realizada a Análise de Componentes Principais para gerar mapas com menor mistura espectral e auxiliar na identificação das classes. A classificação da cobertura da terra foi realizada utilizando o algoritmo de aprendizagem de máquina *K-Nearest Neighbor*. Para avaliar a acurácia da classificação foi empregada a matriz de confusão, por meio dos valores de acurácia do produtor, do usuário e global. Os resultados alcançados foram classificados como satisfatórios. Visando compreender a dinâmica de crescimento das duas cidades, foram geradas curvas de tendência, e, a partir dos dados da classificação da cobertura da terra, foi possível concluir que a Ceilândia Tradicional apresenta tendência de estagnação no avanço da mancha urbana e o Pôr do Sol apresenta tendência de crescimento.

PALAVRAS-CHAVE: Crescimento Urbano. Sensoriamento Remoto. Uso do Solo.

INTRODUÇÃO

A urbanização no Brasil é acompanhada de intensas transformações ambientais, desigualdade social e déficits infraestruturais. O crescimento da mancha urbana constantemente tem ultrapassado os limites legais e naturais, ocupando áreas de interesse ambiental e áreas de risco (MATOS, 2012).

A urbanização no Distrito Federal ocorreu de forma rápida e desordenada devido ao intenso fluxo migratório, diferentemente do planejado para a instalação da nova capital do Brasil, que possuía a ideia de ser o centro do poder e possuir um número populacional controlado e limitado (SEVERO, 2014).

Além da pressão por moradia nas grandes cidades e os ciclos migratórios, o intenso processo de urbanização no Distrito Federal (DF) é caracterizado pelas invasões e pela implantação de uma política de distribuição de lotes realizada pelo governo, onde a terra pública se tornou moeda política de troca. Dessa forma, foi impulsionado o surgimento de novas Regiões Administrativas ao redor da capital (BLANCO, 2018).

Assim, as invasões se proliferaram por todas as regiões do Distrito Federal, a ocupação urbana passou a ser caracterizada pelo avanço das cidades sobre as áreas de chapada de forma não planejada, sendo as encostas consideradas um limite natural impedindo o avanço da mancha urbana para outras áreas (NEVES et. al., 2018).

A política populista de incentivo a criação de setores habitacionais populares, além das crises econômicas que causaram pressão nas grandes, agravou o quadro das ocupações irregulares. Atualmente, diversos setores habitacionais possuem manchas urbanas em Áreas de Preservação Permanente. O Setor Habitacional Pôr do Sol surgiu ocupando as bordas de chapada da Ceilândia Tradicional, convertendo em cidade trechos de sensibilidade ambiental (DIAS; WALDE, 2013).

O crescimento das cidades é um dos principais fatores de mudanças na cobertura da terra, podendo ser avaliado e monitorado através do emprego de detecção de mudanças em séries temporais de imagens de satélite, assim, as técnicas de Sensoriamento Remoto são uma importante ferramenta para análise do crescimento urbano (PASSOS; PINHEIRO, 2021).

O objetivo deste trabalho foi empregar o uso de imagens de média resolução da série temporal Landsat no período de 1984 a 2021, utilizando a técnica de detecção de mudanças e a classificação pixel a pixel com o intuito de descrever a dinâmica de ocupação da cobertura da terra na região da Ceilândia Tradicional e no Pôr do Sol e avaliar as tendências de crescimento da mancha urbana nas duas cidades.

CRESCIMENTO URBANO

O processo de urbanização nas grandes cidades do Brasil ocorreu de forma acelerada e sem planejamento, o que causou diversos problemas ambientais, econômicos e sociais. Os impactos ambientais são cada vez mais perceptíveis nas áreas urbanas do país, causados principalmente pela ocupação desordenada e uso irregular do solo.

A ocupação irregular do solo no DF é um dos principais problemas enfrentados na cidade, com a propagação incontrolável das invasões causando danos irreversíveis aos recursos hídricos e graves problemas de infraestrutura. O crescimento dos condomínios irregulares em Brasília contou com o governo regularizando áreas alvos de grilagem e a omissão no controle do uso do solo, o que impulsionou o aumento de áreas de invasão na capital (SENA, 2015).

Dentre os diversos problemas causados pela ocupação irregular do solo no Distrito Federal, destacam-se os relacionados à ocupação e crescimento da mancha urbana em direção às Áreas de Preservação Permanente, onde as cidades construídas nessas áreas ambientalmente frágeis foram regularizadas pelo Poder Público e continuam se expandindo.

SENSORIAMENTO REMOTO

O Sensoriamento Remoto (SR) é uma ciência de aquisição e monitoramento de informações sobre uma área sem contato físico. A obtenção de informações se dá através da detecção da resposta da interação da radiação eletromagnética com materiais terrestres. Os sensores instalados em aeronaves e satélites permitem a captura de várias faixas de energia da onda eletromagnética emitida pelos objetos na superfície do planeta (USGS, 2018; MENESES; ALMEIDA, 2012).

Entre os diversos usos, o SR é uma ferramenta para analisar o crescimento do espaço urbano e os problemas ambientais causados por essa expansão, facilitando o entendimento das dinâmicas das cidades. Através do uso das imagens de satélite, é possível visualizar a expansão das manchas urbanas e identificar a direção de crescimento de uma cidade (DA SILVA; ARAKI, 2015; KAWASHIMA, 2015).

Para estudos sobre a análise do crescimento urbano, é imprescindível a aplicação da técnica de detecção de mudanças na cobertura da terra. O processo de detecção é realizado a partir da identificação de mudanças entre dois conjuntos de imagens de uma mesma área em épocas diferentes. Essas mudanças são resultado das alterações nos valores de radiância entre as imagens de diferentes épocas e constituem um modo rápido e eficiente de análise (SILVA, 2004; HAYES; SADER, 2001; SINGH, 1989).

No emprego da classificação da cobertura da terra, as técnicas que fazem parte da classificação supervisionada utilizam algoritmos que desempenham a identificação dos padrões espectrais da imagem de satélite com base nas amostras previamente selecionadas e realizam a classificação (MOREIRA, 2005).

Nesse sentido, surge o aprendizado de máquina inserido no contexto de classificação supervisionada. O aprendizado de máquina faz parte da área de estudo da Inteligência Artificial, possuindo o intuito de desenvolver métodos e ferramentas para criar máquinas inteligentes com capacidade de produzir tarefas nas mais diversas áreas. Essas máquinas inteligentes precisam treinar sobre determinado conjunto de amostras, assim, possuem capacidade de adquirir o conhecimento de forma automatizada (MITCHELL, 1977).

Na utilização do aprendizado de máquina para classificação, os algoritmos de aprendizado de máquina apresentam maior precisão e melhores resultados em comparação aos algoritmos utilizados tradicionalmente. O *K-Nearest Neighbor* é

um classificador que se baseia em um conjunto de exemplos de treinamento previamente definidos, identificando os K vizinhos mais próximos (LI et al., 2021; ZHANG et al., 2021; MAXWELL et al., 2018).

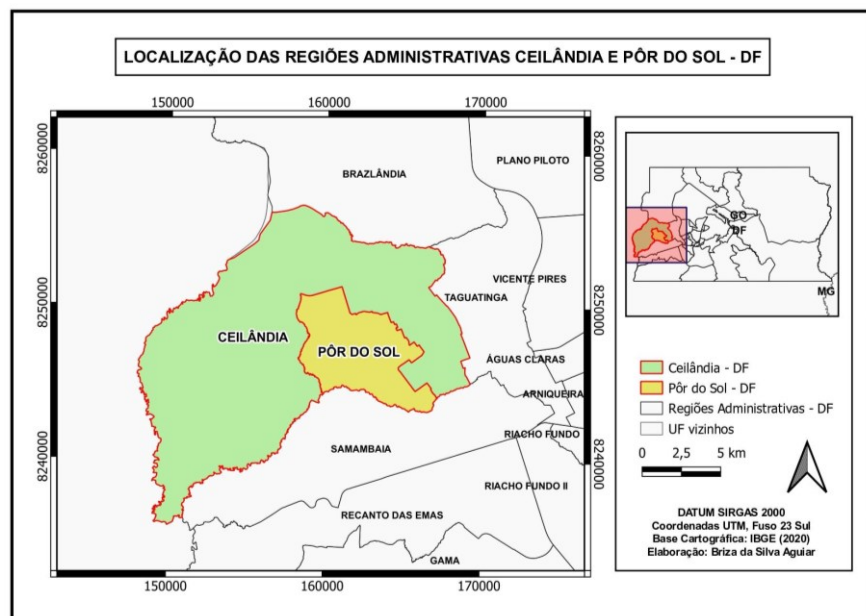
MATERIAIS E MÉTODOS

O Distrito Federal está localizado no Brasil Central com a extensão territorial de 5.760,784 km², sendo dividido em 33 regiões administrativas e população estimada de 3.094.325 habitantes (IBGE, 2021).

A Ceilândia é uma das 33 regiões administrativas do Distrito Federal brasileiro, localizada na porção oeste e situada a 26 quilômetros do centro de Brasília. Sua extensão territorial é de aproximadamente 230,3 km² e detém a maior população urbana em região administrativa do DF, com população estimada em 443.824 habitantes (CODEPLAN, 2018).

As ocupações do Setor Habitacional Sol Nascente e a Área de Regularização de Interesse Social Pôr do Sol foram criadas em 2008 até então pertencendo à Região Administrativa de Ceilândia, havendo uma expansão descontrolada de áreas irregulares. Em agosto de 2019, o Pôr do Sol tornou-se a 32ª Região Administrativa do Distrito Federal, sendo desmembrada da Ceilândia Tradicional. A Região Administrativa (RA) possui extensão territorial de 40,49 km² e população estimada em 83.102 habitantes (CODEPLAN, 2018). Na Figura 1 está demonstrada a localização da área investigada no estudo.

Figura 1 – Mapa de localização da área de estudo



Fonte: Autoria própria com base nos dados do Geoportal do DF (GEOPORTAL, 2022).

A Região Administrativa do Pôr do Sol ocupa áreas com risco alto de erosão e apresenta alta susceptibilidade a alagamentos e processos erosivos acelerados, contrariando grande parte do Distrito Federal onde essas áreas não são urbanizadas (CODEPLAN, 2017).

PROCEDIMENTOS OPERACIONAIS

A aquisição de imagens Landsat 5 e Landsat 8 foi realizada na ferramenta *Earth Explorer* da USGS (*United States Geological Survey*) utilizando os dados da Coleção 2, Nível 2, que estão com correção atmosférica e são geradas a partir dos dados da Coleção 2, Nível 1, que possuem correções geométricas e radiométricas (USGS, 2020).

A correção atmosférica é um procedimento necessário em estudos multitemporais, possuindo o objetivo de minimizar a influência dos componentes atmosféricos, sendo uma forma de normalizar os dados para fins de comparação entre as mais diferentes datas (MOGRANE et al., 2019).

Para a realização do estudo foram selecionadas cenas dos satélites Landsat 5 e Landsat 8, sensores TM (*Thematic Mapper*) e OLI (*Operational Land Imager*) respectivamente, considerando sua extensa coleção de dados. Desta forma, foram adquiridas cinco imagens com as informações descritas na Tabela 1.

As cenas foram selecionadas durante o período de junho, transição do período chuvoso para a seca, com o objetivo de amenizar os efeitos da seca, considerando que alguns estratos de vegetação da região possuem água como fator limitante e ao mesmo tempo seguir o critério de ausência de cobertura de nuvens.

Tabela 1 – Lista de imagens utilizadas no estudo

Data	Satélite	Órbita	Sensor	Resolução, Espacial (m)	Resolução Radiométrica (bits)
11/06/1984	LANDSAT 5	221/71	TM	30	8
10/06/1995	LANDSAT 5	221/71	TM	30	8
24/06/2006	LANDSAT 5	221/71	TM	30	8
09/06/2016	LADNSAT 8	221/71	TM	30	12
17/06/2021	LANDSAT 8	221/71	OLI	30	12

Fonte: Autoria própria (2022).

PRÉ-PROCESSAMENTO

Primeiramente, as imagens foram convertidas em reflectância da superfície da atmosfera (*Bottom Of Atmosphere - BOA*) no *software* QGIS 3.24, usando os coeficientes de redimensionamento fornecidos nos arquivos de metadados.

Para facilitar o processamento, as imagens foram empilhadas utilizando o *software* QGIS 3.24, assim, as bandas que se encontravam em arquivos separados foram inseridas em um único arquivo. Nas cenas do Landsat 5, foram utilizadas as bandas multiespectrais 1, 2, 3, 4, 5 e 7 e no Landsat 8, as bandas 1, 2, 3, 4, 5, 6 e 7. Posteriormente, as cenas foram cortadas no *software* QGIS 3.24 para a área de interesse do estudo.

A correção geométrica foi realizada no *software TerraView 5.6* empregando o registro imagem a imagem, utilizando como referência a cena do ano de 2021 do Landsat 8 por possuir maior resolução espectral e radiométrica.

PROCESSAMENTO

A primeira etapa do processamento de imagens consistiu na realização da Análise de Componentes Principais (ACP) em cada cena utilizando o *software TerraView 5.6*, com o objetivo de gerar novas bandas não correlacionadas com o máximo de informação possível nas primeiras componentes.

As bandas no SR são altamente correlacionadas entre si e espectralmente próximas, dessa forma, parte da informação contida nas bandas apresentam redundância. O método de análise de componentes principais tem a função de determinar a extensão dessa correlação e através de uma transformação matemática, remover a correlação existente entre essas bandas (CROSTA, 1992; CENTENO, 2004).

Essa técnica é utilizada para entender a variabilidade espectral dos alvos, realizando um agrupamento das informações espectrais e gerando um resultado com menor mistura espectral, possuindo o objetivo de olhar a componente principal para definir a amostra, auxiliando na identificação e separação dos alvos por classe.

CLASSIFICAÇÃO

O algoritmo *K-Nearest Neighbor* (KNN) é um classificador que realiza sua função a partir das amostras de treinamento. Com base na influência do vizinho mais próximo, ele calcula a distância entre cada objeto da imagem e todos os objetos de treinamento que foram definidos e com isso decide se o pixel a ser classificado pertence a uma ou outra classe. É uma técnica empregada no reconhecimento de padrões onde o software exige a seleção de valores para o parâmetro *k*, sendo este o número de vizinhos usados para a classificação (CAMPOS et.al., 2013; XU et. Al., 2013).

O processamento digital das imagens para a classificação foi realizado no *software Qgis 3.24* com a utilização do algoritmo KNN com a finalidade de extrair informações e reconhecer padrões homogêneos. Foram definidas oito classes de cobertura da terra: os Corpos D'água (reservatórios, lagos e rios), Mata (formações florestais), Cerrado (formações savânicas), Campo (formações campestres), Solo (áreas erodidas, cascalheiras e solo exposto), Urbano (áreas urbanas consolidadas e não consolidadas), Agropastoril (culturas agrícolas e pastagem) e Reflorestamento (eucalipto e pinus).

Após a definição, foram criados vetores selecionando as regiões de interesse de cada classe em cada cena, com as amostras seguindo o critério de possuírem o conjunto de pixels homogêneos e tamanho semelhantes, percepção tida a partir da análise de componentes principais. O algoritmo foi treinado utilizando 30% dos pixels das amostras de interesse selecionadas e em seguida o modelo treinado obtido foi usado para a geração da classificação.

Para verificar a precisão da classificação foi utilizado o *plugin* AcATAMa no Qgis para a avaliação de precisão de mapas temáticos. Os mapas gerados na classificação foram validados empregando a amostragem aleatória estratificada, foram definidos 20 pontos para cada classe, totalizando 140 pontos para as imagens que não possuem a classe reflorestamento e 160 pontos para as que possuem, criados em coordenadas aleatórias na área dos mapas temáticos com distância mínima de 500 metros entre os pontos gerados. A avaliação de precisão foi realizada utilizando a comparação entre o mapa temático classificado e a imagem de referência.

Após a verificação da qualidade dos mapas gerados, os dados referentes ao crescimento urbano foram inseridos em uma planilha no Excel com o intuito de examinar as tendências de crescimento ao longo da série histórica analisada.

RESULTADOS E DISCUSSÃO

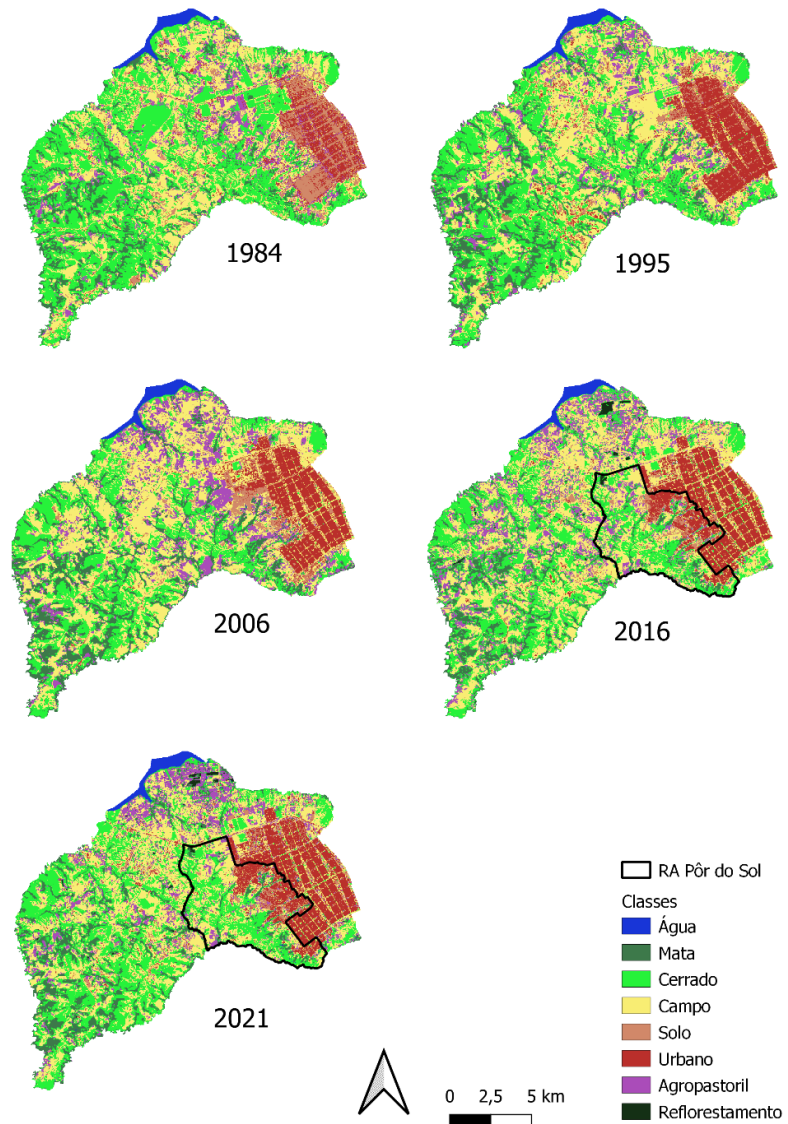
No ajuste de imagem utilizando o *software TerraView* 5.6 com o intuito de realizar a correção geométrica foram selecionados 5 pontos de controle automáticos na imagem de 2016 com erro final de 0,15 pixel. No ajuste das imagens de 1984, 1995 e 2006 foram gerados 8 pontos de controle e todas ficaram com erro final menor do que 0,50 pixel.

O resultado da classificação da cobertura para cada ano está apresentado na Figura 2. A Figura 2 apresenta o resultado para o ano de 1984 que atingiu 98% de acurácia global, para o ano de 1995 que atingiu 97% de acurácia global, para o ano de 2006 que atingiu 95% de acurácia global, para o ano de 2016 que atingiu 97% de acurácia global e para o ano de 2021 que atingiu 97% de acurácia global.

A matriz de confusão é uma maneira eficaz de representar a acurácia na precisão da classificação de cada categoria, é uma ferramenta com arranjo quadrado de números definidos em linhas e colunas, que apresenta o grau de similaridade entre os dados de referência (colunas) e os dados de classificação (linhas), gerados a partir das classificações de cobertura da terra (CONGALTON; GREEN, 2009).

A partir da matriz de confusão é possível calcular medidas específicas como a acurácia global, que corresponde a soma total dos pixels corretamente classificados dividida pelo número total de pixels da matriz de confusão. Dessa forma, todas as classificações de cobertura da terra alcançaram resultados satisfatórios (CONGALTON, 1991).

Figura 2 – Classificação da cobertura da terra na Ceilândia e Pôr do Sol



Fonte: Autoria própria (2022).

Ainda com relação à qualidade do treinamento do algoritmo KNN de classificação, foi observado o coeficiente Kappa, que é dado após a construção da matriz da confusão. O valor de Kappa é um índice referente ao grau de concordância dos dados, medindo a coerência entre a classificação realizada na imagem e seus valores reais (LANDIS; KOCH, 1977).

Tabela 2 – Valores do Índice Kappa

Índice Kappa	Qualidade da classificação
< 0,00	Péssimo
0,0 – 0,2	Ruim
0,2 – 0,4	Razoável
0,4 – 0,6	Boa
0,6 – 0,8	Muito boa
0,8 – 1,0	Excelente

Fonte: Adaptado de Landis e Koch (1977).

O treinamento dos algoritmos utilizados na classificação para os anos de 1984, 1995 e 2006 obtiveram Índice Kappa de 0,95, para o ano de 2016 obteve 1,0 e para o ano de 2021 obteve 0,99. Os resultados alcançados no treinamento utilizando o algoritmo de aprendizagem KNN foram classificados como excelentes.

A análise de acurácia avalia a qualidade do mapeamento realizado, onde são dadas as taxas de acerto geral, taxas de acerto e erro para cada classe definida. A acurácia do produtor define quais são as frações de pixels de cada classe corretamente designados às suas classes pelo classificador. A acurácia do usuário se refere às estimativas das frações de pixels do mapeamento corretamente classificados (MAPBIOMAS, 2019; PONTIUS; MILLONES, 2011).

As matrizes de confusão referentes aos mapas de classificação da cobertura da terra da Ceilândia e Pôr do Sol estão nas Tabelas 3, 4, 5, 6 e 7. Cada valor se refere a uma classe: 1 (Água), 2 (Mata), 3 (Cerrado), 4 (Campo), 5 (Solo), 6 (Urbano), 7 (Agropastoril) e 8 (Reflorestamento).

A Tabela 3 traz a matriz de confusão da classificação da cobertura da terra para o ano de 1984 com 95% de acurácia do usuário. Ao analisar é possível observar que houve confusão na classificação da classe Solo e Urbano. A classe Solo no momento da classificação foi confundida com Urbano e isso se dá pelo fato do solo exposto possuir resposta espectral parecida com a de alguns materiais que existem nas cidades. A confusão ocorrida na classe Urbano, sendo classificada como Campo, Solo e Agropastoril, é causada pela variedade de materiais que as cidades possuem e suas feições espectrais variadas, o que causa confusão espectral no momento da classificação.

Tabela 3 – Matriz de confusão da classificação da cobertura da terra para o ano de 1984

	1	2	3	4	5	6	7
1	1,0	-	-	-	-	-	-
2	-	1,0	-	-	-	-	-
3	-	-	1,0	-	-	-	-
4	-	-	-	1,0	-	-	-
5	-	-	-	-	0,93	0,07	-
6	-	-	-	0,05	0,05	0,73	0,17
7	-	-	-	-	-	-	1,0
Produtor	1,0	1,0	1,0	0,94	0,93	0,92	0,88

Fonte: Autoria própria (2022).

Na classificação para o ano de 1995, apresentada na Tabela 4, ocorreu a mesma confusão da classe Urbano com a classe Solo, também explicada pela composição de materiais das cidades, para as outras classes não houve confusão espectral durante a classificação, gerando a matriz de confusão que indica o melhor desempenho na classificação de todo o período estudado. A acurácia do usuário atingiu 97% de acerto.

Tabela 4 – Matriz de confusão da classificação da cobertura da terra para o ano de 1995

	1	2	3	4	5	6	7
1	1,0	-	-	-	-	-	-
2	-	1,0	-	-	-	-	-
3	-	-	1,0	-	-	-	-
4	-	-	-	1,0	-	-	-
5	-	-	-	-	1,0	-	-
6	-	-	-	-	0,2	0,8	-
7	-	-	-	-	-	-	1,0
Produtor	1,0	1,0	1,0	1,0	0,8	1,0	1,0

Fonte: Autoria própria (2022).

A Tabela 5 apresenta a matriz de confusão da classificação da cobertura da terra para o ano de 2006. A confusão da classe Urbano se deu com a classe Campo, motivada pelos materiais com resposta espectral semelhantes que as cidades possuem com os que compõem a classe Campo. Ainda, o erro apresentado entre as classes Cerrado e Mata pode ocorrer quando uma classe apresenta maior área e consequentemente um número maior de amostras em relação a outra, isso porque durante o processo de previsão de novas amostras, a maioria dos vizinhos pode pertencer a classe maior de amostras, gerando confusão no momento da

classificação. Dessa forma, a acurácia do usuário resultou em 93% de acerto (TANG; HAIBO, 2015).

Tabela 5 – Matriz de confusão da classificação da cobertura da terra para o ano de 2006

	1	2	3	4	5	6	7
1	1,0	-	-	-	-	-	-
2	-	0,95	0,05	-	-	-	-
3	-	-	1,0	-	-	-	-
4	-	-	-	1,0	-	-	-
5	-	-	-	-	1,0	-	-
6	-	-	-	0,2	-	0,8	-
7	-	-	-	-	-	-	1,0
Produtor	1,0	0,82	0,90	0,83	1,0	1,0	1,0

Fonte: Autoria própria (2022).

Avaliando a matriz de confusão para a classificação de cobertura do ano de 2016, conforme a Tabela 6, houve também a confusão na classificação das classes Campo e Urbano. A partir do ano de 2016, surge a classe Reflorestamento composta por eucaliptos e pinus, durante a classificação a classe foi confundida com a Mata, pois ambas são formações florestais e durante a classificação houve confusão espectral entre as classes, resultando em 90% na acurácia do usuário

Tabela 6 – Matriz de confusão da classificação da cobertura da terra para o ano de 2016

	1	2	3	4	5	6	7	8
1	1,0	-	-	-	-	-	-	-
2	-	1,0	-	-	-	-	-	-
3	-	-	1,0	-	-	-	-	-
4	-	-	-	0,95	-	0,05	-	-
5	-	-	-	-	1,0	-	-	-
6	-	-	-	0,2	-	0,8	-	-
7	-	-	-	-	-	-	1,0	-
8	-	0,2	-	-	-	-	-	0,8
Produtor	1,0	0,5	0,8	0,8	1,0	0,95	0,95	1,0

Fonte: Autoria própria (2022).

Ao analisar a Tabela 7, é possível constatar que além das confusões esperadas e ocorridas no período estudado, ocorreu a confusão da classe Agropastoril com a classe Urbano, sendo explicada pelo fato dos diferentes tipos de cobertura e uso do solo englobados em ambas as classes, causando assim misturas e semelhanças

de respostas espectrais. Essa ocorrência de confusões espectrais resultou em 93% de acurácia do usuário.

Tabela 7 – Matriz de confusão da classificação da cobertura da terra para o ano de 2021

	1	2	3	4	5	6	7	8
1	1,0	-	-	-	-	-	-	-
2	-	1,0	-	-	-	-	-	-
3	-	-	1,0	-	-	-	-	-
4	-	-	-	0,95	-	0,05	-	-
5	-	-	-	-	1,0	-	-	-
6	-	-	-	0,2	0,1	0,7	-	-
7	-	-	-	-	-	0,05	0,95	-
8	-	0,2	-	-	-	-	-	0,8
Produtor	1,0	0,8	1,0	0,8	0,9	0,95	0,9	1,0

Fonte: Autoria própria (2022).

Ao avaliar as matrizes de confusão apresentadas, é possível observar que a maior confusão ocorrida durante as classificações se refere a classe Urbano, essa confusão ocorre porque as cidades são compostas de vários materiais como o asfalto, concreto, telhados, coberturas metálicas, entre outras. Dessa forma, o conjunto desses materiais apresenta diversas respostas espectrais, conjunto de pixels divergentes e respostas espectrais que se assemelham com as outras classes, o que causa a confusão entre as classes (FORESTI; PEREIRA, 1987).

A análise das matrizes mostra que o algoritmo utilizado na classificação apresentou satisfatórios resultados, mesmo com a confusão ocorrida em algumas classes, foram confusões pequenas e esperadas, o que justifica o nível de acurácia e exatidão elevados. A confiabilidade da classificação foi satisfatória e aponta a veracidade das áreas classificadas.

A análise comparativa temporal na classe Urbana no período analisado permite constatar o crescimento da mancha urbana, no ano de 1984 as construções urbanas ocupavam 5% da cobertura da terra na região, o equivalente a 13,93 km². Já no ano de 2021, a mancha urbana ocupava 12% da cobertura da região, o equivalente a 28,42 km², o crescimento mais que dobrou nos 37 anos analisados. Na Figura 3 está representado no gráfico o crescimento urbano em cada período estudado e a curva de tendência do crescimento.

Figura 3 – Expansão da área urbanizada na Ceilândia e Pôr do Sol



Fonte: Autoria própria (2022).

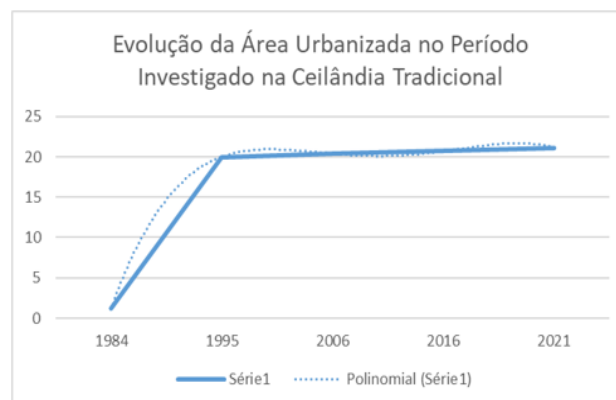
A curva do gráfico representado na Figura 3 traz a mancha urbana das duas Regiões Administrativas e mostra a intensidade do avanço do crescimento urbano no período avaliado, sendo registrado um crescimento de 104%. Na Figura 2, é possível observar que no ano de 1984, a densidade da ocupação urbana é menor e a predominância é de solo exposto. Em 1995, onde é apresentado a maior taxa de crescimento em todo o período investigado, a mancha urbana está adensada, consolidada e ocupando novas áreas.

Seguindo a tendência geral do Distrito Federal de apresentar estagnação no crescimento da mancha urbana, a evolução da área urbanizada no período de 2016 apresenta um ritmo menor do que no início da série temporal estudada, a linha de tendência começa a apresentar uma estabilização (BAPTISTA; NASCIMENTO, 2016).

Considerando que as duas regiões estudadas possuem características distintas de ocupação urbana, a Ceilândia Tradicional tendo sido planejada e implantada nos anos 70 e o Pôr do Sol tendo surgido a partir de ocupações informais e sendo regularizado no ano de 2019, ambas apresentam dinâmicas de crescimento diferentes.

Objetivando a análise mais precisa do crescimento urbano das duas regiões, gerou-se a curva de crescimento e tendência para cada uma das cidades. Na Figura 4 está apresentada a evolução da área urbanizada na Ceilândia Tradicional.

Figura 4 – Evolução da área urbanizada no período investigado na Ceilândia Tradicional



Fonte: Autoria própria (2022).

Conforme observado, a Ceilândia Tradicional segue a tendência de estagnação no crescimento da mancha urbana, por se tratar de uma região mais antiga e considerando que o avanço da mancha urbana encontrou limitadores legais e naturais. Entretanto, a tendência de estagnação não significa que o tamanho da população também apresenta estabilidade, considerando que pode ocorrer a verticalização do espaço urbano e conseqüentemente o aumento populacional (BAPTISTA; NASCIMENTO, 2016).

Seguindo a tendência do Distrito Federal, no período de 1984 a 1995 é demonstrado um crescimento vertiginoso da ocupação urbana na Ceilândia Tradicional, esse intenso crescimento se deu pela construção de novos setores habitacionais e a expansão desses setores, caracterizada pela época de maior crescimento da cidade.

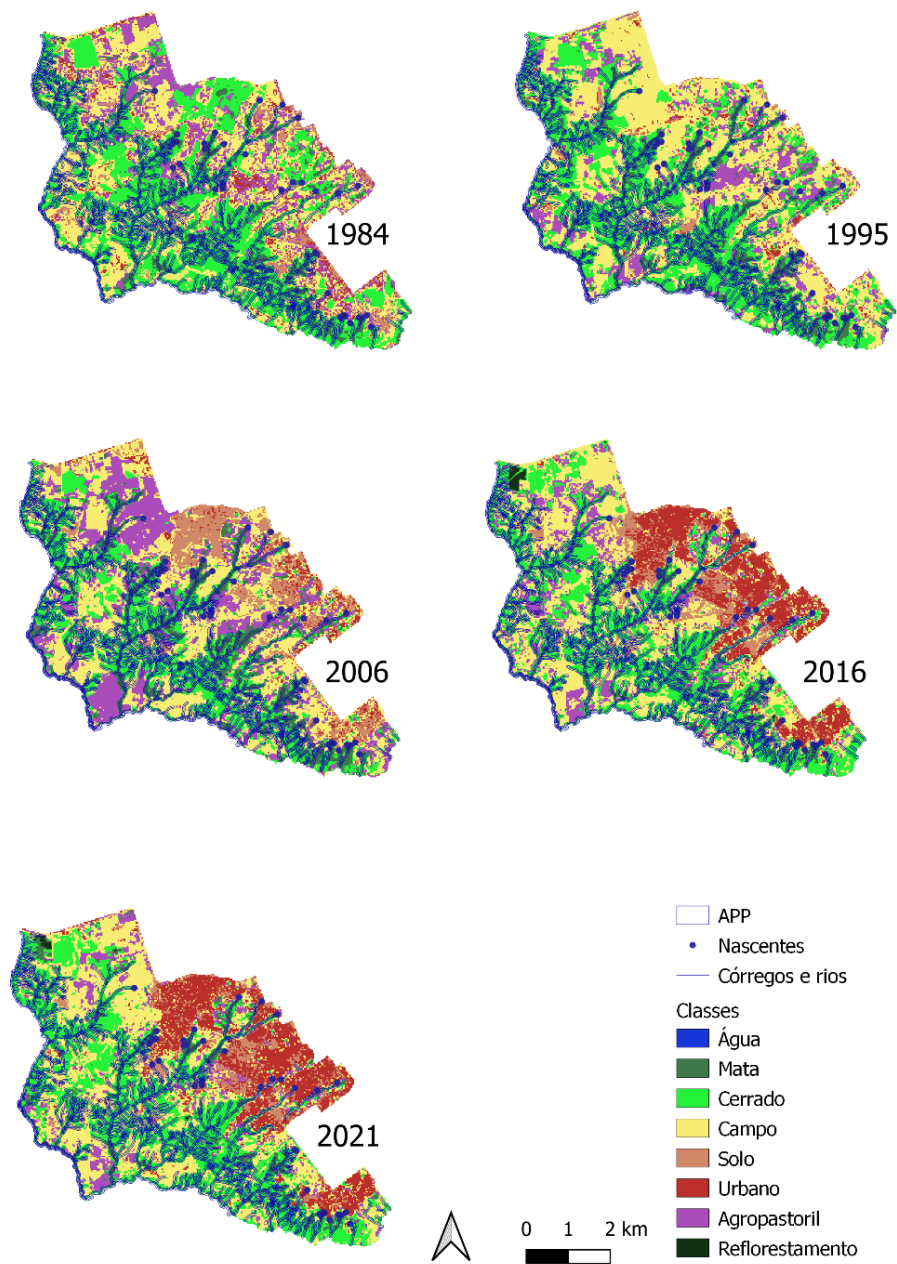
No ano de 1994, as ocupações irregulares na Ceilândia começaram a ocupar áreas ambientalmente frágeis, o início da criação do Setor Habitacional Sol Nascente e a Área de Interesse Social - ARIS Pôr do Sol marcam a intensa ocupação urbana em Áreas de Preservação Permanente. Atualmente, áreas de grande sensibilidade ambiental têm sofrido remoção de vegetação para a implantação de áreas urbanas. O resultado da mudança de cobertura da terra das Áreas de Preservação Permanente na região onde hoje é RA Pôr do Sol está na Figura 5.

Os mapas de mudanças na cobertura da terra da Região Administrativa Pôr do Sol mostram o avanço da mancha urbana sobre as Áreas de Preservação Permanente, as imagens evidenciam que a transformação na cobertura do solo vem acontecendo de forma intensiva, convertendo a vegetação em ocupações urbanas.

Na Figura 6, é demonstrado o crescimento da mancha urbana na Região Administrativa Pôr do Sol, desde o período em que a região ainda era Ceilândia e começou a ser ocupada de forma irregular até a sua regularização a partir do ano de 2019.

A ocupação irregular da região começou no ano de 1994, no Gráfico é possível observar o crescimento ascendente de áreas urbanizadas durante o período analisado. Dessa forma, conclui-se que o crescimento da área urbanizada no setor Pôr do Sol influenciou a tendência de crescimento demonstrada no gráfico da Figura 3, considerando que é uma cidade nova e em expansão.

Figura 5 – Cobertura da terra nas Áreas de Preservação Permanente no Pôr do Sol



Fonte: Autoria própria (2022).

Figura 6 – Expansão da área urbanizada no Pôr do Sol



Fonte: Autoria própria (2022).

Observando a Figura 6 é possível notar o aumento do crescimento urbano no ano de 1995, período em que começaram a ser registradas as ocupações irregulares na região. O gráfico na maior parte do período estudado apresenta o crescimento da mancha urbana e a partir do ano de 2016 começa a apresentar uma tendência de menor crescimento, causado pelas limitações naturais da região, a existência de encostas. Apesar das restrições e limites legais que existem, esse fator não impediu o crescimento da cidade sobre as áreas legalmente protegidas.

Na região do Pôr do Sol, além do impacto ambiental, a construção de casas nessas áreas representam risco de vida, as casas apresentam perigo de desabamentos, além da ocorrência de alagamentos.

Contrariando a estabilização no avanço da mancha urbana no Distrito Federal, a RA Pôr do Sol surge em uma demanda por moradia e apresenta tendência de crescimento. O crescimento apresentado possui a característica de ser menor e se encontra em desaceleração, devido ao fato de a região apresentar delimitadores geográficos como as encostas.

CONCLUSÃO

O método de classificação supervisionada utilizando o algoritmo KNN apresentou desempenho satisfatório, se mostrando como um classificador simples e eficiente com capacidade de classificar cada elemento do conjunto de amostras de maneira correta. Os dados apresentaram confiabilidade para estudar o crescimento urbano.

Dos dois cenários de crescimento apresentados, a Ceilândia apresenta tendência de estagnação e o Pôr do Sol apresenta tendência de crescimento por se tratar de uma área emergente por demanda de moradia. Ainda, mesmo com a tendência de crescimento, é observada uma desaceleração no avanço da mancha urbana na RA Pôr do Sol devido as características ambientais e geográficas da região.

Até a década de 90, a área onde está localizada a RA Pôr do Sol era ocupada por chácaras, somente a partir de 1997 surgiram focos de parcelamentos de lotes.

No período de 2000 até 2010 foram registradas as marcas de maior crescimento populacional. Esses dados concordam com os mapas de cobertura gerados e com os gráficos de expansão da área urbanizada (CODEPLAN, 2019).

Ainda, o período de 1984 e 1995 na RA Ceilândia foi marcado pela criação de novos setores habitacionais, o que implicou no crescimento vertiginoso da população. Esse crescimento também foi observado nos mapas de cobertura gerados para a Região Administrativa estudada e é possível notar o crescimento ascendente no gráfico de expansão da área urbanizada (CODEPLAN, 2017).

Destaca-se que os dados obtidos pela classificação são coerentes com as informações encontradas durante a revisão bibliográfica. Considerando a concordância dos mapas de cobertura com os dados de crescimento urbano obtidos na revisão para a realização do estudo, pode-se concluir que o uso do algoritmo apresentou resultados satisfatórios.

Considerando os dados de crescimento obtidos, é possível observar que as ocupações informais se tornam cidades consolidadas em um curto período, isso se dá pelo processo de regularização das áreas irregulares, o que incentiva a especulação imobiliária dessas regiões.

Dadas as circunstâncias de ocupações na área de estudo, a demanda por moradia, a omissão do Poder Público em fiscalizar as áreas legalmente protegidas e a rápida consolidação de cidades nas áreas invadidas, a regularização dessas áreas se apresenta como a melhor alternativa de modo a garantir o direito social à moradia e o desenvolvimento das funções sociais.

Ressalta-se a necessidade de que sejam elaborados estudos com o intuito de abordar a expansão do setor Pôr do Sol e seus impactos ambientais, para que haja mensuração e mitigação desses impactos. Também é necessário elaborar instrumentos de políticas habitacionais com o objetivo de regularizar ocupações precárias a fim de oferecer serviços essenciais para o bem-estar humano, além de que ocorra nas cidades o desenvolvimento socioeconômico.

Machine learning in analysis of urban growth: case of Ceilândia and Pôr do Sol in Brazil's Federal District

ABSTRACT

This work sought to evaluate the growth of the urban area in the Administrative Regions of Ceilândia and Pôr do Sol, in the Federal District, Brazil, and to determine their growth trends. The analysis was performed through pixel-by-pixel classification, in TM/Landsat-5 and OLI/Landsat-8 images, acquired in 1984, 1995, 2006, 2016, and 2021. The scenes were submitted to surface reflectance conversion and geometric correction. Principal Component Analysis was carried out to generate maps with less spectral mixing and assist in the identification of classes. Land cover classification was performed using the K-Nearest Neighbor machine learning algorithm to assess the accuracy of the classification, the confusion matrix was used, through the values of global, producer and user accuracy. The results achieved were classified as excellent. In order to understand the growth dynamics of the two cities, trend curves were generated and, from the data of the land cover classification, it was possible to conclude that the traditional Ceilândia presents a tendency of stagnation in the advance of the urban area and Pôr do Sol shows a growth trend.

KEYWORDS: Urban Growth. Remote Sensing. Land Cover.

REFERÊNCIAS

BAPTISTA, G. M. M.; NASCIMENTO, M. E. M.. Geração de cenários de crescimento da mancha urbana da Área Metropolitana de Brasília por meio da série histórica DMSP-OLS de dados de luz noturna. **Revista brasileira de Cartografia**, v. 68, n. 5, 2016. <https://doi.org/10.14393/rbcv68n5-44424>.

BLANCO, K. C.. As peculiaridades das ocupações irregulares no Distrito Federal. Paranoá, n. 21, 2018.

CAMPOS, A. R.; FONSECA, D. F.; SAMPAIO, D. M.; FARIA, S. D.; ELMIRO, M. A. T.. Classificação orientada a objeto para mapeamento do uso e cobertura do solo do município de Rio Acima - MG: por crescimento das regiões por extração de objetos. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE.

CENTENO, J. A. S.. Sensoriamento Remoto e Processamento de Imagens Digitais. Curitiba: Ed. Curso de Pós-graduação em Ciências Geodésicas, Universidade Federal do Paraná, 2004.

CODEPLAN. Caracterização urbana e ambiental unidade de planejamento territorial. 2017. Disponível em <<https://www.codeplan.df.gov.br/wp-content/uploads/2018/02/Estudo-UrbanoAmbiental-Unidade-de-Planejamento-Territorial-UPT-OESTE.pdf>>. Acessos em 15 fev. 2022.

CODEPLAN. Pesquisa Distrital por Amostra de Domicílio 2018. 2018. Disponível em <https://www.codeplan.df.gov.br/wp-content/uploads/2019/03/PDAD_DF-Grupo-de-Renda-compactado.pdf>. Acessos em 11 fev. 2022.

CODEPLAN. Sol Nascente/Pôr do Sol: Um Retrato Demográfico e Socioeconômico. 2019. Disponível em <https://www.codeplan.df.gov.br/wp-content/uploads/2018/03/NT_Sol_Nascente_Por_do_Sol-compactado.pdf>. Acessos em 10 jan. 2023.

CONGALTON, R. G.. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35-46. 1991. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B).

CONGALTON, R. G.; GREEN, K.. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. 2. ed, CRC Press, p. 40-60. 2009.

CROSTA, A. P.. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas, São Paulo, p. 170. 1992.

DA SILVA, F. M. M.; ARAKI, H.. Integração de dados de Imagens Orbitais de Alta Resolução e ALS para detecção semi-automática de Edificações em áreas urbanas. **Boletim de Ciências Geodésicas**, Universidade Federal do Paraná, Curitiba, v. 21, n.3, p. 548-517, 2015. <http://dx.doi.org/10.1590/S1982-21702015000300031>.

DIAS, L. T.; WALDE, D. H. G.. Modelagem da dinâmica espacial do uso e ocupação do solo na Bacia Hidrográfica do Lago Paranoá-DF. **Revista Brasileira de Cartografia**, v. 65, n. 1, p. 77-94, 2013. <https://doi.org/10.14393/rbcv65n1-44784>.

FORESTI, C.; PEREIRA, M. D. B.. Utilização de índices vegetativos obtidos com dados do sistema TM Landsat no estudo da qualidade ambiental urbana: cidade de São Paulo. 1987, São José dos Campos, INPE.

Geoportal/DF – Infraestrutura de Dados Espaciais do Distrito Federal. 2022. Disponível em <seduh.df.gov.br>. Acessos em 02 fev. 2022.

HAYES, J. D.; SADER, A. S.. Change detection techniques for monitoring forest clearing and regrowth in a tropical moist forest. **Photogrammetric Engineering and Remote Sensing**, v. 67, n. 9, p. 1067-1075, 2001.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Cidades e Estados. 2021. Disponível em <<https://www.ibge.gov.br/cidades-e-estados/df/brasil.html>>. Acessos em 02 fev. 2022.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Malha Municipal - IBGE. 2020. Disponível em <<https://www.ibge.gov.br/geociencias/organizacao-doterritorio/malhas-territoriais/15774-malhas.html?=&t=downloads>>. Acessos em 02 fev. 2022.

KAWASHIMA, R. S. **Modelagem da dinâmica espacial das mudanças na cobertura da terra na região portuária da Baixada Santista**. 190 p. Dissertação (Mestrado em Ciências) - Escola Politécnica da Universidade Estadual de São Paulo, São Paulo, 2015. <http://dx.doi.org/10.11606/D.3.2016.tde-31122015-111819>.

LANDIS, J. R.; KOCH, G. G.. The measurement of observer agreement for categorical data. **Biometric**, v. 33, p. 159-174, 1977.

LI, R. et al.. Phenology-based classification of crop species and rotation types using fused MODIS and Landsat data: The comparison of a random-forest-based model and a decisionrule-based model. **Soil and Tillage Research**, v. 206, p. 104838, 2021.

MAPBIOMAS. Análise de Acurácia. Disponível em <<https://mapbiomas.org/analisedeacuracia#:~:text=Acur%C3%A1cias%20do%20produtor%3A%20S%C3%A3o%20as,pixel%20da%20classe%20j%20corretamente>>. Acessos em 15 fev . 2022.

MATOS, R. Migração e urbanização no Brasil. **Revista Geografias**, p. 7-23, 2012.

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. <https://doi.org/10.1080/01431161.2018.1433343>.

MENESES, P. R.; ALMEIDA, T.. de. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Universidade de Brasília, Brasília, 2012.

MITCHELL, T.. Machine Learning. New York: McGraw Hill. 1977.

MOGRANE, A. M.; JAMET, C.; LOISEL, H.; VANTREPOTTE, V.; MÉRIAUX, X.; CAUVIN, A.. Evaluation of five atmospheric correction algorithms over french optically-complex Waters for the Sentinel-3^a OLCI ocean color sensor. **Remote Sensing**, v. 11, n. 3, p.1-25. 2019. <https://doi.org/10.3390/rs11060668>.

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

NEVES, G.; SENA-SOUZA, J. P.; VASCONCELOS, V.; MARTINS, E. S.; COUTO JÚNIOR, A. F.. Dinâmica da cobertura da terra do Distrito Federal dentro de suas unidades geomorfológicas. **Sociedade & Natureza**, v. 29, n. 3, p. 383-396, 2018. <https://doi.org/10.14393/SN-v29n3-2017-2>.

PASSOS, J. S.; PINHEIRO, A. V. R.. Sensoriamento remoto aplicado à análise da evolução da mancha urbana em Marabá – PA entre os anos de 1999 a 2019: implicações socioambientais. **Revista Geografia em Atos**, v. 5, p. 1-17, 2021. <https://doi.org/10.35416/geoatos.2021.8095>.

PONTIUS JR., R. G.; MILLONES, M.. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**. v. 32, n. 15, p. 4407-4429, 2011. <https://doi.org/10.1080/01431161.2011.552923>.

SENA, J. V. Análise das Políticas de Controle do Uso do Solo e de Regularização Fundiária no Distrito Federal. 2015.

SEVERO, D. S.. Planejamento urbano no Distrito Federal: o caso da Ceilândia. 73 f. Monografia (Bacharelado em Geografia) – Universidade de Brasília, Brasília, 2014.

SILVA, A. A. **Uso de dados MODIS para o monitoramento ambiental nos cerrados: Um estudo a partir de dados MOD13Q1 realçados através de um modelo de mistura espectral.** Dissertação (Mestrado em Geografia) – Instituto de Estudos Socioambientais da Universidade Federal de Goiânia, Goiânia, 2004.

SINGH, A.. Digital change detection techniques using remotely-sensed data. **International Journal of Remote Sensing**, v.10, p. 989-1003, 1989. <https://doi.org/10.1080/01431168908903939>.

TANG, B; HAIBO, H.. ENN: Extended Nearest Neighbor Method for Pattern Recognition. **IEEE Computational intelligence magazine**, v.10, n. 3, p. 52-60, 2015. <https://doi.org/10.1109/MCI.2015.2437512>.

USGS. United States Geological Survey. Landsat 8 Collection 2 (C2) Level (2) Science Product (L2SP) Guide. 2020. Department of the Interior U.S. Geological Survey. Disponível em <<https://www.usgs.gov/media/files/landsat-8-collection-2-level-2-science-productguide>>. Acessos em 02 fev. 2022.

USGS. United States Geological Survey. What is remote sensing and what is it used for?. 2018. Disponível em <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used?qt-news_science_products=7#qt-news_science_products>. Acessos em 10 jan. 2023.

XU, Y.; ZHU, Q.; FAN, Z.; QIU, M.; CHEN, Y.; LIU, H. Coarse to fine K nearest neighbor classifier. **Pattern Recognition Letters**, North-Holland, v. 34, p. 980-986, Feb. 2013. <https://doi.org/10.1016/j.patrec.2013.01.028>

ZHANG, T.; SU, J.; XU, Z.; LUO, Y.; LI, J. Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. **Applied Sciences**, v. 11, n. 2, p. 543, 2021. <https://doi.org/10.3390/app11020543>

Recebido: 08 set. 2022

Aprovado: 13 fev. 2023

DOI: 10.3895/rbgeo.v11n2.15936

Como citar: AGUIAR, B. S.; BAPTISTA, G. M. M.. Aplicação de aprendizado de máquina para análise do crescimento urbano: o caso das regiões administrativas Ceilândia e Pôr do Sol no Distrito Federal. **R. bras. Geom.**, Curitiba, v. 11, n. 2, p. 465-487, abr./jun. 2023. Disponível em: <<https://periodicos.utfpr.edu.br/rbgeo>>. Acesso em: XXX.

Correspondência:

Briza da Silva Aguiar

QNM 22, conjunto G, 21, CEP 72210-227, Ceilândia, Distrito Federal, Brasil.

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.