



Deep Learning para Classificação Supervisionada de Imagens CBERS-4A da Área Urbana de Rio Claro (SP)

Deep Learning for Supervised Classification of CBERS-4A Imagery in Rio Claro (SP) Urban Area

Danilo Marques de Magalhães¹, Julya Paes de Souza², Edgar Auler Galvão de França³

¹ Universidade Estadual Paulista “Júlio de Mesquita Filho”, Departamento de Geografia e Planejamento Ambiental, Rio Claro (SP), Brasil. danilo.magalhaes@unesp.br

ORCID: <https://orcid.org/0000-0001-9306-4326>

² Universidade Estadual Paulista “Júlio de Mesquita Filho”, Departamento de Engenharia Ambiental, Rio Claro (SP), Brasil. julya.paes@unesp.br

ORCID: <https://orcid.org/0009-0005-4162-0024>

³ Universidade Estadual Paulista “Júlio de Mesquita Filho”, Departamento de Estatística Matemática Aplicada e Computação, Rio Claro (SP), Brasil. Edgar.galvao@unesp.br

ORCID: <https://orcid.org/0009-0002-5625-8011>

Recebido: 09.2024 | Aceito: 04.2025

Resumo: O presente artigo tem por objetivo avaliar a acurácia do mapeamento do uso e cobertura da terra, realizado em um trecho da área urbana de Rio Claro (SP), a partir de técnicas de *Deep Learning* e utilizando uma imagem CBERS-4A (WPM) com 2 m de resolução espacial. Foi estruturada uma rede neural convolucional U-Net a partir de script em Python, utilizando as bibliotecas *Keras* e *Tensor Flow*. A verdade terrestre, utilizada para treinamento e verificação da acurácia do modelo, foi elaborada por meio de classificação supervisionada da mesma imagem no software ArcGIS Pro, utilizando o algoritmo *Support Vector Machine* (SVM) e procedimentos de pós-classificação, incluindo aplicação de filtro majoritário e edição manual de pixels. O resultado obtido pela U-Net foi comparado ao resultado obtido pelo SVM (sem pós-classificação), visando compreender se há ganhos de acurácia, tendo em vista o maior esforço humano, para a criação do *ground truth*, e computacional, para processamento dos dados, inerente às técnicas de *Deep Learning*. Para isso, ambos os resultados foram submetidos a avaliação de acurácia utilizando as métricas *Overall Accuracy*, *Precision*, *Recall*, *F1 Score* e *Kappa*. Constatou-se que o modelo U-Net apresenta melhor acurácia em todas elas, destacando-se o aumento da *Precision* de 0,48 (SVM) para 0,78 (U-Net). Tais resultados indicam o potencial das técnicas de *Deep Learning* para o mapeamento do uso e cobertura da terra em áreas urbanas a partir de imagens de alta resolução, o que pode contribuir, de modo significativo, para ações de planejamento e gestão territorial nos municípios brasileiros.

Palavras-chave: Deep Learning. CBERS-4A. Uso e Cobertura da Terra. U-Net. Classificação Supervisionada.

Abstract: This study evaluates the accuracy of land use and land cover (LULC) mapping in an urban area of Rio Claro (SP) using Deep Learning techniques and a CBERS-4A (WPM) image with 2 m spatial resolution. A U-Net convolutional neural network was developed using Python and the Keras and TensorFlow libraries. Ground truth data for training and accuracy assessment were generated through supervised classification of the same image in ArcGIS Pro, employing the Support Vector Machine (SVM) algorithm, followed by post-classification procedures, including majority filtering and manual pixel editing. U-Net results were compared with SVM results (pre-refinement) to evaluate potential accuracy improvements associated with the greater computational and human effort required by Deep Learning techniques. Both approaches were assessed using Overall Accuracy, Precision, Recall, F1 Score, and Kappa metrics. The U-Net model demonstrated superior performance across all metrics, with a notable increase in Precision from 0.48 (SVM) to 0.78 (U-Net). These findings highlight the potential of Deep Learning methods for high-resolution urban LULC mapping, providing valuable tools for urban planning and territorial management in Brazilian municipalities.

Keywords: Deep Learning. CBERS-4A. Land Use and Land Cover. U-Net. Supervised Classification.

1 INTRODUÇÃO

O uso e cobertura da terra é reconhecidamente uma variável essencial para realização de diagnósticos ambientais e para o planejamento territorial de modo geral (Pabi, 2007; Yu, et al., 2022). Desde o lançamento do primeiro satélite da série Landsat, em 1972, recursos computacionais vêm sendo empregados para classificação dos alvos identificáveis nas imagens com o objetivo de mapear o uso e a cobertura da terra. Essas inovações proporcionaram a realização de mapeamentos em larga escala com maior agilidade, com satisfatória acurácia, e abriram possibilidades para o monitoramento multitemporal das alterações na paisagem (Odenyo & Pettry, 1977).

Atualmente, o avançado estágio tecnológico e científico do Sensoriamento Remoto possibilita a identificação e o mapeamento de feições e fenômenos terrestres a partir de imagens orbitais com altíssima resolução espacial (superiores a 1 m). Essas imagens são, na maioria das vezes, registradas por bandas multiespectrais da região visível e do infravermelho próximo do espectro eletromagnético, possuem um tempo de revisita de poucos dias e registram os valores de radiância dos elementos da superfície terrestre em até 16 bits (65.536 níveis de cinza). Todo esse aporte de dados gera desafios para sua utilização, exigindo maior capacidade de processamento dos hardwares e de separação de classes por parte dos algoritmos classificadores e segmentadores de imagens (Amani et al., 2020; Chi et al., 2016; Parente et al., 2019; Picoli et al., 2020).

Para lidar com esses desafios, métodos de *Deep Learning* (DL) vêm sendo empregados para extrair informações de imagens de sensoriamento remoto tomadas em nível aéreo e orbital e têm demonstrado ótimos resultados para solução de problemas de classificação em cenas complexas. Citam-se, como exemplo, os estudos de caso de Wagner et al. (2020a), que utilizaram a rede neural convolucional U-Net para realizar o mapeamento de edificações em imagens WorldView-3; e de Braga et al. (2020), que utilizaram uma rede neural convolucional para individualizar espécies de vegetação em florestas tropicais a partir de uma imagem WorldView-2. Em ambos os casos, foram apresentados resultados satisfatórios, indicando que os métodos de DL conseguem identificar objetos com elevada acurácia mesmo em imagens de alta a altíssima resolução espacial.

Acredita-se que essa recente utilização de algoritmos de DL esteja mudando o paradigma do processamento de imagens de sensoriamento remoto, com base na grande quantidade de estudos publicados nos últimos anos, como mostra Li et al. (2022). Estudos recentes têm mostrado o potencial dessas ferramentas para o mapeamento do uso e cobertura da terra utilizando imagens de média, alta e altíssima resolução espacial. De modo geral, os resultados apresentam avanços em relação aos algoritmos tradicionais de *Machine Learning* (ML), tanto em relação aos métodos de classificação por pixel quanto por regiões, quando são realizadas segmentações das imagens (Carranza-García et al., 2019; Ge et al., 2019; Karimian et al. 2024; Nigar et al., 2024; Vali et al., 2020; Yu et al., 2022; Zhang et al., 2023; Zhao et al., 2023). Esses avanços representam, na prática, a geração de informações essenciais para realização de diagnósticos ambientais e criação de planos de ordenamento territorial, com maior acurácia e menor intervenção humana nas etapas de pós-classificação.

Embora a literatura especializada evidencie o potencial dessas técnicas para o mapeamento do uso e cobertura da terra a partir de imagens de alta resolução espacial, até o momento da redação deste artigo não foram encontradas publicações semelhantes a esta nas principais bases de dados de periódicos, como Scopus, da Web of Science, Scielo e outras. Além disso, é relevante frisar que se trata de uma imagem de alta resolução espacial, fruto de colaboração sino-brasileira e distribuída gratuitamente pelo Instituto Nacional de Pesquisas Espaciais (INPE). Diante dessas observações, foi proposto este estudo que tem por objetivo avaliar a acurácia do mapeamento do uso e cobertura da terra em um trecho da área urbana de Rio Claro (SP), realizado por meio da arquitetura U-Net, a partir da imagem CBERS-4A (WPM) com 2 m de resolução espacial. A seguir, são apresentados o referencial teórico, os principais aspectos metodológicos, as características da área de estudo, a acurácia do classificador para as classificações feitas por ML e DL, a comparação dos resultados e as considerações acerca da continuidade dessa pesquisa.

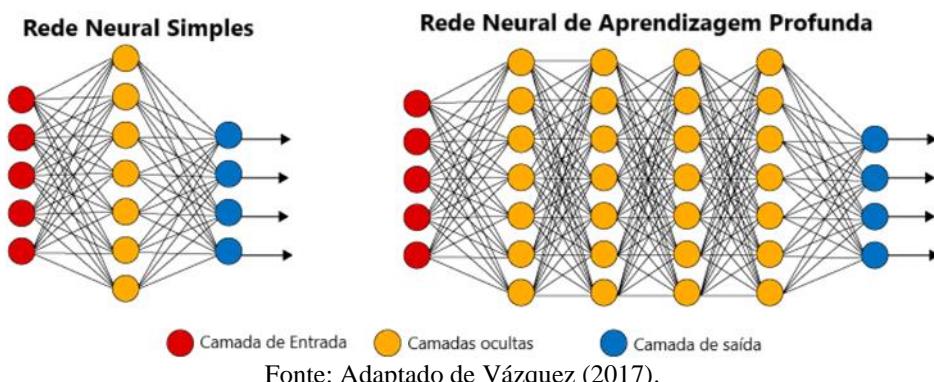
2 DEEP LEARNING

De acordo com Chollet (2021), DL é uma subárea do conhecimento que integra um conjunto de

conceitos e métodos de ML que, por sua vez, está aninhado no ramo da Inteligência Artificial (IA). O desenvolvimento das técnicas de ML, sobretudo durante a década de 1980, alterou o paradigma da programação computacional da época, pois permitiu o treinamento dos softwares para a execução de tarefas. Sistemas que inicialmente recebiam dados e regras como entrada e, por meio de uma programação prévia, definiam as respostas lógicas possíveis, passaram a receber dados e respostas como valores de entrada para retornar regras como saídas.

O que diferencia DL de ML é o método por meio do qual a máquina aprende a realizar a tarefa. Nos métodos de ML, podem ser utilizadas uma ou mais funções matemáticas que, a partir dos parâmetros definidos e do conjunto de dados de entrada, realizam o processamento para geração dos resultados. Esses modelos podem ou não ser formados por camadas, uma vez que existem abordagens neurais e não neurais. Dentre os modelos de ML considerados neurais, esses são, por vezes, chamados de “rasos” (*shallow learning*), pois a aprendizagem é realizada em apenas uma ou duas camadas do modelo. Cada camada é responsável pela aplicação de uma função matemática, sendo que ela recebe dados de entrada e os processa para gerar novas informações. Já os modelos de DL são estruturados por composições de funções, e cada uma delas gera resultados intermediários que servem como entrada para camadas sucessivas, mais profundas, também chamadas de camadas ocultas (*hidden layers*). Modelos modernos de DL oferecem até centenas de camadas, e todas elas aprendem a partir dos dados de treinamento utilizados como entrada (Chollet, 2021; Johnson & Khoshgoftaar, 2019; Ponti & Costa, 2017; Zhao et al., 2019). A Figura 1 apresenta uma comparação entre uma rede neural simples (rasa) e uma rede neural profunda.

Figura 1 – Comparativo entre uma rede neural simples (esq.) e uma rede neural profunda (dir.). Os círculos representam os neurônios da rede, organizados em camadas. Os vermelhos representam as entradas; os laranjas, os neurônios posicionados em camadas ocultas da rede; e os azuis, as saídas.



Fonte: Adaptado de Vázquez (2017).

As redes neurais artificiais se baseiam no funcionamento dos neurônios biológicos. Todo o processo é estruturado sequencialmente e a cada camada (ou neurônio) ocorre a aplicação de uma função de ativação que tem por objetivo determinar se o sinal de entrada seguirá adiante ou não. Isso é feito por meio da determinação de pesos, que serão aplicados sobre os dados de entradas, e por um limiar (*threshold*) para a ativação. Caso seja atingido o limiar mínimo, o resultado da função será 1, e o sinal seguirá adiante para ser processado pelo próximo neurônio. Caso contrário, o resultado será 0, e o sinal será interrompido. Esse processo é análogo ao funcionamento da mente humana que, a partir de critérios, condições e interesses pré-definidos, toma decisões acerca da realização de tarefas. Nesse processo, são comumente utilizadas funções sigmoidais que geram saídas apenas com sinais positivos, o que implica a propagação positiva do sinal (*feedforward*) (Chollet, 2021; Ponti & Costa, 2017; Zhang et al., 2021). Ponti e Costa (2017) explicam que, em DL, a Função Retificadora Linear (do inglês *Rectified Linear Function* – ReLU) tem sido amplamente utilizada, uma vez que torna mais simples a tarefa de treinamento do modelo, pois as funções sigmoidais saturam a partir de um determinado ponto, enquanto a ReLU funciona simplesmente como uma identidade para valores positivos.

Dentre os modelos de DL, destacam-se as Redes Neurais Convolucionais (do inglês *Convolutional Neural Network* – CNN), que consistem em um modelo de DL amplamente utilizado para a segmentação semântica de imagens (Ponti & Costa, 2017).

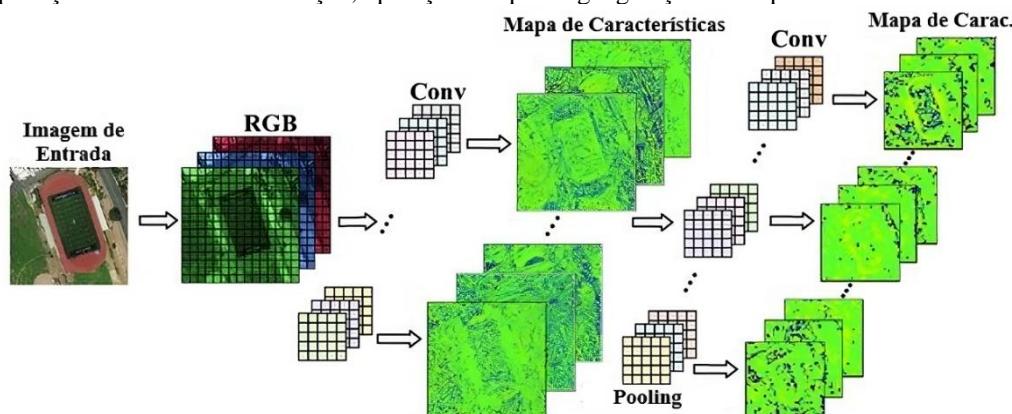
2.1 Redes Neurais Convolucionais

Conforme mencionado, as CNNs foram inspiradas no funcionamento do cérebro humano, sobretudo em relação ao processo de aprendizagem por meio da experiência e de tomada de decisões. Assim, uma rede neural estabelece que o conjunto de dados de entrada será analisado por meio de funções e critérios que estabelecerão as premissas para o aprendizado e, consequentemente, guiarão a tomada de decisão sobre como classificar os dados de entrada.

Essas redes são estruturadas por camadas convolucionais que são, basicamente, camadas de filtros espaciais lineares feitos através de matrizes denominadas máscaras ou *kernels*. Esses filtros processam as imagens de entrada gerando novas informações as quais são chamadas de mapas de características (*feature maps*). Essas informações geradas serão novamente processadas pelas camadas sucessivas, gerando um acúmulo de informações que é base para o aprendizado da rede (Goodfellow et al., 2016; Prakash & Rao, 2017). Além disso, às camadas convolucionais podem ser adicionadas operações de *pooling* que promovem a redução da dimensão espacial das imagens de entrada. Essa técnica vem sendo empregada para reduzir o tempo de processamento computacional, pois para cada aplicação de filtros são geradas novas informações, o que eleva muito o volume de dados a serem processados (Brownlee, 2019; Chollet, 2021; Ponti & Costa, 2017).

A Figura 2 ilustra, esquematicamente, o processo de aplicação de filtros convolucionais, as operações de *pooling* e a geração de mapas de características em uma rede neural convolucional.

Figura 2 – Aplicação de filtros de convolução, operações de pooling e geração de mapa de características em uma CNN.



Fonte: Adaptado de Deng, Liu & Mao (2022).

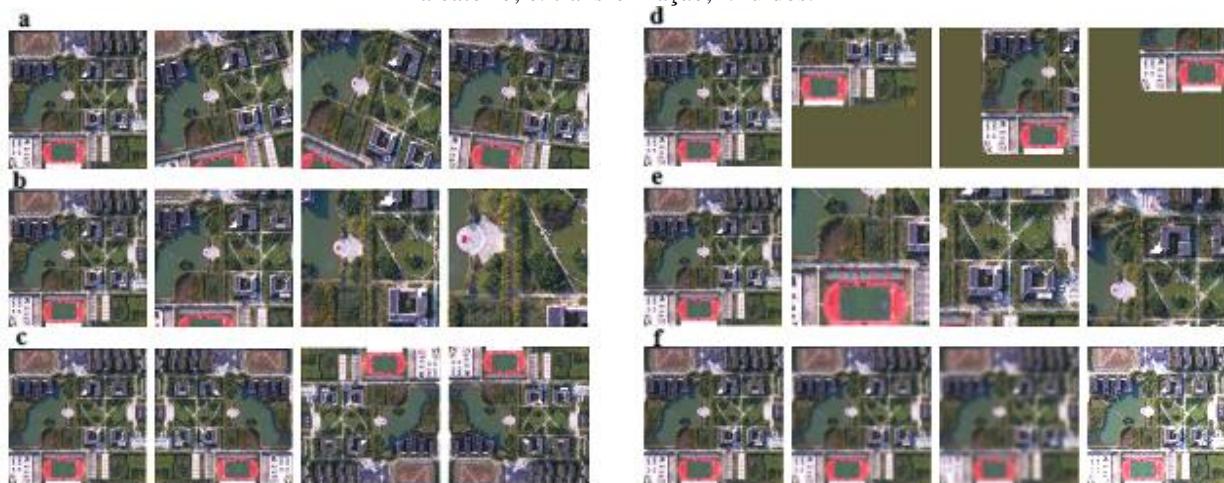
Dentre as diversas CNNs, destaca-se a U-Net, que vem apresentando excelentes resultados para a classificação de imagens de sensoriamento remoto, tanto para segmentação e extração de feições de objetos individuais, como edificações (Wagner et al., 2020a) e espécies de árvores (Wagner et al., 2020b), quanto para mapeamento do uso e cobertura da terra (Solórzano et al., 2021; Yu et al., 2022; Zhang et al., 2021).

O modelo U-Net recebe esse nome em função de sua arquitetura possuir uma estrutura em “U” que indica dois caminhos de fluxo de execução da rede. O primeiro segue uma estrutura semelhante às CNNs e consiste em um caminho de codificação dos dados, onde ocorrem operações de convolução e *pooling*. Na segunda parte, ocorrem operações de deconvolução e restauração da resolução inicial dos dados (*upconvolution*), que pode ser entendida como fase de decodificação. É realizada, ainda, a concatenação dos mapas de características criados na fase de codificação por meio de conexões diretas com os dados gerados na segunda etapa do modelo. Essas conexões são conhecidas como *skip connections* e são responsáveis pela recuperação da informação espacial da imagem original, ou seja, retornam o pixel classificado para o seu local exato em relação à imagem original e reconstruem uma imagem final com a mesma resolução da imagem inicial. Isso possibilita distinguir os elementos não somente pela característica espectral, mas, também, com base no seu contexto espacial, proporcionando segmentação e classificação com elevada precisão (Malik et al., 2021; Vali et al., 2020; Wagner et al., 2020a; Yan et al., 2022; Yu, et al., 2022).

No que diz respeito às imagens de sensoriamento remoto, a literatura tem demonstrado a necessidade de se utilizar técnicas de *data augmentation* para melhoria do treinamento do algoritmo. Essas técnicas promovem alterações e distorções na imagem original, como rotação, espelhamento, expansão e achatamento

e, como consequência, geram uma quantidade ainda maior de imagens de entrada na rede. Esse processo tem por objetivo treinar o algoritmo para reconhecer a informação desejada mesmo que ela pareça distorcida e/ou invertida (Ma et al., 2019; Vali et al., 2020; Yu et al., 2017), conforme apresenta a Figura 3. Em imagens de sensoriamento remoto, é constante a existência de elementos de mesma tipologia, como edificações ou vias, por exemplo, porém com formas e tamanhos bastante distintos entre si.

Figura 3 – Exemplo da criação de imagens aumentadas (*augmented images*). As sequências horizontais demonstram as transformações aplicadas à imagem original (primeira de cada linha), que são: a. rotação; b. zoom; c. inversão; d. corte aleatório; e. transformação; f. ruídos.



Fonte: adaptado de Hao et al. (2023).

Essa complexa arquitetura de processamento de imagens vem apresentando resultados bastante satisfatórios para os processos de extração de informações de imagens de sensoriamento remoto, conforme será demonstrado a seguir.

2.2 Aplicação do Deep Learning para o mapeamento de uso e cobertura da terra

Nos últimos anos, tem sido perceptível um aumento significativo no uso de algoritmos de DL em aplicações de sensoriamento remoto – para segmentação de imagens, extração de feições e classificação de uso e cobertura da terra (Wang et al., 2024). Esse fato se deve à melhoria na acurácia dos mapeamentos em face às técnicas predecessoras, mas também à possibilidade de adaptação dos algoritmos de DL de acordo com a demanda do usuário, o que motiva a realização de novos testes e leva a novas aplicações (Kuras et al., 2021; Ma et al., 2019; Vali et al., 2020).

Vali et al. (2020) destacam que um dos maiores desafios do emprego do DL no Sensoriamento Remoto está na criação, disponibilização e utilização da informação de *ground truth*, que consiste na informação de referência que será utilizada pelo usuário para treinar e avaliar o algoritmo classificador. Alguns autores geram essas informações por meio de processos exaustivos de vetorização manual das tipologias (Wagner et al., 2020b), enquanto outros vêm testando aplicações semiautomatizadas, como os *autoencoders* (Jozdani et al., 2019), para segmentação das imagens, o que não exclui a necessidade de uma posterior validação e classificação manual das áreas segmentadas.

Ma et al. (2019) indicam que a maior parte dos trabalhos publicados sobre mapeamento de uso e cobertura da terra, por meio de técnicas de DL, empregaram CNNs como método para classificar imagens com resolução espacial igual ou superior a 10 m. A partir da síntese organizada pelos autores, pode-se compreender o potencial dessas técnicas para realizar classificações supervisionadas de cenas complexas, como aquelas encontradas em imagens de média a alta resolução em ambientes urbanos.

Yu et al. (2022) também apresentam os desafios da classificação do uso e cobertura da terra em ambientes urbanos, indicando que a baixa qualidade do *ground truth* para o treinamento pode comprometer os resultados dos algoritmos de DL. Para lidar com essa situação, os autores utilizaram SIG para produzir um conjunto robusto de dados de referência e propuseram uma adaptação da rede U-Net que conseguiu gerar uma classificação do uso do solo urbano com, aproximadamente, 90% de precisão das classes em relação ao *ground*

truth.

Abdollahi et al. (2021) também realizaram uma adaptação da rede U-Net, mas, nesse caso, com o objetivo de realizar a segmentação de vias e edificações em imagens aéreas. Os autores implementaram uma função específica para reconhecimento dos limites dos objetos (*Boundary-Aware Loss – BAL*), o que permitiu que a rede se concentrasse em áreas complexas para a segmentação, como objetos pequenos e/ou com formas complexas. Esse processo proporcionou uma melhoria da acurácia em torno de 2% em relação à U-Net tradicional.

Carranza-García et al. (2019) compararam a performance de classificadores tradicionais de ML, como Random Forest, Support Vector Machines e K-Nearest-Neighbors com uma rede neural desenvolvida pelos próprios autores. Eles testaram o desempenho para classificação de imagens hiperespectrais e de radar e concluíram que a CNN desenvolvida apresenta maior acurácia do que os demais métodos testados em todos os testes de validação empregados.

Diante do exposto, entende-se que há um consenso na literatura de que o modelo U-Net pode gerar resultados acurados para a classificação de uso e cobertura da terra a partir de diferentes fontes de dados. Algumas adaptações, que vêm sendo testadas, demonstram pequenas melhorias em relação ao modelo tradicional, como exposto por Lv et al. (2023), que utilizaram uma rede adversarial generativa para detecção de mudanças em imagens bitemporais de sensoriamento remoto. O modelo, chamado de U-Net aprimorado (*Enhanced U-Net – E-UNet*), combinou transformação de imagens, convoluções multiescala e um módulo de auto atenção polarizada para aprimorar a identificação de áreas alteradas. Como resultado, os autores perceberam uma melhoria de, aproximadamente, 4% na acurácia global com os bancos de dados avaliados.

3 MATERIAIS E MÉTODOS

Nesta seção são apresentados os materiais e métodos utilizados neste trabalho, que consistiu na classificação supervisionada de uma imagem CBERS-4A em uma área de, aproximadamente, 9 km² localizada no município de Rio Claro (SP), visando ao mapeamento do uso e cobertura da terra.

A área está localizada na porção sul desse município e foi selecionada em função da diversidade das tipologias de uso e cobertura da terra presentes, que impõem desafios para a sua classificação em função da similaridade da resposta espectral desses elementos. Além disso, os autores do estudo vivem nessa região, o que contribuiu para a construção da verdade terrestre a partir das observações cotidianas no campo de estudo.

Inicialmente foi realizada a seleção da imagem do sensor WPM, do satélite CBERS-4A, do mês de maio de 2023, pois consistia na imagem mais recente disponível sem cobertura de nuvens. De posse da imagem, foi realizado o processo de melhoria da resolução espacial por meio da técnica *pansharpening*, que combina informações de detalhes espaciais extraídas da banda pancromática do sensor visando a produzir uma imagem de mais alta resolução, preservando as informações espectrais das bandas do sensor (Yang et al., 2017). Esse procedimento foi realizado por meio do plugin *CBERS4A downloader*, disponível no software QGIS, que permite, além da seleção e download das imagens, realizar a composição das bandas RGB+NIR e o procedimento *pansharpening* aplicando a transformação de Brovey (Klippel, 2021).

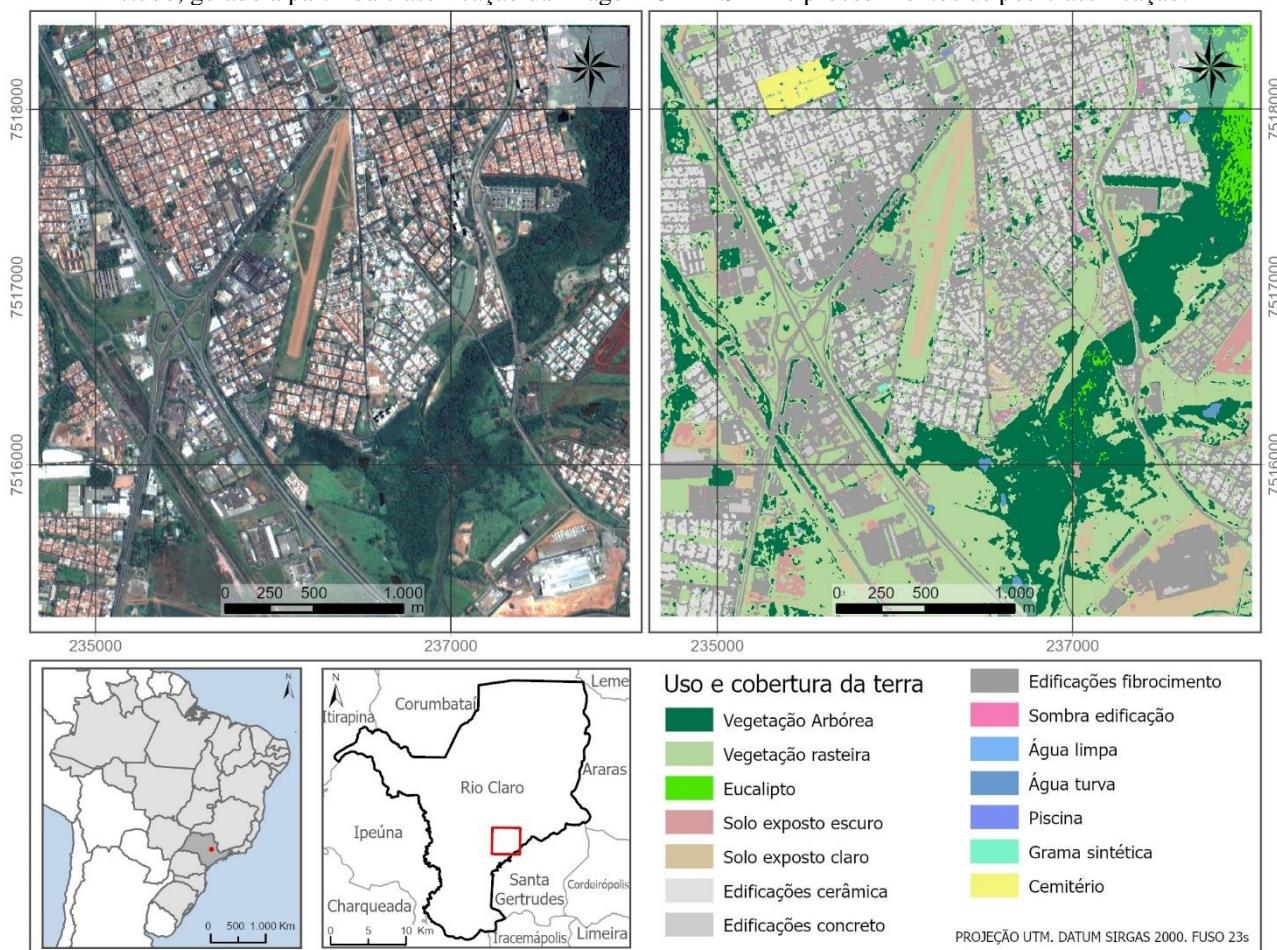
As tipologias de uso e cobertura da terra foram determinadas a partir da inspeção visual da imagem, da comparação com outras imagens gratuitas de sensores com maior resolução e de inspeção em campo sob demanda, em datas e locais variados. As tipologias passíveis de identificação na imagem CBERS-4A foram: Vegetação arbórea, Vegetação rasteira, Eucalipto, Solo exposto (escuro), Solo exposto (claro), Edificações (telhas de cerâmica), Edificações (lajes de concreto), Edificações (telhas de fibrocimento) / Asfalto (agrupados na mesma classe), Sombras de edificação, Água limpa (sem turbidez), Água turva, Piscina, Grama sintética e Cemitério. Essas tipologias podem ser observadas na Figura 4 que também apresenta a localização da área de estudo.

O arquivo de referência para a classificação da imagem, bem como para a validação dos resultados – ou *ground truth* (Figura 4) –, foi gerado a partir da classificação supervisionada da mesma imagem, e de procedimentos de pós classificação, incluindo a aplicação de filtro majoritário (máscara de 3 x 3 pixels) e a edição manual dos pixels classificados. Para isso, considerou-se a experiência relatada por Magalhães (2024), que identificou um melhor desempenho do algoritmo *Support Vector Machine* (SVM) frente ao *Random Trees*

(RT), ambos disponíveis no software ArcGIS Pro. A partir dessa observação, foram realizados testes de classificação com o SVM utilizando baixa amostragem (10 polígonos amostrais por classe), média amostragem (50 polígonos amostrais por classe) e alta amostragem (100 polígonos amostrais por classe). Além disso, para cada uma das amostragens, foram testados os valores 125, 250, 500 (*default*) e 1.000 para o “Número Máximo de Amostras por Classe” (NMAC), parâmetro a ser definido pelo usuário. Ao definir um NMAC, o algoritmo evita que classes com muitas amostras dominem o processo de treinamento, garantindo um equilíbrio no aprendizado. Isso foi importante neste trabalho, pois havia desequilíbrio entre as classes presentes na área de estudo, como o Cemitério, a Piscina e a Grama sintética, que ocupavam áreas ínfimas, e a Vegetação arbórea e a Vegetação rasteira, que ocupavam grandes áreas. Dessa maneira, foram realizadas, ao todo, 12 classificações supervisionadas da imagem (3 níveis de amostragem x 4 NMAC).

Os resultados obtidos nessa etapa foram avaliados visualmente, comparando a classificação com a imagem CBERS-4A, e quantitativamente, por meio da verificação da acurácia a partir de pontos aleatórios na imagem. Essa estratégia foi adotada, neste momento, pois ainda não se dispunha do *ground truth* para uma verificação mais robusta. O objetivo desse processo, foi o de identificar o melhor resultado do classificador SVM para este pudesse ser utilizado como base de comparação com a classificação gerada pela U-Net e, com isso, verificar se o modelo de DL promove avanços em relação ao algoritmo de ML. Além disso, esse resultado serviu como um dado inicial a ser editado para a construção do *ground truth*.

Figura 4 – Localização e características da Área de Estudo. À esquerda, apresenta-se a composição colorida de cores verdadeiras da imagem CBERS-4A (WPM) com 2 m de resolução espacial. À direita, o *ground truth* da Área de Estudo, gerado a partir da classificação da imagem CBERS-4A e procedimentos de pós-classificação.



Elaboração: Os autores (2025).

Essa análise inicial da acurácia foi realizada com auxílio da ferramenta *Accuracy Assessment Points*, adotando-se a estratégia de amostragem randômica estratificada, na qual cada classe recebe um número de pontos proporcional à sua área total. Foi indicado um número mínimo de 323 pontos, definidos com base na formulação de Foody (2009) que está apresentada na Equação 1. Para esta pesquisa, foi estabelecido um nível

de confiança de 95%, correspondendo a um valor Z de 1,96. Devido à quantidade de classes, área total analisada e estratégia adotada, foram gerados, automaticamente, 394 pontos distribuídos randomicamente sobre a imagem classificada.

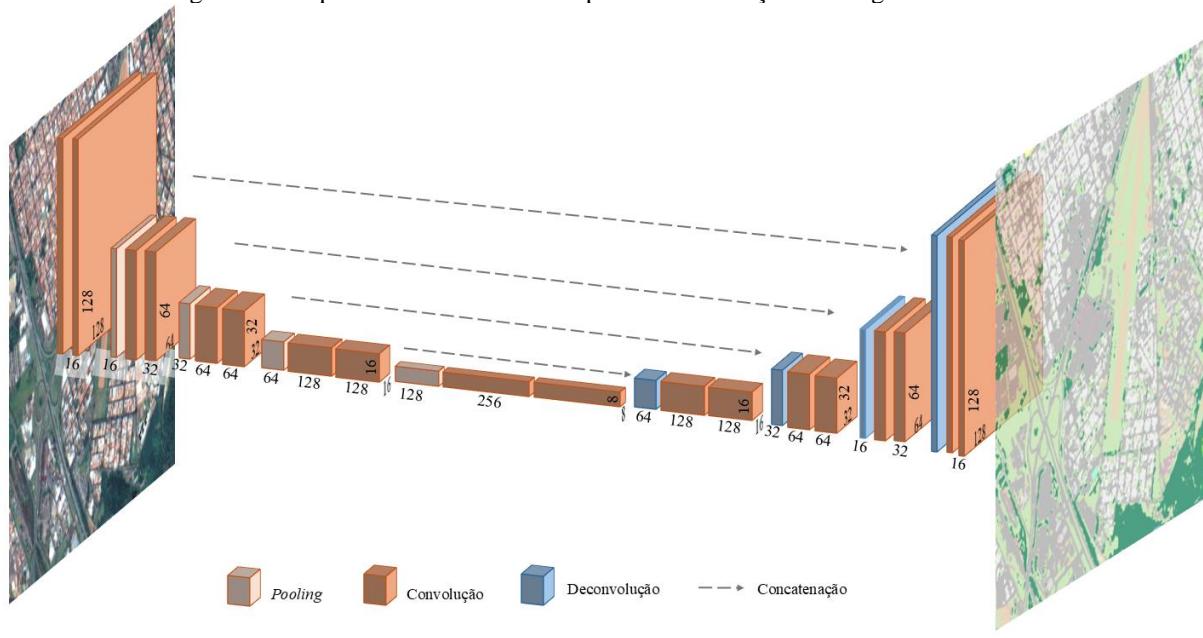
$$N = \left(\frac{Z^2 \cdot P \cdot (1 - P)}{E^2} \right) = \left(\frac{1,96^2 \cdot 0,70 \cdot (1 - 0,70)}{0,05^2} \right) \approx 323 \quad (1)$$

Onde N é o tamanho de amostras necessário; Z é o valor crítico da distribuição normal padrão para o nível de confiança desejado; P é a proporção esperada de casos corretamente classificados; E é o erro máximo tolerável na estimativa da proporção.

Os resultados obtidos nessas análises indicaram uma melhor acurácia para a classificação gerada com média amostragem (50 amostras por classe) e com o NMAC igual a 500, o que gerou uma acurácia global igual a 0,68 e coeficiente Kappa igual a 0,64. De posse desse dado, foram realizados os procedimentos de pós-classificação visando construir o *ground truth*. O primeiro passo foi a aplicação do Filtro Majoritário (*Majority Filter*), o que proporcionou a redução de efeito salpicado na imagem, causado pela classificação de pixels isolados. Na sequência, foi realizado um exaustivo processo de edição manual dos pixels visando ajustar a imagem classificada ao máximo às tipologias observadas na imagem do sensor WPM. Esse processo de edição considerou a imagem CBERS-4A como referência principal, determinando os limites das classes. Contudo, contou ainda com a utilização de imagens de melhor resolução, de visualização livre via web, de imagens de *street view* de aplicativos de navegação e de observações *in loco* em momentos diversos, uma vez que os autores vivem nessa região.

De posse do *ground truth*, foi realizada a classificação da imagem por meio da arquitetura U-Net. A rede foi estruturada através de um script em Python, utilizando-se as bibliotecas de DL *Keras* e *Tensor Flow*. A imagem original, de 3 canais (RGB), foi subdividida em fragmentos (*tiles*) de 128 x 128 pixels para viabilizar o processamento computacional. Na fase de codificação do modelo, foram aplicados filtros de convolução de máscara 3 x 3 e operações de *pooling* para a redução da dimensão das imagens, gerando novas imagens de 64², 32², 16² e 8² pixels. Na fase de decodificação, foram empregadas as operações de deconvolução e concatenação dos dados àqueles gerados na fase inicial do modelo, o que garante o posicionamento correto do pixel classificado dentro da imagem. A arquitetura elaborada está representada na Figura 5.

Figura 5 – Arquitetura U-Net utilizada para a classificação da imagem CBERS-4A.

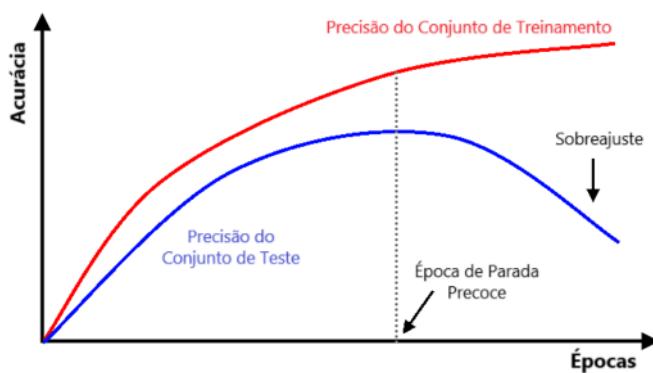


Elaboração: Os autores (2025).

O treinamento do algoritmo é realizado por “épocas” que, de um modo simplificado, consiste no número de passagens completas do conjunto de dados dentro do processo de aprendizagem. O uso de poucas

épocas pode causar o *underfitting*, ou seja, o algoritmo pode não aprender toda a informação possível e gerar uma classificação com baixa acurácia. Por outro lado, o uso de muitas épocas pode levar ao *overfitting*, ou seja, o algoritmo pode “aprender demais” e começar a classificar os ruídos dos dados e não o sinal desejado e, com isso, reduzir a acurácia da classificação (DSA, 2022). Portanto, a quantidade de épocas ideal deve ser definida a partir da observação da acurácia *versus* a quantidade de épocas, buscando interromper o treinamento no ponto máximo de acurácia, antes de se iniciar o processo de *overfitting*. Esse processo é conhecido como *early stopping* e está ilustrado na Figura 6. Neste estudo, o valor ideal identificado para o treinamento, obtido a partir de testes, foi o de 200 épocas.

Figura 6 – Definição da Época de Parada Precoce (*Early Stopping*).



Fonte: Adaptado de DSA (2022).

O processamento das imagens no ambiente Python não requer a informação de georreferenciamento que é utilizada nos softwares de SIG, como o ArcGIS Pro, sendo que as imagens são classificadas individualmente. Portanto, para reconstituir o georreferenciamento da imagem classificada, de modo a se posicionar corretamente sobre a imagem CBERS-4A, foram preservados os arquivos auxiliares de georreferenciamento que são gerados automaticamente pelo ArcGIS Pro, a saber, as extensões .pgw, .xml e .ovr. A associação das imagens classificadas com os arquivos auxiliares se deu por meio da preservação das nomenclaturas dos arquivos, de modo que os resultados puderam ser espacializados corretamente.

O resultado da classificação, gerado também em *tiles* de 128 x 128 pixels, foi espacializado no ArcGIS Pro de modo a criar um mosaico da classificação com georreferenciamento correspondente à imagem original. Esse processo foi realizado para permitir a comparação dessa classificação com aquela gerada por meio do algoritmo SVM (sem as edições da pós-classificação), com o objetivo de verificar se a classificação pelo modelo U-Net proporciona resultados mais acurados do que o modelo de ML. A avaliação da acurácia das classificações foi realizada por meio das métricas Acurácia Global (*Overall Accuracy – OA*), *Precision*, *Recall*, *F1 Score* e *Intersect Over Union* (IoU), geradas a partir da comparação da verdade terrestre com a imagem classificada por meio de uma matriz de confusão. Tais métricas são amplamente utilizadas, atualmente, para a verificação da acurácia de classificação de imagens de sensoriamento remoto e estão descritas em trabalhos como os de Maxwell, Warner e Guillén (2021a, 2021b), dentre outros. De modo simplificado, a partir da matriz de confusão, os pixels da imagem classificada são identificados como: Verdadeiros Positivos (VP), que são aqueles previstos corretamente; Verdadeiros Negativos (VN), que são aqueles previstos erroneamente; Falsos Positivos (FP), que não foram previstos de maneira correta; e os Falsos Negativos (FN), que não foram previstos de maneira equivocada. A partir dessa matriz, são estabelecidas formulações que calculam a assertividade do classificador em relação a um conjunto de dados.

A Acurácia Global, por exemplo, calcula a quantidade de acertos do modelo em relação ao total da amostra, por meio da divisão de todos os pixels classificados como verdadeiros (VP + VN) pelo conjunto total. Essa métrica dá uma noção do quanto certo o modelo é. Contudo, apresenta limitações em cenários com desbalanceamento de classes, o que pode mascarar problemas de desempenho, especialmente quando uma ou mais classes dominam o conjunto de dados (Morales-Barquero et al., 2019). Para lidar com esse problema, outras métricas podem ser utilizadas, como a *Precision*, que avalia a proporção de previsões corretas em relação ao conjunto das previsões positivas por meio do cálculo da razão entre VP e a soma de VP e FP. Essa medida nos dá a noção do quanto podemos confiar em um modelo quando ele prevê que um pixel pertence a

uma determinada classe. Já a *Recall* é uma razão entre o número total de pixels classificados corretamente (VP), para uma determinada classe, e o total dessa mesma classe presente no *ground truth* (VP + FN). Isso nos fornece uma noção da taxa de detecção do modelo, ou seja, a capacidade que ele tem para identificar corretamente uma classe em relação à realidade. O *F1 Score* combina as métricas *Precision* e *Recall* em uma métrica única usando sua média harmônica e é útil em problemas onde há desequilíbrio entre classes, equilibrando omissões e comissões. Por fim, a IoU avalia a sobreposição entre a predição e o *ground truth*, através do cálculo da razão entre a área de interseção (VP) e a união das áreas predita e verdadeira (VP + FN + FP) (Maxwell, Warner e Guillén, 2021a, 2021b; Morales-Barquero et al., 2019).

4 RESULTADOS E DISCUSSÕES

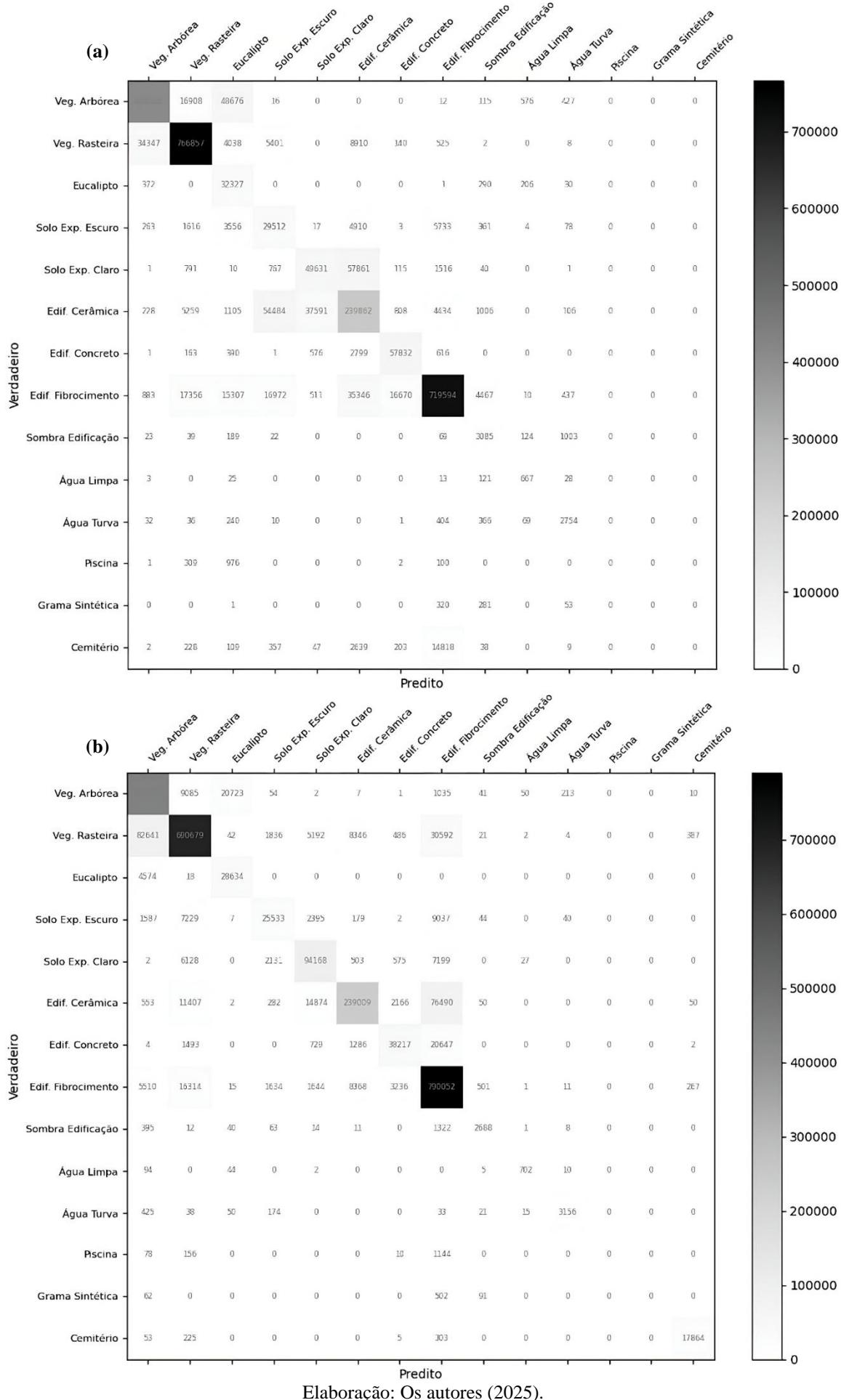
A Figura 7 apresenta as matrizes de confusão obtidas para as classificações realizadas com o SVM (7a) e a U-Net (7b) em relação ao *ground truth*. No eixo Y estão indicadas as classes verdadeiras, registradas no *ground truth*; no eixo X estão indicadas as classes preditas pelos classificadores. Os valores presentes nas matrizes correspondem ao número de pixels classificados em cada classe. Os valores mais altos estão registrados em tons mais escuros e, portanto, indicam as classes com maior predominância na área de estudo. Quando posicionados na diagonal, no encontro das classes homônimas, indicam acertos dos classificadores. Os valores registrados em outras posições das matrizes, indicam erros do classificador e contribuem para identificar as classes onde houve maior confusão para a classificação.

A comparação entre as matrizes nos indica uma assertividade ligeiramente superior da U-Net em relação ao SVM, para classificação da imagem testada. Destaca-se a capacidade do classificador por DL para detectar classes muito específicas, como o Cemitério, o que se mostrou inviável para o SVM. Atribui-se esse resultado ao fato de a U-Net trabalhar, também, com informações contextuais e texturais da imagem, o que gera uma capacidade superior para distinguir classes com características espectrais muito semelhantes, como o caso do Cemitério e das Edificações com lajes de concreto. Nessa mesma linha de interpretação, chama atenção, ainda, uma maior assertividade da classe Solo exposto claro, que na classificação do SVM houve elevada confusão com as Edificações com telhas de cerâmica. Novamente, elementos com características espectrais semelhantes, porém com formas, tamanhos e textura diversas. Observa-se que as classes Grama sintética e Piscina não foram identificadas por ambos os classificadores, provavelmente por ocuparem uma área ínfima dentro da área de estudo.

Já para as classes com características espectrais distintas das demais, como as Vegetações arbórea e rasteira e as Edificações com telhas de fibrocimento (agrupada com o Asfalto), o desempenho dos classificadores é bastante semelhante, sendo que em alguns casos há uma maior assertividade do SVM em relação à U-Net. Essa situação fica clara quando são analisados os resultados das métricas *Precision*, *Recall*, *F1 Score* e IoU para cada classe individualmente, que são apresentadas na Figura 8.

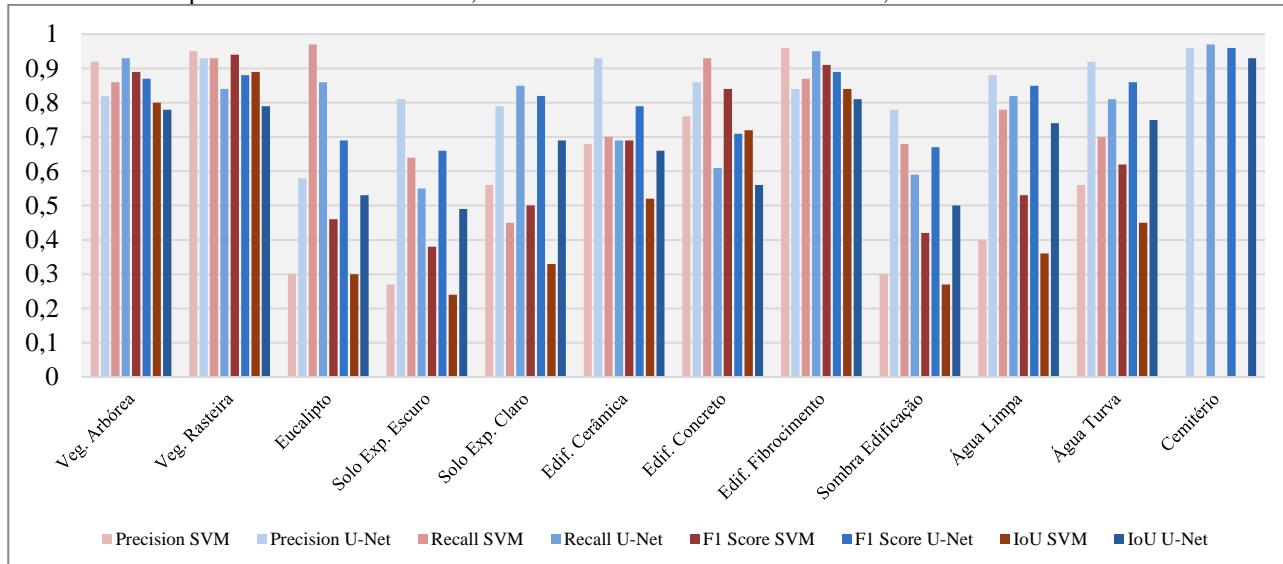
Das doze classes identificadas por algum dos classificadores, em quatro delas o desempenho do SVM é igual ou superior ao da U-Net, o que reforça a qualidade desse classificador para mapeamento do uso e cobertura da terra, como detalhado por Sheykhou et al (2020). Por outro lado, explicita sua limitação para distinção de classes com características espectrais muito semelhantes, o que dificulta sua aplicação para mapeamento de uso da terra em áreas urbanas a partir de imagens de alta resolução espacial. A análise do gráfico (Figura 8) nos permite perceber, ainda, que o SVM apresenta um desempenho irregular, desbalanceado, para o conjunto analisado, ao passo que a U-Net apresenta um resultado mais constante e homogêneo. Isso reforça o seu potencial para classificação de cenas complexas, compostas por imagens de alta resolução com heterogeneidade de classes, como as áreas urbanizadas, assim como identificado por Ma et al. (2019). A irregularidade observada para o SVM compromete os resultados para a análise da performance geral do classificador, conforme apresenta a Tabela 1. Dentre as métricas utilizadas, destaca-se o resultado 0,48 (SVM) e 0,72 (U-Net), obtido na métrica *Precision*, a qual indica a capacidade do classificador de prever corretamente uma classe.

Figura 7 – Matriz de confusão das classificações supervisionadas da imagem CBERS-4A. (a) Apresenta o resultado do SVM; (b) Apresenta o resultado da U-Net.



Elaboração: Os autores (2025).

Figura 8 – Resultados das métricas *Precision*, *Recall*, *F1 Score* e *IoU* para cada classe de uso e cobertura da terra mapeada. Em tons vermelhos, os resultados do SVM. Em tons azuis, os resultados da U-Net.



Elaboração: Os autores (2025).

Tabela 1 – Síntese da acurácia das classificações pela U-Net e pelo SVM.

Dado	OA	IoU	Precision	Recall	F1 Score
U-Net	0,86	0,59	0,72	0,68	0,69
SVM	0,84	0,41	0,48	0,61	0,51

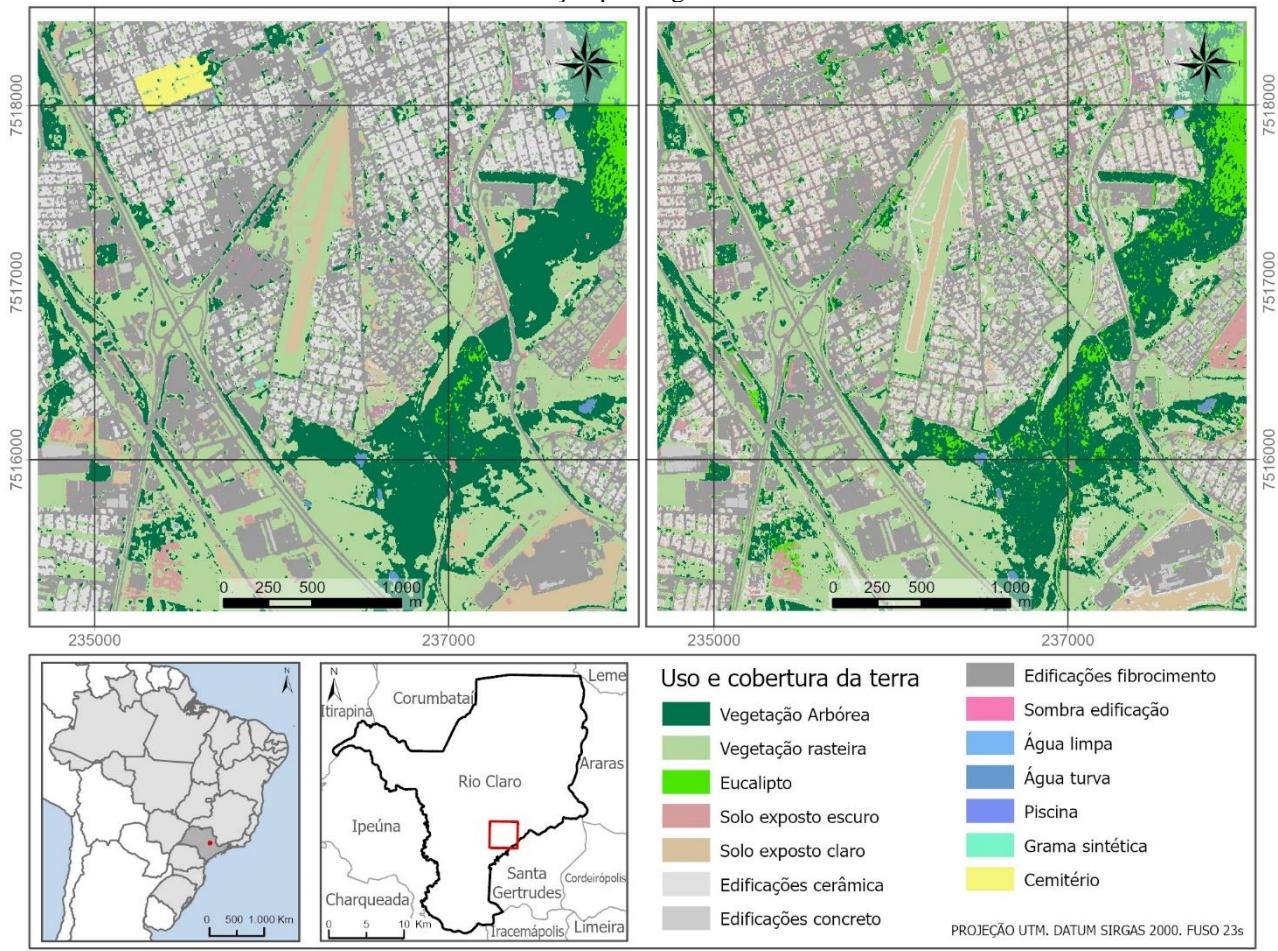
Elaboração: Os autores (2025).

Avaliando graficamente o resultado produzido pelo SVM (Figura 9), percebe-se que as tipologias que ocupam grandes áreas e de modo contínuo foram bem representadas, como as vegetações arbóreas e rasteiras. Em contraposição, os elementos menores e com características espectrais semelhantes a outros materiais, como o caso das telhas de cerâmica e dos solos expostos, obtiveram erros importantes na identificação pelo classificador, o que compromete o uso visando ações de planejamento e gestão territorial.

Atribui-se esse resultado à grande quantidade de informações espaciais disponíveis na imagem, em função de sua resolução espacial de 2 m, aliada a uma baixa quantidade de informações espectrais do sensor, que é composto por bandas do visível e do infravermelho próximo. Nesse sentido, o incremento das bandas espectrais da região do infravermelho de ondas médias, como aquelas disponíveis no satélite Landsat, poderia contribuir para um melhor resultado da classificação supervisionada desse tipo de imagem.

Ao analisar, graficamente, os resultados da classificação pelo modelo U-Net (Figura 10), reforça-se a percepção da maior assertividade desse classificador para a tarefa proposta. A imagem corrobora os dados observados nas métricas de acurácia, pois, visualmente, há uma forte coincidência das classes de uso e cobertura da terra preditas em relação ao *ground truth*. Destaca-se a separação das classes onde houve maior confusão no SVM, como as Edificações com telhas de cerâmica e o Cemitério. Essas tipologias, que ocupavam pequenas áreas e que possuíam características espectrais muito semelhantes a outros materiais, foram representadas de maneira bastante precisa por esse algoritmo.

Figura 9 – Classificação do uso e cobertura da terra a partir do algoritmo SVM. À esquerda o *ground truth* e à direita a classificação pelo algoritmo SVM.

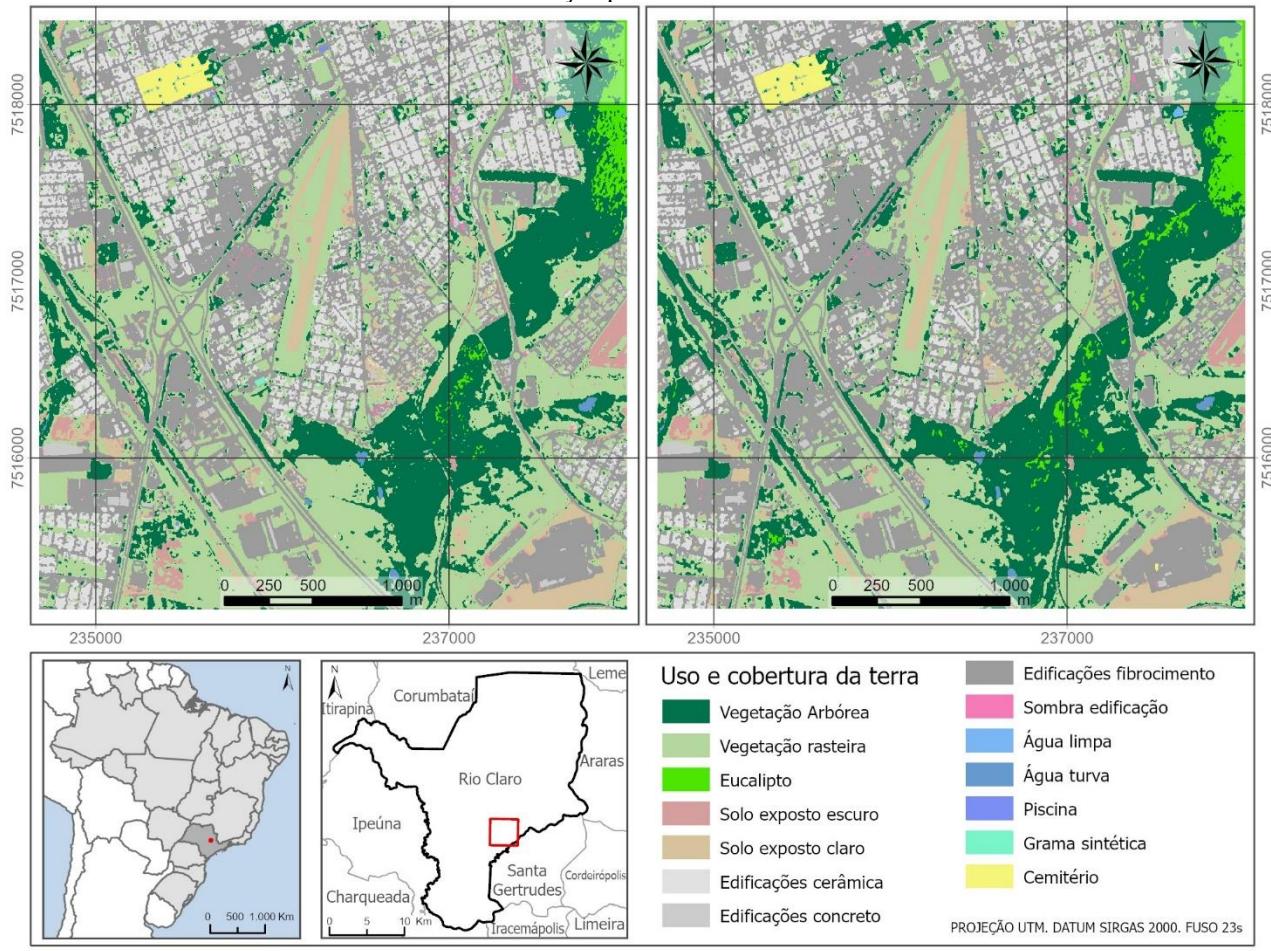


Elaboração: Os autores (2025).

Entende-se que os resultados aqui apresentados são, ainda, preliminares, frutos de testes realizados em um trecho da área urbana do município de Rio Claro (SP) e, por isso, acredita-se que a acurácia do modelo proposto poderá ser incrementada a partir da utilização de um volume maior de dados de entrada para treinamento do algoritmo. Contudo, considera-se que os resultados são promissores e que demonstram o potencial dessa técnica para mapeamento de uso e cobertura da terra em áreas urbanas a partir das imagens gratuitas geradas pelo sensor WPM, do CBERS-4A. Os resultados encontrados estão de acordo com aqueles identificados na literatura especializada, como apresentado por Carranza-García et al. (2019), que, ao comparar o desempenho de algoritmos de ML e DP para a classificação supervisionada de imagens, concluem que a CNN supera as demais técnicas testadas, alcançando um alto nível de desempenho em todos os conjuntos de dados testados. Jozdani et al. (2019) também realizaram uma comparação do desempenho de diferentes modelos de DL e de ML para o mapeamento do uso e cobertura da terra em ambiente urbano, a partir de imagens de altíssima resolução espacial (50 cm e 30 cm) e concluíram que a arquitetura *Multilayer Perceptron* obteve os resultados de classificação mais precisos. Os autores ressaltam, ainda, que o SVM produziu resultados de classificação altamente precisos, demonstrando a versatilidade de algoritmos ML. Na mesma direção, Solórzano et al. (2021) afirmam que o uso da CNN para aplicações de observação da Terra tem proporcionado melhorias para a classificação de uso e cobertura da terra detalhadas, destacando que, embora tenham treinado a U-net com um pequeno conjunto de dados, sua performance superou à do algoritmo RT.

Por fim, Zhang et al. (2021) discutem que o mapeamento do uso e cobertura da terra é um trabalho complexo no âmbito do sensoriamento remoto, sobretudo a partir do incremento da resolução espacial das imagens. Diante do desafio de se classificar imagens por métodos baseados em pixel (que geram muito ruído) ou baseados em objetos (que requerem grande intervenção humana), os autores propõem a utilização de uma CNN de ponta a ponta e os resultados mostram que o modelo CNN proposto superou os métodos tradicionais baseados em pixels na classificação de imagens de alta resolução.

Figura 10 – Classificação do Uso e Cobertura da Terra pelo modelo U-Net. À esquerda o *ground truth* e à direita a classificação pelo modelo U-Net.



Elaboração: Os autores (2025).

5 CONSIDERAÇÕES FINAIS

O artigo apresentou os primeiros resultados da pesquisa que estuda a utilização de algoritmos de DL para a classificação supervisionada de imagens de sensoriamento remoto, tendo como objeto inicial de estudo as imagens do sensor WPM, do satélite CBERS-4A, com resolução espacial de 2 m.

Os resultados apontam para um potencial de utilização dessa técnica para o mapeamento de um significativo número de tipologias de uso e cobertura da terra a partir de imagens de alta resolução, pois obteve acurácia superior ao algoritmo SVM, disponível no ArcGIS Pro, para o conjunto de dados testados. Considera-se que são resultados promissores, uma vez que o modelo foi capaz de distinguir algumas classes que possuem características espectrais muito semelhantes, como as Edificações com lajes de concreto e o Cemitério, e os Solos expostos e as Edificações com telhas de cerâmica. Entende-se que os resultados são, ainda iniciais, e que novos testes com um volume maior de dados de entrada podem proporcionar a geração de resultados mais acurados. Contudo, neste momento, é relevante ressaltar o potencial da técnica empregada em imagens gratuitas e de alta resolução visando ao mapeamento do uso e cobertura da terra de municípios brasileiros, o que pode gerar dados relevantes para apoio ao planejamento e à gestão territorial. Como exemplo, cita-se a possibilidade de realização de monitoramento da cobertura vegetal dentro da malha urbana, de análise da permeabilidade do solo, de estudos da formação de ilhas de calor, de considerações acerca do parcelamento do solo urbano, dentre outros.

A análise da acurácia dos resultados vai de encontro com aqueles identificados na literatura internacional especializada, que demonstram elevada acurácia desses classificadores para imagens de sensoriamento remoto. Destaca-se, ainda, a capacidade de extraer informações em imagens de alta resolução espacial em ambientes urbanos, o que é uma tarefa complexa no processamento de imagens de sensoriamento remoto.

A pesquisa encontra-se, ainda, em andamento e os próximos passos consistem na replicação do processo para mapear a área total do município de Rio Claro (SP). Esse processo demanda grande esforço humano, para a geração do *ground truth*, e computacional, para o processamento de uma grande área. Esses desafios são comuns ao se trabalhar com DL, como pode ser observado em trabalhos correlatos, como os de Vali et al. (2020) e Yu et al. (2022). Assim, a produção desses dados poderá proporcionar a criação de um banco de dados para o treinamento de classificadores DL para imagens CBERS-4A, que poderá ser utilizado para classificar imagens de outras datas e localidades, desde que apresentem características semelhantes. Isso, pois o algoritmo não depende da localização espacial das amostras de treinamento para operar, possibilitando que o processo seja replicado em outras localidades e contribuindo para a produção de dados sobre uso e cobertura da terra a partir de imagens gratuitas de alta resolução.

Referências

- Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2021). Multi-object segmentation in complex urban scenes from high-resolution remote sensing data. *Remote Sensing*, 13(18), 3710. <https://doi.org/10.3390/rs13183710>
- Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Alizadeh Moghaddam, S. H., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., & Brisco, B. (2020). Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5326–5350. <https://doi.org/10.1109/JSTARS.2020.3011956>
- Braga, J. R. G., Peripato, V., Dalagnol, R., Ferreira, M. P., Tarabalka, Y., Aragão, L. E. O. C., Campos Velho, H. F. de, Shiguemori, E. H., & Wagner, F. H. (2020). Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, 12(8), 1288. <https://doi.org/10.3390/rs12081288>
- Brownlee, J. (2019). *Deep learning for computer vision: Image classification, object detection and face recognition in Python*. Machine Learning Mastery.
- Carranza-García, M., García-Gutiérrez, J., & Riquelme, J. (2019). A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), 274. <https://doi.org/10.3390/rs11030274>
- Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207. <https://doi.org/10.1109/JPROC.2016.2598228>
- Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning Publications.
- Data Science Academy (DSA). (2022). *Deep learning book*. <https://www.deeplearningbook.com.br/>
- Deng, T., Liu, X. & Mao, G. (2022). Improved YOLOv5 Based on Hybrid Domain Attention for Small Object Detection in Optical Remote Sensing Images. *Electronics*, 11(17), 2657. <https://doi.org/10.3390/electronics11172657>
- Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, 30(20), 5273-5291. <https://doi.org/10.1080/01431160802582899>
- Ge, P., He, J., Zhang, S., Zhang, L., & She, J. F. (2019). An integrated framework combining multiple human activity features for land use classification. *ISPRS International Journal of Geo-Information*, 8(2), 90. <https://doi.org/10.3390/ijgi8020090>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hao, X., Liu, L., Yang, R., Yin, L., Zhang, L. & Li, X. (2023). A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition. *Remote Sensing*, 15(3), 827. <https://doi.org/10.3390/rs15030827>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big*

- Data*, 6(1), 27. <https://doi.org/10.1186/s40537-019-0192-0>
- Jozdani, S. E., Johnson, B. A., & Chen, D. (2019). Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. *Remote Sensing*, 11(14), 1713. <https://doi.org/10.3390/rs11141713>
- Karimian, R., Rangzan, K., Karimi, D. & Einali, G. (2024). Spatiotemporal Monitoring of Land Use-Land Cover and Its Relationship with Land Surface Temperature Changes Based on Remote Sensing, GIS, and Deep Learning. *J Indian Soc Remote Sens.*, 52, 2461. <https://doi.org/10.1007/s12524-024-01958-3>
- Klippel, S. (2022). *CBERS4A downloader QGIS plugin*. <https://github.com/sandroklippel/cbers4a>
- Kuras, A., et al. (2021). Hyperspectral and Lidar data applied to the urban land cover machine learning and neural-network-based classification: A review. *Remote Sensing*, 13(17), 3393. <https://doi.org/10.3390/rs13173393>
- Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., Ben, G., & Gao, Y. (2022). Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. *Remote sensing*, 14(10), 2385. <https://doi.org/10.3390/rs14102385>
- Lv, Z., Huang, H., Sun, W., Lei, T., Benediktsson, J. A. & Li, J. (2023). Novel Enhanced UNet for Change Detection Using Multimodal Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*, 20, 1. 10.1109/LGRS.2023.3325439
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.05.019>
- Magalhães, D. M. (2024). Avaliação da acurácia da classificação supervisionada de imagens de sensoriamento remoto utilizando o software ArcGIS Pro. In N. I. Ladwig, T. Sutil, C. H. R. da Silva, & B. Giacom (Eds.), *Planejamento e gestão territorial* (1st ed., pp. 141-164). Pedro e João. <http://dx.doi.org/10.51795/9786526514276>
- Malik, K., Robertson, C., Braun, D., & Greig, C. (2021). U-Net convolutional neural network models for detecting and quantifying placer mining disturbances at watershed scales. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102510. <https://doi.org/10.1016/j.jag.2021.102510>
- Maxwell, A. E., Warner, T. A. & Guillén, L. A. (2021a). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sensing*, 13(13), 2450. <https://doi.org/10.3390/rs13132450>
- Maxwell, A. E., Warner, T. A. & Guillén, L. A. (2021b). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sensing*, 13(13), 2591. <https://doi.org/10.3390/rs13132591>
- Morales-Barquero, L., Lyons, M. B., Phinn, S. R., & Roelfsema, C. M. (2019). Trends in Remote Sensing Accuracy Assessment Approaches in the Context of Natural Resources. *Remote Sensing*, 11(19), 2305. <https://doi.org/10.3390/rs11192305>
- Nigar, A., Li, Y., Jat Baloch, M. Y., Alrefaei, A. F. & Almutairi, M. H. (2024). Comparison of machine and deep learning algorithms using Google Earth Engine and Python for land classifications. *Frontiers in Environmental Science*, 12, 01. 10.3389/fenvs.2024.1378443
- Odenyo, V. A. O., & Pettry, D. E. (1977). Land-use mapping by machine processing of LANDSAT-1 data. *Photogrammetric Engineering and Remote Sensing*, 43(4), 515–523.
- Pabi, O. (2007). Understanding land-use/cover change process for land and environmental resources use management policy in Ghana. *GeoJournal*, 68(4), 369–383. <https://doi.org/10.1007/s10708-007-9120-7>
- Parente, L., Taquary, E., Silva, A. P., Souza, C., & Ferreira, L. (2019). Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, 11(23), 2881. <https://doi.org/10.3390/rs11232881>
- Picoli, M. C. A., Simoes, R., Chaves, M., Santos, L. A., Sanchez, A., Soares, A., Sanches, I. D., Ferreira, K. R., & Queiroz, G. R. (2020). CBERS data cube: A powerful technology for mapping and monitoring

- Brazilian biomes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3–2020, 533–539. <https://doi.org/10.5194/isprs-annals-V-3-2020-533-2020>
- Ponti, M. A., & Costa, G. B. P. (2017). Como funciona o deep learning. In: *Tópicos em gerenciamento de dados e informações* (1st ed., p. 31). SBC.
- Prakash, D. P., & Rao, A. S. K. R. (2017). *Deep learning cookbook: Solve complex neural net problems with TensorFlow, H2O, and MXNET*. Packt.
- Sheykhou, M., Mahdianpari, M., Ghanbari, H., Mohammadmanesh, F., Ghamisi, P. & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308. [10.1109/JSTARS.2020.3026724](https://doi.org/10.1109/JSTARS.2020.3026724).
- Solórzano, J. V., Mas, J. F., Gao, Y., & Gallardo-Cruz, J. A. (2021). Land use land cover classification with U-Net: Advantages of combining Sentinel-1 and Sentinel-2 imagery. *Remote Sensing*, 13(18), 3600. <https://doi.org/10.3390/rs13183600>
- Vali, A., Comai, S., & Matteucci, M. (2020). Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2495. <https://doi.org/10.3390/rs12152495>
- Vázquez, F. (2017, December 21). *Deep learning made easy with deep cognition*. Becoming Human: Artificial Intelligence Magazine. <https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351>
- Wagner, F. H., Dalagnol, R., Tarabalka, Y., Segantine, T. Y. F., Thomé, R., & Hirye, M. C. M. (2020a). U-Net-Id, an instance segmentation model for building extraction from satellite images — Case study in the Joanópolis City, Brazil. *Remote Sensing*, 12(10), 1544. <https://doi.org/10.3390/rs12101544>
- Wagner, F. H., Dalagnol, R., Tagle Casapia, X., Streher, A. S., Phillips, O. L., Gloor, E., & Aragão, L. E. O. C. (2020b). Regional mapping and spatial distribution analysis of canopy palms in an Amazon forest using deep learning and VHR images. *Remote Sensing*, 12(14), 2225. <https://doi.org/10.3390/rs12142225>
- Wang, L., Zhang, M., Gao, X., & Shi, W. (2024). Advances and challenges in deep learning-based change detection for remote sensing images: A review through various learning paradigms. *Remote Sensing*, 16, 804. <https://doi.org/10.3390/rs16050804>
- Yan, C., Fan, X., Fan, J., & Wang, N. (2022). Improved U-Net remote sensing classification algorithm based on multi-feature fusion perception. *Remote Sensing*, 14(5), 1118. <https://doi.org/10.3390/rs14051118>
- Yang, Y., Wan, W., Huang, S., Lin, P., & Que, Y. (2017). A novel pan-sharpening framework based on matting model and multiscale transform. *Remote Sensing*, 9(4), 391. <https://doi.org/10.3390/rs9040391>
- Yu, J., Zeng, P., Yu, Y., Yu, H., Huang, L., & Zhou, D. (2022). A combined convolutional neural network for urban land-use classification with GIS data. *Remote Sensing*, 14(5), 1128. <https://doi.org/10.3390/rs14051128>
- Zhang, X., Du, L., Tan, S., Wu, F., Zhu, L., Zeng, Y., & Wu, B. (2021). Land use and land cover mapping using RapidEye imagery based on a novel band attention deep learning method in the Three Gorges Reservoir area. *Remote Sensing*, 13(6), 1225. <https://doi.org/10.3390/rs13061225>
- Zhang, P., Wu, Y., Li, C., Li, R., Yao, H., Zhang, Y., Zhang, G. & Li, D. (2023). National-Standards- and Deep-Learning-Oriented Raster and Vector Benchmark Dataset (RVBD) for Land-Use/Land-Cover Mapping in the Yangtze River Basin. *Remote Sensing*, 14(15), 3907. <https://doi.org/10.3390/rs15153907>
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2019.2897602>
- Zhao, S., Tu, K., Ye, S., Tang, H., Hu, Y. & Xie, C. (2023). Land Use and Land Cover Classification Meets Deep Learning: A Review. *Remote Sensing*, 23(21), 8966. <https://doi.org/10.3390/s23218966>

Contribuição dos Autores

D. M. Magalhães foi responsável pela Conceptualização, Administração do projeto, Supervisão, Redação – minuta inicial e Redação – revisão e edição. J. P. Souza contribuiu com a Curadoria dos dados e Análise formal (SIG). E. A. G. França contribuiu com a Curadoria de dados e Análise formal (Python).

Conflitos de Interesse

Os autores declaram que não há conflitos de interesse.

Biografia do autor principal



Danilo Marques de Magalhães, natural de Belo Horizonte-MG (1983), é Bacharel (2010), Mestre (2013) e Doutor (2021) em Geografia pela UFMG, tendo realizado estágio sanduíche na UniBo (Itália). É Professor Assistente Doutor do Departamento de Geografia e Planejamento Ambiental da Unesp de Rio Claro (SP) e do Programa de Pós-Graduação em Geografia da mesma instituição. Coordena o Grupo MAPEAR de Pesquisa e Extensão, orientando trabalhos na área de Sensoriamento Remoto e Sistemas de Informações Geográficas. Atualmente, coordena o projeto de pesquisa "Uso de Deep Learning para mapeamento de uso e cobertura da terra a partir de imagens CBERS-4A".



Esta obra está licenciada com uma Licença [Creative Commons Atribuição 4.0 Internacional](#) – CC BY. Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original.