

Sistema de detecção de erosões em plantações de cana-de-açúcar utilizando métodos de Inteligência Artificial

Isabella Beatriz Montania Dominguez da Silva^a, Vanessa Reis Abdonor Schmidt Casemiro^a, Eloisa Arnt Oliveira dos Santos^a, Wesley Nunes Gonçalves^a

^a*Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Av. Costa e Silva, s/n, Campo Grande, 79070-900, MS, Brazil*

Abstract

Este trabalho propõe o desenvolvimento de um sistema para detecção de erosão em plantações de cana-de-açúcar, utilizando técnicas de segmentação semântica baseadas em Inteligência Artificial (IA) e aprendizado profundo. A erosão do solo representa uma ameaça significativa para a produtividade agrícola, acarretando perda de nutrientes, comprometimento do desenvolvimento das plantas e impactos econômicos e ambientais. Dessa forma, a identificação precoce de áreas suscetíveis à erosão é essencial para mitigar seus efeitos e assegurar a sustentabilidade das plantações. Para alcançar esse objetivo, imagens aéreas capturadas por drones são processadas para a geração de ortofotos, que, posteriormente, são segmentadas em patches de múltiplas resoluções (256×256 , 512×512 e 1024×1024 pixels). Esses patches são utilizados para o treinamento de modelos de segmentação semântica, capacitando as redes neurais a identificar automaticamente áreas de erosão com alta precisão. O processo de treinamento é estruturado em três etapas principais: geração e organização dos conjuntos de treino, validação e teste, seguido pelo treinamento efetivo dos modelos. Em seguida, são aplicadas métricas quantitativas, como IoU (Intersection over Union) e F1-Score, para avaliação do desempenho dos modelos. Os métodos desenvolvidos neste projeto possibilitam uma análise rápida e precisa das condições do solo, reduzindo a necessidade de inspeções manuais, que geralmente demandam mais tempo e estão sujeitas à subjetividade. Além disso, a aplicação dessas técnicas contribui para o monitoramento contínuo das plantações, otimizando estratégias de manejo e diminuindo os custos associados à manutenção do solo, promovendo maior

sustentabilidade no cultivo de cana-de-açúcar.

This work proposes the development of a system for detecting soil erosion in sugarcane plantations, using semantic segmentation techniques based on Artificial Intelligence (AI) and deep learning. Soil erosion represents a significant threat to agricultural productivity, resulting in nutrient loss, impaired plant development, and economic and environmental impacts. Therefore, the early identification of areas susceptible to erosion is essential to mitigate its effects and ensure the sustainability of plantations. To achieve this goal, aerial images captured by drones are processed to generate orthophotos, which are subsequently segmented into patches of multiple resolutions (256×256 , 512×512 , and 1024×1024 pixels). These patches are used for the training of semantic segmentation models, enabling neural networks to automatically identify erosion areas with high precision. The training process is structured into three main stages: generation and organization of the training, validation, and test sets, followed by the effective training of the models. Subsequently, quantitative metrics such as IoU (Intersection over Union) and F1-Score are applied to evaluate the models' performance. The methods developed in this project allow for a quick and precise analysis of soil conditions, reducing the need for manual inspections, which generally require more time and are subject to subjectivity. Furthermore, the application of these techniques contributes to the continuous monitoring of plantations, optimizing management strategies and reducing costs associated with soil maintenance, thus promoting greater sustainability in sugarcane cultivation.

Keywords: Detecção de Erosão, Plantações de Cana-de-açúcar, Segmentação Semântica, Aprendizado Profundo, Imagens de Drones

1. Introdução

A cana-de-açúcar, no Brasil, destaca-se como uma das principais culturas do agronegócio, exercendo papel essencial na produção de açúcar, etanol e bioenergia, além de contribuir significativamente para a geração de milhões de empregos diretos e indiretos no país [1]. O Brasil é reconhecido como líder mundial na produção de cana, o que reforça sua importância econômica e estratégica. Nesse contexto, práticas adequadas de manejo e monitoramento do solo tornam-se indispensáveis para assegurar a produtividade e a sustentabilidade de toda a cadeia produtiva da cana-de-açúcar.

A erosão do solo é um dos principais desafios enfrentados pela agricultura moderna, impactando diretamente a produtividade e a sustentabilidade das plantações. A erosão em plantações de cana-de-açúcar refere-se ao desgaste, transporte e deposição de partículas de solo causados pela água, vento ou atividades humanas. Esse fenômeno pode resultar em perda de solo superficial rico em nutrientes, diminuição da capacidade de retenção de água, podendo levar a uma redução significativa na fertilidade e profundidade cultivável, e compactação do solo, prejudicando o desenvolvimento das raízes das plantas [2, 3, 4]. Essas alterações, por sua vez, resultam em quedas expressivas na produtividade agrícola [4]. Identificar a erosão no início é crucial para mitigar esses impactos negativos e preservar a integridade do solo e a viabilidade econômica das plantações [5, 6]. Se não controlada, a erosão pode causar perdas econômicas substanciais, exigir investimentos adicionais em fertilizantes e correções de solo, além de impactar negativamente os recursos hídricos locais devido ao assoreamento.

A utilização de drones para a identificação de erosão oferece uma solução inovadora e eficiente, permitindo a coleta de dados precisos sobre as condições do solo. As imagens aéreas capturadas por drones fornecem uma visão abrangente das plantações, possibilitando a identificação rápida de áreas afetadas pela erosão que seriam difíceis de detectar por métodos tradicionais. Esta abordagem melhora a precisão do monitoramento, reduz os custos operacionais e o tempo necessário para a inspeção do solo. Com a crescente demanda por práticas agrícolas mais sustentáveis e eficientes, a implementação de tecnologias de drones para a detecção de erosão em plantações de cana-de-açúcar justifica-se como uma medida essencial para assegurar a produtividade e a sustentabilidade a longo prazo [7, 8].

Atualmente, a identificação da erosão é geralmente realizada por meio de inspeções de campo e análises visuais, métodos que são demorados, trabalhosos e frequentemente subjetivos [9]. Essas metodologias tradicionais dependem fortemente da experiência do avaliador e são limitadas em sua capacidade de cobrir grandes áreas de forma eficiente [10]. Além disso, a coleta de dados em campo pode ser influenciada por condições climáticas adversas e pela dificuldade de acesso a certas áreas, o que pode resultar em uma avaliação incompleta ou imprecisa [11]. As lacunas nestas metodologias incluem a falta de precisão na detecção precoce da erosão, a baixa eficiência na cobertura de grandes áreas e a subjetividade inerente às avaliações visuais. Portanto, há uma necessidade crescente de técnicas mais avançadas e automatizadas, como o uso de drones, para superar essas limitações e fornecer

dados mais precisos e abrangentes sobre a condição do solo [7].

O objetivo geral deste projeto é utilizar métodos de segmentação de imagens baseados em aprendizagem profunda para a identificação e monitoramento de erosão em plantações de cana-de-açúcar utilizando imagens de drones. Apesar dos avanços recentes em aprendizagem profunda, uma lacuna significativa ainda persiste na literatura: compreender como a resolução espacial dos patches utilizados em modelos de segmentação afeta o desempenho de diferentes arquiteturas, especialmente aquelas baseadas em transformers, que exploram contextos globais, em comparação às convolucionais, que enfatizam contextos locais, na identificação precisa de áreas de erosão. Este trabalho busca preencher essa lacuna ao avaliar o desempenho dos modelos em três resoluções distintas (256x256, 512x512 e 1024x1024 pixels). Os resultados demonstram que modelos convolucionais, como FCN e DeepLabV3+, têm um desempenho superior em resoluções menores, onde o detalhamento local é crucial. Em contrapartida, modelos baseados em transformers, como o SegFormer, apresentam maior robustez em resoluções maiores, evidenciando a importância da captura do contexto global. Esses resultados fornecem importantes subsídios para seleção de arquiteturas e resoluções apropriadas para o monitoramento eficiente e preciso da erosão em grandes áreas agrícolas.

2. Materiais e Métodos

2.1. Conjunto de Dados

No presente trabalho, 13 ortofotos foram geradas a partir de imagens capturadas por drones sobrevoando áreas de cultivo de cana-de-açúcar de diferentes regiões no Brasil. Os lugares exatos serão omitidos por questões de privacidade. Os drones estavam equipados com câmeras de alta resolução, permitindo a captura detalhada das áreas de interesse. O processo de ortorretificação foi utilizado para corrigir as distorções geométricas inerentes às imagens aéreas, resultando em ortofotos que mantêm a escala uniforme em toda a extensão da imagem, como se fossem vistas diretamente de cima. O GSD das ortofotos foi de aproximadamente 3cm.

As imagens aéreas retratam áreas agrícolas afetadas por diferentes tipos de erosão, que podem ser classificadas visualmente em *sulcos*, *ravinas* e *erosão laminar*:

- **Erosão laminar:** ocorre de forma mais difusa, removendo camadas finas de solo sem formar canais visíveis, como mostrado na Figura 1a.

Este processo é mais difícil de identificar, mas tem um impacto significativo na qualidade do solo a longo prazo.

- **Erosão em sulcos:** caracteriza-se por canais alongados e estreitos formados pela concentração do fluxo de água em superfícies inclinadas ou mal manejados, onde a água encontra pouca resistência para infiltrar e acaba removendo o solo superficial (Figura 1b). Este tipo de erosão é comum em terrenos agrícolas com declives acentuados.
- **Erosão em ravinas:** Representa um estágio mais avançado da erosão, onde há a formação de valas profundas, como ilustrado na Figura 1c. As ravinas se desenvolvem principalmente em áreas onde a água escoar com maior intensidade, agravando a perda de solo.

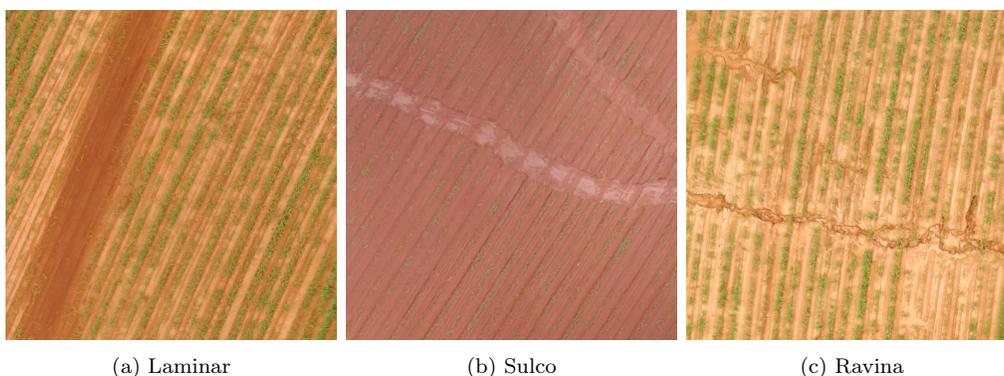


Figure 1: Tipos de erosão presentes nas imagens.

Além dos diferentes tipos de erosão, existem desafios adicionais para a análise dos solos devido a fatores como a baixa resolução e a cobertura vegetal esparsa, como pode ser observado nas Figuras 2a e 2b. Esses fatores dificultam a distinção de detalhes como sulcos e variações na textura. A falta de contraste entre o solo exposto e a erosão torna a identificação dos danos menos clara, especialmente se for uma erosão laminar ou inicial. Portanto, pequenos sulcos ou sinais de degradação podem passar despercebidos, enquanto o ângulo da imagem pode não capturar adequadamente as variações no relevo.



(a) Baixa resolução espacial.

(b) Cobertura vegetal esparsa.

Figure 2: Desafios adicionais presente nas ortofotos.

Após a obtenção das ortofotos, elas foram analisadas e anotadas manualmente por um especialista para identificar regiões afetadas pela erosão do solo. Para isso, foram delineados polígonos para separar as áreas de erosão das áreas intactas, permitindo uma segmentação precisa que foi utilizada em análises subsequentes com técnicas de aprendizado de máquina. Um exemplo é apresentado na Figura 3, que exhibe a ortofoto completa e um destaque ampliado da região afetada pela erosão, anotada em vermelho pelo especialista.

2.2. Metodologia

A metodologia deste trabalho foi estruturada em quatro etapas principais, com o objetivo de preparar, treinar e avaliar modelos de segmentação semântica para detecção de áreas de erosão em imagens aéreas. A seguir, são descritos os passos realizados:

1. Geração dos conjuntos de dados multiescala:

As imagens ortorretificadas foram divididas em patches de diferentes tamanhos (256×256 , 512×512 e 1024×1024), permitindo a análise da influência da escala na detecção da classe de erosão. Cada patch foi gerado com sua respectiva máscara de segmentação.

2. Organização em conjuntos de treino, validação e teste:

Para cada escala, os patches foram organizados em três conjuntos distintos. O conjunto de treino foi utilizado para o ajuste dos pesos



Figure 3: Ortofoto completa da área analisada, destacando-se em detalhe (zoom) a região com erosão identificada e anotada em vermelho por especialista.

dos modelos; o conjunto de validação auxiliou na escolha dos hiperparâmetros e na prevenção de *overfitting*; e o conjunto de teste foi reservado exclusivamente para a avaliação final de desempenho.

3. **Treinamento dos modelos:**

Foram treinados três modelos distintos de segmentação semântica: FCN [12], DeepLabV3+ [13] e SegFormer [14]. Cada modelo foi treinado separadamente para cada uma das escalas de patch, utilizando as mesmas condições de treinamento e critérios de parada.

4. **Geração das métricas usando o conjunto de teste:**

Após o treinamento, os modelos foram avaliados com o conjunto de teste, utilizando métricas como IoU (*Intersection over Union*) e F1-score para quantificar o desempenho na segmentação da classe de erosão.

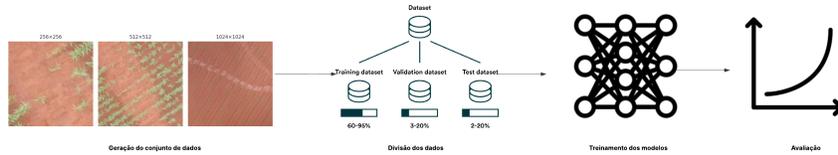


Figure 4: Fluxograma

2.2.1. Conjunto de Dados Multiescala

A partir das ortofotos anotadas, o conjunto de dados (dataset) foi criado e consiste em recortes, chamados de patches, extraídos das ortofotos. Para avaliar a influência da escala nos modelos de aprendizagem profunda, os patches foram gerados em três resoluções: 256x256, 512x512 e 1024x1024 pixels, visando fornecer diferentes níveis de detalhe e contextos espaciais. O total de patches gerados para cada resolução foi distribuído conforme Tabela 1.

| Resolução | Total de Patches |
|-----------|------------------|
| 256x256 | 39.051 |
| 512x512 | 13.058 |
| 1024x1024 | 2.848 |

Table 1: Distribuição total dos patches gerados para cada resolução.

A erosão aparece em poucas regiões das ortofotos, tornando esse problema desbalanceado. Para determinar a proporção de erosão nas imagens, o número médio de pixels classificados como erosão foi comparado ao total de pixels. No conjunto de 256x256 pixels, a porcentagem de pixels de erosão foi de 1,01% a 1,32%, e nas imagens de 512x512 pixels, a erosão ocupou entre 1,09% e 1,25% da imagem. Em contrapartida, no *dataset* de maior resolução, 1024x1024 pixels, o percentual de erosão foi menor, entre 0,92% e 1,02%. Esses resultados sugerem que imagens maiores tendem a ter uma menor concentração de áreas com erosão. Embora o percentual de erosão tenda a diminuir em imagens de maior resolução, a variabilidade da erosão entre as imagens aumenta substancialmente. Esse comportamento pode ser explicado pela captura de áreas mais amplas em imagens maiores, que podem incluir tanto regiões erodidas quanto intactas em maior diversidade. De qualquer forma, a baixa porcentagem de erosão presente em cada patch representa um desafio para os modelos de aprendizagem profunda, especialmente

considerando que o desbalanceamento de classes é um fator crítico para o desempenho desses modelos.

2.2.2. Amostragem

O processo de amostragem dos patches foi conduzido de forma a garantir diversidade e representatividade. Para a divisão do dataset em conjuntos de treinamento, validação e teste, foram utilizados os seguintes passos:

1. **Amostragem Aleatória:** As imagens de cada conjunto, correspondentes as resoluções de patches (256, 512, 1024), foram embaralhadas aleatoriamente para garantir a representatividade e evitar viés nas amostras. Esse embaralhamento foi realizado isoladamente para cada resolução.
2. **Divisão dos Conjuntos:** Após o embaralhamento, as imagens foram divididas em três conjuntos: treinamento (70%), validação (10%) e teste (20%). Essa divisão foi baseada em proporções predefinidas para assegurar que cada conjunto contenha uma quantidade significativa de amostras. Além disso, foi garantido que não houvesse vazamento de dados entre os conjuntos de treinamento, validação e teste, assegurando a independência entre eles.
3. **Organização dos Patches:** As imagens RGB e suas respectivas máscaras de anotação foram organizadas nas pastas correspondentes de treinamento, validação e teste dentro de cada diretório de resolução.

Os patches gerados para cada tamanho de imagem foram distribuídos conforme apresentado a seguir:

| Resolução | Treinamento | Validação | Teste | Total |
|-----------|-------------|-----------|-------|--------|
| 256x256 | 27.759 | 3.764 | 7.528 | 39.051 |
| 512x512 | 8.144 | 1.700 | 3.214 | 13.058 |
| 1024x1024 | 1.774 | 376 | 698 | 2.848 |

Table 2: Distribuição dos patches nos conjuntos de treinamento, validação e teste para cada resolução de dataset.

2.2.3. Modelos de Segmentação Semântica

Para realizar a tarefa de segmentação da classe de erosão nas imagens aéreas, foram utilizados três modelos distintos de segmentação semântica

baseados em redes neurais profundas: FCN, DeepLabV3+ e SegFormer. A seguir, cada modelo é descrito brevemente.

Fully Convolutional Network (FCN) [12]: O modelo FCN (Fully Convolutional Network) foi um dos primeiros a adaptar redes convolucionais para tarefas de segmentação semântica, substituindo camadas totalmente conectadas por convoluções e permitindo a produção de mapas de segmentação de tamanho variável. Sua arquitetura é composta por um caminho de down-sampling, que extrai e interpreta o contexto da imagem, e um caminho de upsampling, que reconstrói espacialmente os mapas de características por meio de camadas de deconvolução. Além disso, o FCN utiliza skip connections para recuperar informações espaciais finas perdidas durante o down-sampling. O FCN utilizado neste trabalho emprega o backbone ResNet101, o que garante extração de características profundas e robustas. Pela sua simplicidade e eficiência, o FCN se tornou uma base importante para o desenvolvimento de arquiteturas modernas de segmentação.

DeepLabV3+ [13]: O DeepLabV3+ é uma evolução do modelo DeepLab, projetada para capturar informações contextuais em múltiplas escalas através de convoluções dilatadas e do módulo ASPP (Atrous Spatial Pyramid Pooling). Assim como o modelo FCN, o DeepLabV3+ também utiliza o backbone ResNet101, potencializando a capacidade de extração de atributos complexos. A principal inovação da versão “plus” é a inclusão de um decoder que refina os contornos segmentados, aumentando a precisão em regiões de borda e eliminando a necessidade de pós-processamento adicional, como o DenseCRF. Além disso, o DeepLabV3+ aprimora o ASPP ao explorar diferentes taxas de amostragem espacial, melhorando a detecção de objetos em múltiplas escalas. Sua robustez e capacidade de lidar com grande variação espacial garantem excelente desempenho em aplicações como a análise de imagens de erosão.

SegFormer [14]: O SegFormer possui uma arquitetura de segmentação semântica baseada em Transformers, reconhecida pela alta eficiência e desempenho competitivo, mesmo utilizando menos parâmetros. Neste trabalho, foi empregado o backbone MiT-B5 (Mix Transformer-B5), uma variante poderosa da família SegFormer, que dispensa o uso de positional encoding, o que aumenta a robustez em cenários com variação de resolução entre treino e teste, como nos conjuntos multiescala usados neste trabalho. Seu decoder leve, baseado em MLPs, combina informações locais e globais de maneira eficaz. Mesmo com uma estrutura simples, o SegFormer apresenta resultados expressivos, como o modelo SegFormer-B5, que atinge 84,0% de mIoU no

dataset Cityscapes, sendo especialmente indicado para aplicações que exigem boa generalização com restrições computacionais.

2.3. Protocolo Experimental

O treinamento dos modelos FCN, DeepLabV3+ e SegFormer foi realizado utilizando a biblioteca MMSegmentation no ambiente Google Colab. Cada modelo foi treinado separadamente para cada conjunto de dados multiescala (patches 256×256 , 512×512 e 1024×1024), respeitando a mesma estrutura experimental para possibilitar uma comparação justa entre eles.

Os treinamentos foram realizados a partir de modelos pré-treinados (pre-trained weights) disponíveis na plataforma OpenMMLab, sendo ajustados para o novo conjunto de dados através de arquivos de configuração modificados. Os modelos foram otimizados utilizando o algoritmo SGD (*stochastic gradient descent*), com taxa de aprendizagem inicial definida em 0.01 com decaimento de 0.0005. Cada modelo foi treinado por um total de 20000 iterações. A cada 2000 iterações, era realizada uma validação no conjunto de validação para acompanhar a evolução do desempenho ao longo do treinamento. Para garantir a reprodutibilidade dos resultados, foi fixado o valor da semente aleatória como zero. A função de perda utilizada foi a entropia cruzada, adequada para a segmentação semântica binária proposta.

Após a conclusão do treinamento, foi realizada a inferência sobre o conjunto de teste. Para isso, os modelos treinados foram carregados utilizando suas respectivas configurações e checkpoints. Cada imagem do conjunto de teste foi processada individualmente, gerando duas saídas: a máscara de segmentação predita e a visualização da segmentação sobreposta à imagem original.

A máscara predita foi salva em formato .png, utilizando uma paleta de cores que preserva a distinção entre o fundo e a erosão, possibilitando posteriormente a avaliação quantitativa dos resultados. Paralelamente, foram geradas imagens visuais com as previsões sobrepostas às imagens originais, permitindo uma avaliação qualitativa do desempenho dos modelos. Todo o processo foi automatizado para percorrer todas as imagens do conjunto de teste, salvando os resultados em diretórios separados para organização.

Para a avaliação quantitativa do desempenho dos modelos, foram utilizadas as métricas IoU (Intersection over Union) e F1-Score. A métrica IoU, elucidada na Equação 1, foi empregada para medir a qualidade da sobreposição entre as áreas de erosão preditas e as áreas reais presentes nas

anotações. Nesse contexto, TP (True Positives) representa os pixels corretamente classificados como erosão, FP (False Positives) os pixels incorretamente classificados como erosão, e FN (False Negatives) os pixels de erosão não detectados.

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

Para a métrica F1-Score, são utilizados dois conceitos base, Precisão (Precision) e Revocação (Recall), conforme as Equações 2 e 3, respectivamente. Nesse cenário, a Precisão mede a proporção de pixels classificados como erosão que realmente pertencem à classe de erosão, enquanto a Revocação indica a proporção de pixels de erosão corretamente identificados em relação ao total real.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

O F1-Score, por sua vez, corresponde à média harmônica entre Precisão e Revocação e foi utilizado para avaliar a capacidade dos modelos em lidar com o desequilíbrio entre a quantidade de pixels de erosão e de fundo nas imagens — aspecto especialmente relevante, considerando que a erosão ocupa uma pequena porção da área total analisada. Na Equação 4, demonstra-se o cálculo do F1-Score .

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

A combinação dessas métricas permitiu uma análise abrangente da capacidade dos modelos em segmentar corretamente as áreas de interesse, considerando a sobreposição espacial e o equilíbrio entre erros de omissão e comissão.

3. Resultados e Discussão

Nesta seção, são apresentados e discutidos os resultados obtidos a partir da metodologia proposta. A discussão está estruturada em duas partes principais, considerando os dados das duas tabelas geradas. A primeira parte apresenta os resultados quantitativos dos modelos avaliados diretamente nos

patches, destacando métricas como F1 e IoU para diferentes tamanhos de dataset. A segunda parte explora uma análise qualitativa, evidenciando as principais vantagens e limitações de cada modelo, bem como sua capacidade de representar áreas de erosão com precisão.

3.1. Resultados Quantitativos

Os resultados quantitativos dos três métodos de segmentação avalia para a identificação de áreas de erosão em imagens capturadas por drones estão apresentados nas Tabelas 3 e 4. A Tabela 3 exibe as métricas gerais, considerando todas as classes presentes nos patches (fundo e erosão), enquanto a Tabela 4 foca exclusivamente nos resultados para as áreas de erosão, permitindo uma análise detalhada do desempenho específico dos modelos na classe de interesse.

Os modelos SegFormer, FCN e DeepLabV3+ apresentam variações de desempenho conforme a resolução dos patches, refletindo diferenças nas arquiteturas e capacidades de generalização. Os resultados foram avaliados sob duas perspectivas: uma geral, considerando todas as classes nos patches, e outra focada exclusivamente nas áreas de erosão.

| Modelo | SegFormer | | FCN | | DeepLabV3+ | |
|-------------------|---------------|---------------|---------------|---------------|------------|--------|
| Resolução Patches | F1 | IoU | F1 | IoU | F1 | IoU |
| 256 | 0.9713 | 0.9646 | 0.9786 | 0.9713 | 0.9773 | 0.9698 |
| 512 | 0.9780 | 0.9681 | 0.9728 | 0.9596 | 0.9693 | 0.9562 |
| 1024 | 0.9489 | 0.9254 | 0.9334 | 0.9089 | 0.8630 | 0.8411 |

Table 3: Média das métricas F1 e IoU para os modelos SegFormer, FCN e DeepLabV3+, considerando todas as classes nos patches, nos tamanhos de dataset 256, 512 e 1024.

| Modelo | SegFormer | | FCN | | DeepLabV3+ | |
|-------------------|---------------|---------------|---------------|---------------|---------------|--------|
| Resolução Patches | F1 | IoU | F1 | IoU | F1 | IoU |
| 256 | 0.6978 | 0.6225 | 0.7856 | 0.7033 | 0.7856 | 0.7014 |
| 512 | 0.8739 | 0.8097 | 0.8397 | 0.7546 | 0.8255 | 0.7403 |
| 1024 | 0.7624 | 0.6528 | 0.6899 | 0.5762 | 0.4293 | 0.3273 |

Table 4: Média das métricas F1 e IoU para os modelos SegFormer, FCN e DeepLabV3+, considerando apenas a classe erosão, nos tamanhos de dataset 256, 512 e 1024.

Na análise geral, o modelo FCN obteve o melhor desempenho com resolução 256, alcançando a maior F1 (0.9786). Porém, conforme a resolução dos

patches aumentou, o SegFormer passou a liderar, alcançando F1 de 0.9780 com resolução 512 e 0.9489 com resolução 1024. O DeepLabV3+, embora competitivo nas resoluções menores, apresentou uma queda acentuada com resolução 1024, com F1 de apenas 0.8630, evidenciando dificuldades em lidar com contextos mais amplos no patch.

Quando analisados os resultados focados exclusivamente em áreas de erosão, o SegFormer demonstrou maior consistência em resoluções maiores, alcançando F1 de 0.8739 com resolução 512 e 0.7624 com resolução 1024. O FCN teve um bom desempenho com resolução 256 (F1 de 0.7856), mas apresentou uma redução significativa em resoluções maiores (F1 de 0.6899 com resolução 1024). O DeepLabV3+, embora competitivo com resolução 256, caiu drasticamente com resolução 1024, com F1 de apenas 0.4293, indicando dificuldades substanciais em lidar com informações contextuais mais amplas.

A comparação entre as resoluções dos patches revela que a resolução 256 oferece maior detalhamento local, favorecendo a identificação de padrões pequenos, como áreas de erosão pontuais. No entanto, sua limitação de contexto prejudica a análise em imagens aéreas maiores. A resolução 512, por sua vez, representa um equilíbrio entre detalhamento e contexto, sendo onde o SegFormer e o FCN alcançam os melhores desempenhos em erosão. Já a resolução 1024, embora ofereça maior contexto global, sacrifica o detalhamento, impactando negativamente os modelos baseados em convoluções, como o DeepLabV3+.

As diferenças de arquitetura explicam o comportamento observado. Modelos baseados em convoluções, como FCN e DeepLabV3+, enfrentam dificuldades em datasets com resoluções maiores devido às limitações de suas janelas fixas, que restringem a integração de informações contextuais globais. Por outro lado, o SegFormer, baseado em transformadores, destacou-se como o modelo mais robusto em datasets com resoluções maiores. Sua capacidade de capturar o contexto global da imagem permite equilibrar contexto e detalhamento de forma eficiente, garantindo um desempenho superior tanto na média geral quanto na detecção específica de áreas de erosão.

Assim, a escolha da resolução do dataset e do modelo ideal depende do equilíbrio necessário entre contexto e detalhamento para cada aplicação. Enquanto resoluções menores e modelos baseados em convoluções são mais indicados para tarefas que requerem maior detalhamento local, modelos baseados em transformadores, como o SegFormer, são mais adequados para análises que exigem maior contexto global, especialmente em imagens aéreas de grande escala.

3.2. Resultados Qualitativos

Os resultados foram também analisados qualitativamente conforme Figura 5 e Figura 6. A análise qualitativa dos modelos SegFormer, FCN e DeepLabV3+ em diferentes resoluções (256, 512 e 1024), reforça os padrões observados nos resultados quantitativos apresentados nas tabelas anteriores. A Figura 5 exemplifica como os modelos segmentam uma mesma imagem em diferentes resoluções de patch, evidenciando variações no nível de detalhamento. Já a Figura 6 ilustra os pontos fortes e fracos de cada dataset, destacando os desafios das arquiteturas para diferentes resoluções.

Na Figura 5, o comportamento dos modelos é evidente ao avaliar a mesma imagem em resoluções variadas. Na resolução 256, o FCN apresentou o melhor resultado, capturando detalhes menores com precisão, mas exibindo falsos positivos moderados devido à ampliação de áreas detectadas. O DeepLabV3+, apesar de razoável, mostrou fragmentações em regiões de bordas, comprometendo a conectividade. Já o SegFormer, dependente de maior contexto, teve dificuldades em predições precisas nesse cenário. Na resolução 512, os modelos se beneficiaram de um equilíbrio entre contexto e detalhamento. O SegFormer destacou-se, capturando padrões amplos e bordas mais precisas, enquanto o DeepLabV3+ e FCN mantiveram um bom desempenho, apesar de aumentarem os falsos positivos e negativos, respectivamente. Na resolução 1024, o contexto global ofereceu vantagens ao SegFormer, que melhorou a conectividade sem perder detalhes. Contudo, o DeepLabV3+ enfrentou queda significativa no desempenho devido às limitações de suas janelas de convolução fixas.

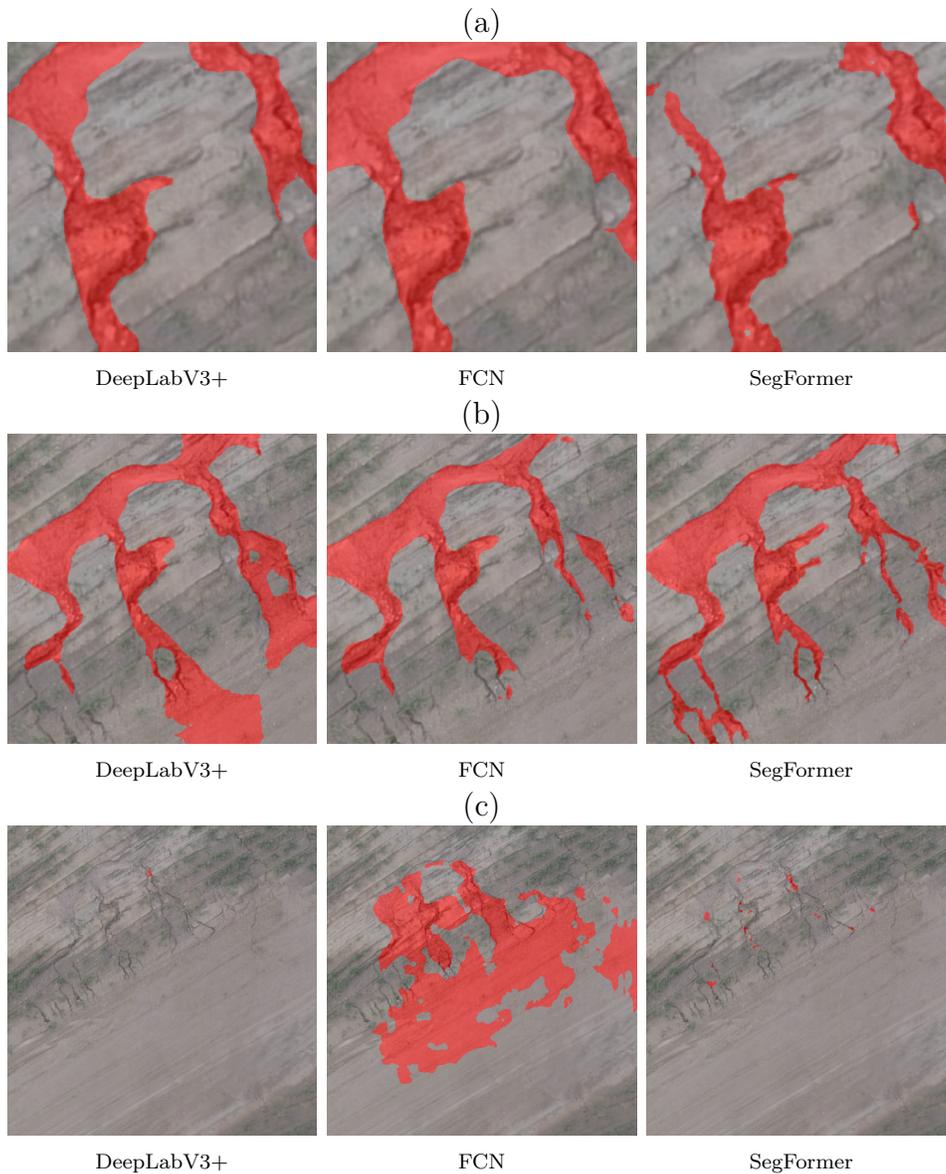


Figure 5: Resultados qualitativos dos métodos de segmentação em três resoluções: (a) 256, (b) 512, e (c) 1024.

A Figura 6 complementa a análise ao discutir os pontos fortes e fracos das resoluções. A resolução 256 (Figura 6a), destaca-se pelo detalhamento local, ideal para identificar padrões pequenos. Entretanto, a ausência de contexto global prejudicou a conectividade entre áreas de erosão, como observado na

Figura 6a à direita. A resolução 512 (Figura 6b) proporcionou um equilíbrio entre detalhamento e contexto. Na Figura 6b à esquerda, os modelos demonstraram boa precisão em bordas e conectividade, enquanto a Figura 6b à direita mostrou a capacidade de lidar com áreas complexas. A resolução 1024 (Figura 6c) evidencia como o contexto global favorece a identificação de padrões amplos e conectados. No entanto, a perda de detalhamento afetou a precisão de bordas e áreas menores, como na Figura 6c à direita, sendo um desafio para modelos como o DeepLabV3+.

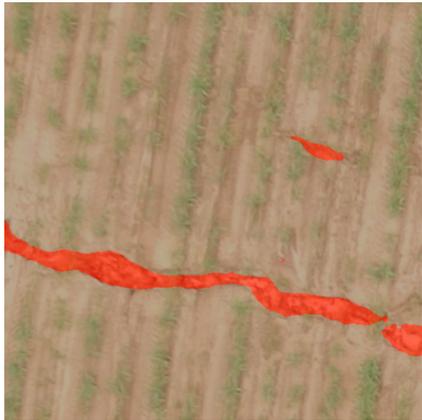
Os erros cometidos pelos modelos têm impactos distintos na prática, especialmente no contexto do monitoramento de áreas suscetíveis à erosão. Dois tipos de erro são particularmente relevantes: os falsos positivos (FP) que ocorrem quando o modelo identifica uma área como erosão, quando na realidade ela está em boas condições; e os falsos negativos (FN) que ocorrem quando o modelo falha em detectar uma área realmente afetada pela erosão.

Os FPs podem levar a intervenções desnecessárias em áreas intactas, aumentando custos e esforços para o manejo do solo. Por outro lado, os FNs, mais prejudiciais em termos de impacto ambiental, falham em alertar sobre áreas críticas de erosão que necessitam de intervenção urgente, permitindo a deterioração contínua. Modelos como o FCN apresentam maior incidência de FPs devido à ampliação das áreas detectadas, enquanto o DeepLabV3+, apesar de ser mais conservador, gera mais FNs devido à omissão de pequenos detalhes. O SegFormer, por sua vez, consegue equilibrar melhor esses erros, minimizando tanto FP quanto FN.

Essa análise qualitativa reforça que a escolha do modelo e do tamanho do dataset deve ser feita com base no objetivo da aplicação. Modelos baseados em convoluções, como o FCN e o DeepLabV3+, se mostram mais eficazes para análises focadas em detalhamento local, enquanto o SegFormer, com sua arquitetura baseada em transformadores, é mais indicado para aplicações que exigem contexto global, especialmente em áreas amplas e conectadas. Esses resultados oferecem subsídios importantes para a definição de estratégias em análises de solo, destacando a necessidade de considerar tanto as características do modelo quanto os desafios impostos pelo cenário analisado.



(a) 256



(b) 512



(c) 1024

Figure 6: Resultados qualitativos dos métodos de segmentação nos datasets, destacando pontos fortes e negativos.

4. Conclusão

Este trabalho propôs e avaliou um sistema de detecção de áreas de erosão em imagens aéreas utilizando técnicas de segmentação semântica com os modelos FCN, DeepLabV3+ e SegFormer. A metodologia foi estruturada nas seguintes etapas: geração dos patches em múltiplas resoluções (256×256 , 512×512 e 1024×1024), organização dos conjuntos (treino, validação e teste), treinamento dos modelos para cada escala e avaliação quantitativa e qualitativa dos resultados. As métricas utilizadas, IoU (Intersection over Union) e F1-Score, permitiram quantificar o desempenho dos modelos na segmentação da classe de erosão, evidenciando a capacidade de cada arquitetura em capturar tanto detalhes locais quanto contexto global nas áreas analisadas.

Os resultados demonstraram que a resolução dos patches influencia diretamente o desempenho dos modelos. Para patches de 256×256 , o FCN apresentou o melhor desempenho geral, com um F1-Score de 0.9786. No entanto, conforme a resolução aumentou, o SegFormer se destacou, atingindo um F1-Score de 0.8739 para 512×512 e 0.7624 para 1024×1024 , evidenciando uma maior capacidade de representar contextos globais. O modelo DeepLabV3+, embora competitivo em resoluções menores, apresentou uma queda expressiva em 1024×1024 , alcançando apenas 0.4293 de F1-Score, revelando dificuldades em lidar com regiões extensas e conectadas. A análise qualitativa confirmou esses achados, mostrando que o SegFormer, por ser baseado em Transformers, equilibrava melhor os erros de falsos positivos e falsos negativos em regiões amplas, enquanto o FCN capturava melhor os detalhes locais. O DeepLabV3+, por outro lado, apresentou dificuldades em manter a conectividade em bordas, especialmente em resoluções mais elevadas.

Para trabalhos futuros, recomenda-se a ampliação do dataset, incorporando novas regiões geográficas e diferentes condições climáticas, de modo a aumentar a robustez dos modelos. Adicionalmente, a implementação de estratégias de balanceamento de classes pode mitigar os desafios impostos pela baixa representatividade da erosão nos patches, contribuindo para um aprendizado mais equilibrado das redes. Outra possibilidade é explorar técnicas de pós-processamento para refinar as bordas das áreas segmentadas, principalmente em resoluções maiores, onde a perda de detalhamento se mostrou significativa. Essas melhorias podem ampliar a aplicabilidade dos modelos para monitoramento ambiental em larga escala, tornando o sistema mais robusto e preciso.

References

- [1] S. M. F. Maia, A. C. Corsi, J. E. Cunha, Sustentabilidade da produção de cana-de-açúcar no brasil, *Revista de Política Agrícola* 19 (3) (2010) 5–16.
- [2] L. A. Martinelli, S. Filoso, Expansion of sugarcane ethanol production in brazil: Environmental and social challenges, *Ecological Applications* 18 (4) (2008) 885–898. doi:<https://doi.org/10.1890/07-1813.1>.
- [3] S. A. Bezerra, J. R. B. Cantalice, Interrill erosion under different conditions of soil cover of sugarcane, *Revista Brasileira de Ciencia do Solo* 30 (3) (2006) 565–573, disponível em: <https://www.rbcjournal.org/article/interrill-erosion-under-different-conditions-of-soil-cover-of-sugarcane/>. doi:10.1590/S0100-06832006000300016.
- [4] R. P. C. Morgan, *Soil erosion and conservation*, 3rd Edition, Blackwell Publishing Ltd, Oxford, 2005.
- [5] T. G. Andualem, G. A. Hewa, B. R. Myers, S. Peters, J. Boland, Erosion and sediment transport modeling: A systematic review, *Land* 12 (7) (2023). doi:10.3390/land12071396.
- [6] A. S. Pambudi, Determination of the most priority conservation areas based on population pressure and erosion hazard levels in lesti sub-watershed, malang regency, indonesia, in: A. Vieira, S. C. Rodrigues (Eds.), *Soil Erosion*, IntechOpen, 2020, Ch. 4, disponível em: <https://doi.org/10.5772/intechopen.95386>. doi:10.5772/intechopen.95386.
- [7] L. Qi, P. Shi, K. Dvorakova, K. Van Oost, Q. Sun, H. Yu, B. van Wesemael, Detection of soil erosion hotspots in the croplands of a typical black soil region in northeast china: Insights from sentinel-2 multispectral remote sensing, *Remote Sensing* 15 (5) (2023). doi:10.3390/rs15051402.
- [8] V. Carabassa, P. Montero, J. M. Alcañiz, J.-C. Padró, Soil erosion monitoring in quarry restoration using drones, *Minerals* 11 (9) (2021). doi:10.3390/min11090949.

- [9] P. Domlija, S. Bernat Gazibara, Z. Arbanas, S. Mihalic Arbanas, Identification and mapping of soil erosion processes using the visual interpretation of lidar imagery, *ISPRS International Journal of Geo-Information* 8 (10) (2019). doi:10.3390/ijgi8100438.
- [10] V. Ferro, A. Nicosia, Soil erosion measurement techniques and field experiments, *Water* 15 (15) (2023). doi:10.3390/w15152846.
- [11] P. Tu, Q. Zhou, M. Qi, Soil erosion type and risk identification from the perspective of directed weighted complex network, *Sustainability* 15 (3) (2023). doi:10.3390/su15031939.
URL <https://www.mdpi.com/2071-1050/15/3/1939>
- [12] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015. doi:10.1109/CVPR.2015.7298965.
- [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818, disponível em: <https://arxiv.org/abs/1802.02611>.
- [14] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers, in: Advances in Neural Information Processing Systems (NeurIPS), 2021.