Universidade de Brasília
Instituto de Ciências Humanas
Departamento de Geografia
Programa de Pós-graduação em Geografia

Aplicações de Modelos de Deep Learning Para Monitoramento Ambiental e Agrícola no Brasil

Tese de Doutorado

Pablo Pozzobon de Bem

Brasília, fevereiro de 2022

Universidade de Brasília
Instituto de Ciências Humanas
Departamento de Geografia
Programa de Pós-graduação em Geografia

Aplicações de Modelos de Deep Learning Para Monitoramento Ambiental e Agrícola no Brasil

Autor: Pablo Pozzobon de Bem

Tese de Doutorado submetida ao Departamento de Geografia da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do Grau de Doutor em Geografia.

Área de concentração: Geoprocessamento

Orientador: Dr. Osmar Abílio de Carvalho Júnior

Brasília, fevereiro de 2022

Universidade de Brasília

Instituto de Ciências Humanas
Departamento de Geografia
Programa de Pós-graduação em Geografia

Aplicações de Modelos de Deep Learning Para Monitoramento Ambiental e Agrícola no Brasil

Pablo Pozzobon de Bem

Tese de Doutorado submetida ao Departamento de Geografia da Universidade de Brasília, como parte dos requisitos necessários para a obtenção do Grau de Doutor em geografia, área de concentração Geoprocessamento, opção Acadêmica.

Aprovado por:

Dr. Osmar Abílio de Carvalho Junior
(Departamento de Geografia – Universidade de Brasília)
(Orientador)

Dr. Eraldo Aparecido Trondoli Matricardi
(Departamento de Engenharia Florestal – Universidade de Brasília)
(Examinador Interno)

Dra. Carla Bernadete Madureira Cruz
(Instituto de Geociências – Universidade Federal do Rio de Janeiro)
(Examinadora Externa)

Dra. Liana Oighenstein Anderson
(Centro Nacional de Monitoramento e Alerta de Desastres Naturais – CEMADEN)
(Examinadora Externa)

Brasília, 28 de fevereiro de 2022

Ficha catalográfica

PB455a

POZZOBON DE BEM, PABLO

Aplicações de Modelos de Deep Learning Para Monitoramento Ambiental e Agrícola no Brasil / Pablo Pozzobon de Bem; orientador Osmar Abílio de Carvalho Júnior – Brasília, 2022.

95 p.

Tese de Doutorado – Programa de Pós-graduação em Geografia – Universidade de Brasília, 2022

1. Uso e cobertura da terra
2. Aprendizado de máquina
3. Sensoriamento remoto
4. Detecção de mudança

_____

Pablo Pozzobon de Bem

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

ADAM	Adaptative Moment
AUPRC	Area Under Precision Recall Curve
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BCE	Binary Cross Entropy
CONAB	Companhia Nacional de Abastecimento
CPU	Central Processing Unit
CNN	Convolutional Neural Network
DARNN	Domain-adapted Recurrent Neural Network
DETER	Sistema de Detecção de Desmatamento em Tempo Real
DL	Deep Learning
DT	Decision Tree
ESA	European Space Agency
FCRNN	Fully Connected Recurrent Neural Network
GEE	Google Earth Engine
GLCM	Grey Level Co-occurrence Matrix
GPU	Graphical Processing Unit
GRD	Ground Range Detected
GT	Ground Truth
INPE	Instituto Nacional de Pesquisas Espaciais
IRGA	Instituto Rio-Grandense do Arroz
IW	Interferometric Wide
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
mIoU	Mean Intersection Over Union
ML	Machine Learning
MLP	Multilayer Perceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
NADAM	Nesterov Adaptative Moment
NASA	North American Space Agency
NBR	Normalized Burn Ratio
OBIA	Object Based Image Analysis
OLI	Operational Land Imager
PRODES	Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite
RF	Random Forest
RGB	Red-Green-Blue
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RS	Remote Sensing
SAR	Synthetic-aperture Radar
SVM	Support Vector Machine
UAV	Unnmanned Aerial Vehicle
USGS	United States Geological Survey
VH	Vertical–Horizontal
VV	Vertical–Vertical

Resumo

Algoritmos do novo campo de aprendizado de máquina conhecido como *Deep Learning* têm se popularizado recentemente, mostrando resultados superiores a modelos tradicionais em métodos de classificação e regressão. O histórico de sua utilização no campo do sensoriamento remoto ainda é breve, porém eles têm mostrado resultados similarmente superiores em processos como a classificação de uso e cobertura da terra e detecção de mudança. Esta tese teve como objetivo o desenvolvimento de metodologias utilizando estes algoritmos com um enfoque no monitoramento de alvos críticos no Brasil por via de imagens de satélite a fim de buscar modelos de alta precisão e acurácia para substituir metodologias utilizadas atualmente. Ao longo de seu desenvolvimento, foram produzidos três artigos onde foi avaliado o uso destes algoritmos para a detecção de três alvos distintos: (a) áreas queimadas no Cerrado brasileiro, (b) áreas desmatadas na região da Amazônia e (c) plantios de arroz no sul do Brasil. Apesar do objetivo similar na produção dos artigos, procurou-se distinguir suficientemente suas metodologias a fim de expandir o espaço metodológico conhecido para fornecer uma base teórica para facilitar e incentivar a adoção destes algoritmos em contexto nacional. O primeiro artigo avaliou diferentes dimensões de amostras para a classificação de áreas queimadas em imagens Landsat-8. O segundo artigo avaliou a utilização de séries temporais binárias de imagens Landsat para a detecção de novas áreas desmatadas entre os anos de 2017, 2018 e 2019. O último artigo utilizou imagens de radar Sentinel-1 (SAR) em uma série temporal contínua para a delimitação dos plantios de arroz no Rio Grande do Sul. Modelos similares foram utilizados em todos os artigos, porém certos modelos foram exclusivos a cada publicação, produzindo diferentes resultados. De maneira geral, os resultados encontrados mostram que algoritmos de *Deep Learning* são não só viáveis para detecção destes alvos mas também oferecem desempenho superior a métodos existentes na literatura, representando uma alternativa altamente eficiente para classificação e detecção de mudança dos alvos avaliados.

Abstract

Algorithms belonging to the new field of machine learning called Deep Learning have been gaining popularity recently, showing superior results when compared to traditional classification and regression methods. The history of their use in the field of remote sensing is not long, however they have been showing similarly superior results in processes such as land use classification and change detection. This thesis had as its objective the development of methodologies using these algorithms with a focus on monitoring critical targets in Brazil through satellite imagery in order to find high accuracy and precision models to substitute methods used currently. Through the development of this thesis, articles were produced evaluating their use for the detection of three distinct targets: (a) burnt areas in the Brazilian Cerrado, (b) deforested areas in the Amazon region and (c) rice fields in the south of Brazil. Despite the similar objective in the production of these articles, the methodologies in each of them was made sufficiently distinct in order to expand the methodological space known. The first article evaluated the use of differently sized samples to classify burnt areas in Landsat-8 imagery. The second article evaluated the use of binary Landsat time series to detect new deforested areas between the years of 2017, 2018 and 2019. The last article used continuous radar Sentinel-1 (SAR) time series to map rice fields in the state of Rio Grande do Sul. Similar models were used in all articles, however certain models were exclusive to each one. In general, the results show that not only are the Deep Learning models viable but also offer better results in comparison to other existing methods, representing an efficient alternative when it comes to the classification and change detection of the targets evaluated.

Sumário

Capítulo I

Introdução

Deep Learning (DL) é um ramo do campo mais abrangente conhecido como *Machine Learning* (ML), ou aprendizado de máquina, que tem ganhado forte tração nos últimos anos devido ao potencial dos algoritmos utilizados e a avanços no campo da computação que facilitaram sua implementação e acessibilidade. Melhorias em hardware, particularmente o advento de unidades de processamento gráficas (GPUs) com maiores capacidade de processamento reduziram o tempo necessário de treinamento e aplicação destes algoritmos, e a criação de frameworks dedicados para DL tais como *Tensorflow* (Abadi et al. 2015), *PyTorch* (Paszke et al. 2019) e *Caffe* (Jia et al. 2014), junto da disponibilização de bibliotecas de alto nível de abstração como Keras (Chollet et al. 2015) simplificou a construção de modelos, tornando o campo mais acessível para o público leigo.

Estes algoritmos se diferenciam de algoritmos clássicos de ML tais como regressões, árvores de decisão e redes neurais rasas em função de sua maior profundidade, complexidade e capacidade de aprender em um contexto multidimensional. Suas estruturas de caráter modular podem conter uma variedade de camadas disponíveis para identificar diferentes tipos de padrões em um conjunto de dados. Um fator comum dentre as arquiteturas de DL é o uso de camadas com filtros convolucionais para a extração de padrões e feições que geralmente são utilizadas para formar arquiteturas conhecidas como *Convolutional Neural Networks* (CNN), ou redes neurais convolucionais. Devido à capacidade de identificar padrões em âmbito multidimensional, este tipo de arquitetura se popularizou e vêm sendo utilizado para a solução de diversos problemas tais como análise e classificação textual (Zhang, Zhao et al. 2015), reconhecimento e classificação de imagens (Simonyan e Zisserman 2015; Chan et al. 2015), reconhecimento de voz (Zhang, Chan et al. 2017) e análise de vídeos (Frizzi et al. 2016). Modelos de DL como CNNs aprendem através de processos iterativos de treinamento nos quais eles realizam previsões que são comparadas à realidade (*ground truth*), produzindo valores de erros que são utilizados para ajustar e otimizar os valores dos filtros e pesos utilizados nas camadas convolucionais e entre camadas — processo denominado *backpropagation* — ao longo de diversas gerações.

Pesquisadores também têm encontrado aplicações da tecnologia no campo do sensoriamento remoto, onde ela tem suplantando as metodologias previamente utilizadas para a solução de certos problemas como classificação de uso e cobertura da terra (Carranza-García et al. 2019; Scott et al. 2017), detecção de objetos (Guo et al. 2018) e detecção de mudanças (Shi et al. 2020).

A classificação de uso e cobertura da terra, onde são associadas classes semânticas a frações da superfície terrestre, é atualmente a aplicação mais comum de algoritmos de DL (Ball et al. 2017), possivelmente em função da grande variedade de alvos a serem identificados e da necessidade de se obter classificações de alta fidelidade em relação à realidade. Estudos têm utilizado estes algoritmos para a classificação de diversos tipos de alvos tais como corpos d'água (Yu et al. 2017), nuvens (Liu, Zhang et al. 2019), deslizamentos de terra (Ghorbanzadeh et al. 2019) e paineis e usinas de produção de energia solar (Imamoglu et al. 2017).

A detecção de mudanças da superfície terrestre em séries temporais de imagens de satélite também se encontra entre os usos mais comuns destes algoritmos dadas as suas capacidades de detecção de padrões e, consequentemente, a capacidade de detecção de mudanças nesses padrões. Existe uma grande variedade metodológica em estudos de detecção de mudanças, em função do tipo de imagem, da quantidade de imagens e do objeto de estudo (Tewkesbury et al. 2015), porém o principal fator comum é o que há uma comparação entre estados temporais da superfície. A dimensão temporal nesse tipo de problema representa uma fonte adicional de informação para algoritmos de DL, permitindo a identificação de padrões temporais em adição aos padrões espaciais e espectrais. A detecção de mudanças em estruturas urbanas (Liu, Kuffer et al. 2019), mudanças de uso da terra (Cao et al. 2019) e detecção de desastres naturais (Liu e Wu 2016) são alguns exemplos do uso destes algoritmos neste campo.

Atualmente uma grande variedade de informações pode ser extraída de dados de diferentes sensores e pesquisadores têm procurado utilizar aquelas que melhor se adaptem a seus objetos de estudo. Sensores hiperespectrais como o *Airborne Visible/Infrared Imaging Spectrometer* (AVIRIS) da Agência Espacial Norte-Americana (NASA) oferecem imagens ricas em informação espectral que têm sido amplamente utilizadas em estudos com DL (Zhang, Chan et al. 2017; Santara et al. 2017; Zhong et al. 2018). O sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) capta imagens com média resolução espectral e baixa resolução espacial, porém com resolução temporal diária, para estudos com objetos onde se necessita de um acompanhamento detalhado da variação temporal dos dados (Wang et al. 2018). Os sensores a bordo dos satélites *Landsat* 5 a 8 da NASA têm média reincidência e produzem uma a duas imagens por mês sobre o mesmo ponto na terra, porém com média resolução espacial e espectral, sendo utilizadas quando o objeto em foco necessita de detalhe espacial e a informação espectral possui certa relevância (Yu et al. 2017; Zeng et al. 2018). Imagens detectadas via radar, tais como Lidar e o *Synthetic Aperture Radar* (SAR) a bordo do satélite Sentinel-1 da Agência Espacial Europeia (ESA) são particularmente úteis em casos onde informação textural é relevante, além de oferecerem imagens livres de nuvens devido às características inerentes a esse tipo de sensoriamento (Cozzolino et al. 2017; Kang et al. 2017).

Os principais trabalhos de monitoramento por satélite no país têm sido realizados pelo Instituto Nacional de Pesquisas Espaciais (INPE), através de programas como o Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES), o Sistema de Detecção de Desmatamento em Tempo Real (DETER), o Sistema de Mapeamento do uso e Ocupação da Terra, (TerraClass) e o Programa de Monitoramento de Queimadas. Recentemente o projeto

Mapbiomas também tem contribuído com mapeamentos de uso e cobertura da terra no Brasil, particularmente sobre os impactos do desmatamento na florestal Amazônica. Apesar da confiabilidade e uso frequente dos dados produzidos por agências nacionais, existe ainda espaço para o aprimoramento dos métodos de monitoramento ambiental, e estudos mostram que produtos como o PRODES e mapas de distribuição de áreas queimadas usados amplamente em um contexto global possuem uma tendência a subestimar a extensão real do problema, seja por razões metodológicas ou operacionais (Nogueira et al. 2016; Milodowski et al. 2017). Estas novas tecnologias como DL surgem como possíveis candidatos para a produção de mapas mais compatíveis com a realidade.

O monitoramento de plantios agrícolas também é um aspecto importante do monitoramento do Brasil via sensoriamento remoto, seja para fins produtivos ou para fins de conservação ambiental. No entanto, levantamentos de plantios agrícolas no país são realizados em grande parte através de mapeamentos manuais por via de interpretação visual e visitas *in situ*, resultando em processos longos e dispendiosos. O arroz produzido no estado do Rio Grande do Sul é exemplo disto. Sendo um dos componentes principais da sexta básica brasileira, é um dos produtos agrícolas mais importantes no país e portanto justifica-se o monitoramento e acompanhamento deste tipo de plantio. Este processo é realizado atualmente por órgãos como a Companhia Nacional de Abastecimento (CONAB), que realiza mapeamentos de diversos tipos de plantios no país. O uso de algoritmos de DL pode surgir como uma alternativa automatizada ou semi-automatizada com elevado grau de acurácia para substituir os processos utilizados atualmente no país. Estudos têm mostrado que estes modelos são capazes de fornecer mapas confiáveis e de baixo custo de produção em plantios de grãos como soja (Xu, Zhu et al. 2020), milho (Xu, Yang et al. 2021) e arroz (Park et al. 2018).

Dado o contexto até então apresentado, levanta-se a hipótese de que estes algoritmos oferecem uma alternativa de alta acurácia e precisão para a classificação e demarcação de áreas críticas nos contextos ambiental e agrícola do país tais como áreas desmatadas, áreas queimadas e plantios de arroz — possivelmente produzindo resultados superiores aos métodos utilizados atualmente. A fim de testar esta hipótese, o objetivo principal deste trabalho foi a investigação e aplicação de diferentes algoritmos de DL para a detecção destes alvos e, ademais, compará-los com métodos já existentes.

Sob uma ótica global, o número de estudos utilizando algoritmos de DL junto de dados de sensoriamento remoto tem crescido cada vez mais. No entanto, estudos realizados no Brasil — particularmente com um enfoque ambiental e agrícola — ainda são comparativamente escassos e uma parte considerável dos estudos em áreas nacionais têm partido de iniciativa internacional apesar da importância deste tópico para o desenvolvimento ambiental e agrícola do país. Portanto, além de apresentar uma base científica para futuros estudos, esta tese surge ainda como uma fonte de auxílio para incentivar estes passos iniciais na adoção desta tecnologia no país.

Além deste capítulo introdutório, neste trabalho constam três capítulos contendo publicações produzidas em paralelo ao longo do desenvolvimento desta tese a fim de testar a hipótese levantada e explorar a utilização destes algoritmos no campo do sensoriamento remoto (Figura

I.1). Estes artigos tiveram como foco três alvos importantes nos contextos ambiental e agrícola no Brasil. No Capítulo II, a seguir, é apresentado o artigo intitulado "*Performance Analysis of Deep Convolutional Autoencoders with Different Patch Sizes for Change Detection from Burnt Areas*", no qual foram utilizados algoritmos para a detecção de áreas queimadas no Cerrado brasileiro. Em seguida, no Capítulo 3, é apresentado o artigo "*Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks*", onde foi explorada sua utilização para a detecção de áreas desmatadas na Amazônia. O último artigo, de título "*Irrigated Rice Crop Identification in Southern Brazil Using Convolutional Neural Networks and Sentinel-1 Time Series*", difere dos demais no tipo de alvo a ser detectado, dando um enfoque aos plantios de arroz no estado do Rio Grande do Sul. Existem similaridades entre as metodologias dos três artigos apresentados, porém procurou-se diferencia-las a fim de expandir o espaço metodológico conhecido. Finalmente, encerra-se esta tese com um capítulo de discussão e um capítulo final onde são discutidas as conclusões encontradas e possíveis procedências no campo do uso de DL no sensoriamento remoto.

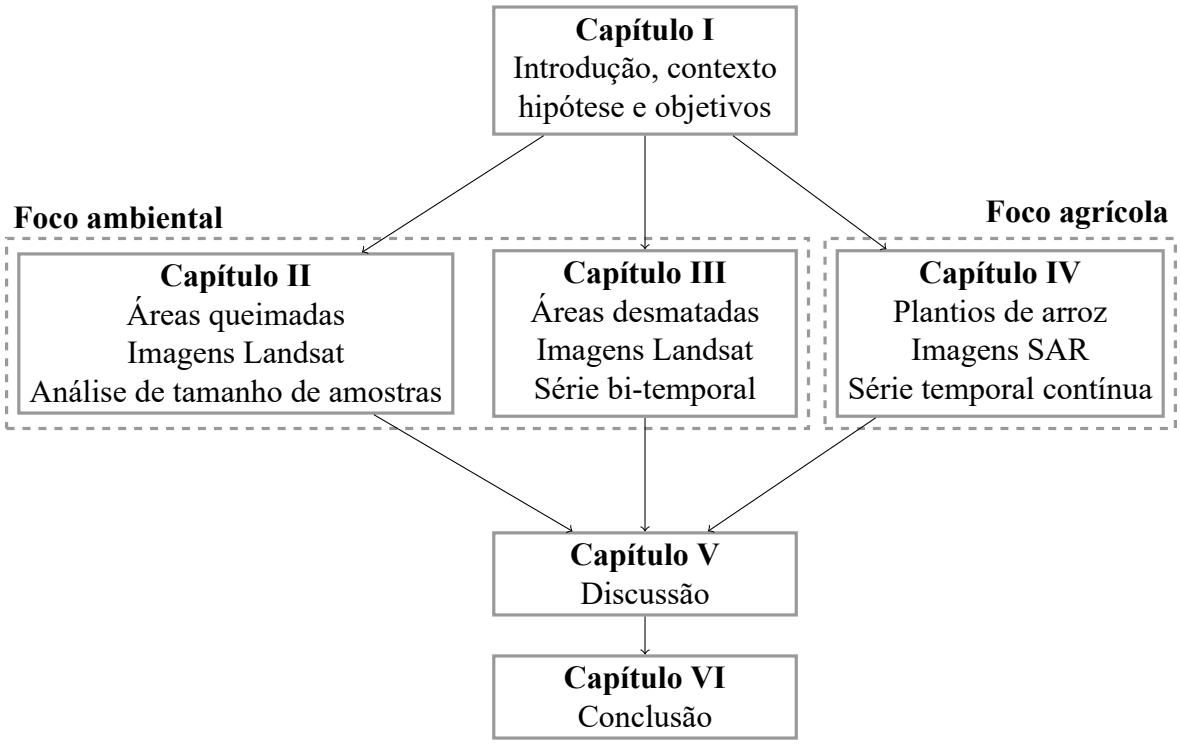

Figura I.1. Estruturação do desenvolvimento desta tese.

Referências

Abadi, Martín et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*.

Ball, John E., Derek T. Anderson e Chee Seng Chan (set. de 2017). "Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools, and Challenges for the Community". Em: *Journal of Applied Remote Sensing* 11.04, p. 1. ISSN: 1931-3195. DOI: 10.1117/1.JRS.11.042609.

Cao, Cong, Suzana Dragićević e Songnian Li (fev. de 2019). “Land-Use Change Detection with Convolutional Neural Network Methods”. Em: *Environments* 6.2, p. 25. ISSN: 2076-3298. DOI: 10.3390/environments6020025.

Carranza-García, Manuel, Jorge García-Gutiérrez e José Riquelme (jan. de 2019). “A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks”. Em: *Remote Sensing* 11.3, p. 274. ISSN: 2072-4292. DOI: 10.3390/RS11030274.

Chan, Tsung-Han et al. (dez. de 2015). “PCANet: A Simple Deep Learning Baseline for Image Classification?” Em: *IEEE Transactions on Image Processing* 24.12, pp. 5017–5032. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2015.2475625.

Chollet, François et al. (2015). *Keras*. https://keras.io.

Cozzolino, Davide et al. (jul. de 2017). “A Fully Convolutional Neural Network for Low-Complexity Single-Stage Ship Detection in Sentinel-1 SAR Images”. Em: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX: IEEE, pp. 886–889. ISBN: 978-1-5090-4951-6. DOI: 10.1109/IGARSS.2017.8127094.

Frizzi, Sebastien et al. (out. de 2016). “Convolutional Neural Network for Video Fire and Smoke Detection”. Em: *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*. Florence, Italy: IEEE, pp. 877–882. ISBN: 978-1-5090-3474-1. DOI: 10.1109/IECON.2016.7793196.

Ghorbanzadeh, Omid et al. (jan. de 2019). “Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection”. Em: *Remote Sensing* 11.2, p. 196. ISSN: 2072-4292. DOI: 10.3390/rs11020196.

Guo, Wei et al. (jan. de 2018). “Geospatial Object Detection in High Resolution Satellite Images Based on Multi-Scale Convolutional Neural Network”. Em: *Remote Sensing* 10.1, p. 131. ISSN: 2072-4292. DOI: 10.3390/rs10010131.

Imamoglu, Nevrez et al. (abr. de 2017). “Solar Power Plant Detection on Multi-Spectral Satellite Imagery Using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion”. Em: *arXiv:1704.06410 [cs]*. arXiv: 1704.06410 [cs].

Jia, Yangqing et al. (2014). *Caffe: Convolutional Architecture for Fast Feature Embedding*. arXiv: 1408.5093 [cs.CV].

Kang, Miao et al. (ago. de 2017). “Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection”. Em: *Remote Sensing* 9.8, p. 860. ISSN: 2072-4292. DOI: 10.3390/rs9080860.

Liu, Cheng-Chien, Yu-Cheng Zhang et al. (jan. de 2019). “Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation”. Em: *Remote Sensing* 11.2, p. 119. ISSN: 2072-4292. DOI: 10/gftgqn.

Liu, Ruoyun, Monika Kuffer e Claudio Persello (nov. de 2019). “The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach”. Em: *Remote Sensing* 11.23, p. 2844. ISSN: 2072-4292. DOI: 10.3390/rs11232844.

Liu, Ying e Linzhi Wu (2016). “Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning”. Em: *Procedia Computer Science* 91, pp. 566–575. ISSN: 18770509. DOI: 10.1016/j.procs.2016.07.144.

Milodowski, D T, E T A Mitchard e M Williams (set. de 2017). “Forest Loss Maps from Regional Satellite Monitoring Systematically Underestimate Deforestation in Two Rapidly Changing Parts of the Amazon”. Em: *Environmental Research Letters* 12.9, p. 094003. ISSN: 1748-9326. DOI: 10.1088/1748-9326/aa7e1e.

Nogueira, Joana et al. (dez. de 2016). “Can We Go Beyond Burned Area in the Assessment of Global Remote Sensing Products with Fire Patch Metrics?” Em: *Remote Sensing* 9.1, p. 7. ISSN: 2072-4292. DOI: 10.3390/rs9010007.

Park, Seonyoung et al. (2018). “Classification and Mapping of Paddy Rice by Combining Landsat and SAR Time Series Data”. Em: *Remote Sensing* 10.3, p. 447. DOI: 10.3390/rs10030447.

Paszke, Adam et al. (2019). “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. Em: *Advances in Neural Information Processing Systems 32*. Ed. por H. Wallach et al. Curran Associates, Inc., pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Santara, Anirban et al. (set. de 2017). “BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification”. Em: *IEEE Transactions on Geoscience and Remote Sensing* 55.9, pp. 5293–5301. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2017.2705073.

Scott, Grant J. et al. (abr. de 2017). “Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery”. Em: *IEEE Geoscience and Remote Sensing Letters* 14.4, pp. 549–553. ISSN: 1545-598X. DOI: 10.1109/LGRS.2017.2657778.

Shi, Wenzhong et al. (mai. de 2020). “Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges”. Em: *Remote Sensing* 12.10, p. 1688. ISSN: 2072-4292. DOI: 10.3390/rs12101688.

Simonyan, Karen e Andrew Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs.CV].

Tewkesbury, Andrew P. et al. (abr. de 2015). “A Critical Synthesis of Remotely Sensed Optical Image Change Detection Techniques”. Em: *Remote Sensing of Environment* 160, pp. 1–14. ISSN: 00344257. DOI: 10.1016/j.rse.2015.01.006.

Wang, Anna X. et al. (2018). “Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data”. Em: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) - COMPASS ’18*. Menlo Park and San Jose, CA, USA: ACM Press, pp. 1–5. ISBN: 978-1-4503-5816-3. DOI: 10.1145/3209811.3212707.

Xu, Jinfan, Jie Yang et al. (out. de 2021). “Towards Interpreting Multi-Temporal Deep Learning Models in Crop Mapping”. Em: *Remote Sensing of Environment* 264, p. 112599. ISSN: 00344257. DOI: 10.1016/j.rse.2021.112599.

Xu, Jinfan, Yue Zhu et al. (set. de 2020). “DeepCropMapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean

Mapping". Em: *Remote Sensing of Environment* 247, p. 111946. ISSN: 00344257. DOI: 10.1016/j.rse.2020.111946.

Yu, Long et al. (mar. de 2017). "Convolutional Neural Networks for Water Body Extraction from Landsat Imagery". Em: *International Journal of Computational Intelligence and Applications* 16.01, p. 1750001. ISSN: 1469-0268, 1757-5885. DOI: 10.1142/S1469026817500018.

Zeng, Xiaoshuang et al. (2018). "Cloud detection of remote sensing images on Landsat-8 by deep learning". Em: *Tenth International Conference on Digital Image Processing (ICDIP 2018)*. Ed. por Xudong Jiang e Jenq-Neng Hwang. Vol. 10806. International Society for Optics e Photonics. SPIE, pp. 1388–1392. DOI: 10.1117/12.2503034. URL: https://doi.org/10.1117/12.2503034.

Zhang, Xiang, Junbo Zhao e Yann LeCun (2015). "Character-Level Convolutional Networks for Text Classification". Em: *Advances in Neural Information Processing Systems 28*. Ed. por C Cortes et al. Curran Associates, Inc., pp. 649–657.

Zhang, Yu, William Chan e Navdeep Jaitly (mar. de 2017). "Very Deep Convolutional Networks for End-to-End Speech Recognition". Em: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA: IEEE, pp. 4845–4849. ISBN: 978-1-5090-4117-6. DOI: 10.1109/ICASSP.2017.7953077.

Zhong, Zilong et al. (fev. de 2018). "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework". Em: *IEEE Transactions on Geoscience and Remote Sensing* 56.2, pp. 847–858. ISSN: 0196-2892. DOI: 10.1109/TGRS.2017.2755542.

Capítulo II

Performance Analysis of Deep Convolutional Autoencoders with Different Patch Sizes for Change Detection from Burnt Areas

Resumo: Fogo é uma das principais causas de danos a ambientes naturais em um nível global. Estimativas mostram que aproximadamente 4 milhões de km² de florestas queimam anualmente. Estudos mostram que no entanto essas estimativas tendem a subestimar a extensão real de áreas queimadas, evidenciando a necessidade de se encontrar métodos mais confiáveis para detectar e classificar estas áreas. Esse estudo teve como objetivo analisar o uso de arquiteturas conhecidas como Deep Convolutional Autoencoders na classificação de áreas queimadas considerando diferentes dimensões de amostras. Um Autoencoder simples juntamente das arquiteturas U-Net e ResUnet foram avaliados. Imagens Landsat 8 OLI+ em três cenas diferentes em quatro datas consecutivas foram utilizadas no processo de detecção de áreas queimadas. Os dados foram amostrados segundo quatro estrategias de amostragem diferentes para avaliar diferenças na performance dos modelos. A fase de treino utilizou duas das cenas disponíveis, enquanto a fase de validação utilizou a cena restante. A máscara de mudança demonstrando as áreas queimadas foi criada utilizando o índice espectral Normalized Burn Ratio (NBR) como base. As classificações dos modelos foram avaliadas segundo os índices F1, Kappa e mean Intersection over Union (mIoU). Os resultados mostram que as arquiteturas U-Net e ResUnet ofereceram as melhores classificações com valores médios de F1, Kappa e mIoU de aproximadamente 0.96, representando classificações excelentes. Também se verificou que uma amostragem utilizando um tamanho de janela de 256 por 256 pixels produziu os melhores resultados.

Artigo publicado na revista *Remote Sensing*, 2020. DOI: 10.3390/rs12162576

Performance Analysis of Deep Convolutional Autoencoders with Different Patch Sizes for Change Detection from Burnt Areas

Abstract: Fire is one of the primary sources of damages to natural environments globally. Estimates show that approximately 4 million km² of land burns yearly. Studies have shown that such estimates often underestimate the real extent of burnt land, which highlights the need to find better, state-of-the-art methods to detect and classify these areas. This study aimed to analyze the use of deep convolutional Autoencoders in the classification of burnt areas, considering different sample patch sizes. A simple Autoencoder and the U-Net and ResUnet architectures were evaluated. We collected Landsat 8 OLI+ data from three scenes in four consecutive dates to detect the changes specifically in the form of burnt land. This data was sampled according to four different sampling strategies to evaluate possible performance changes related to sampling window sizes. The training stage used two scenes, while the validation stage used the remaining scene. The ground truth change mask was created using the Normalized Burn Ratio (NBR) spectral index through a thresholding approach. The classifications were evaluated according to the F1 index, Kappa index, and mean Intersection over Union (mIoU) value. Results have shown that the U-Net and ResUnet architectures offered the best classifications with average F1, Kappa, and mIoU values of approximately 0.96, representing excellent classification results. We have also verified that a sampling window size of 256 by 256 pixels offered the best results.

Keywords: deep learning; CNN; classification; fire; multitemporal image

1. Introduction

Deep Learning (DL) is the term that refers to the use of multilayered neural networks to solve complex problems. It is one of the fastest-growing trends in machine learning, data science, and computer vision. The DL field has become more accessible in the latest years due to improvements in the understanding of machine-learning theory, coupled with the increased processing power from the better consumer-grade computer hardware. Within the full range of deep network architectures, Convolutional Neural Networks (CNN) gained the focus in recent DL studies. CNNs are multilayer neural networks capable of identifying patterns in data, both spatial and spectrally wise, and using these learned patterns for inference to classify data. Studies have shown they possess the ability to solve a wide variety of problems, such as text classification [1], image recognition [2], video analysis [3], and speech recognition [4].

Recent DL algorithms have found several applications in geoscience and Remote Sensing (RS) domains [5,6]. Studies have shown the feasibility and performance of CNNs

for common RS tasks such as land cover classification [7–9], object detection [10,11], image pansharpening [12,13], change detection [14–17], and many others. These tasks have been carried successfully with a variety of RS image types, such as hyperspectral [7], multispectral [11,18], and in simple Red-Green-Blue (RGB) images [19].

Within the scope of land cover classification and change detection, the mapping of areas burnt by wildfires is crucial because of their ecological, social, and economic impacts [20]. The Brazilian Savanah (Cerrado) is one of the world's hotspots for biodiversity conservation and the world's richest neotropical savanna [21]. It is characterized by the regular occurrence of fires, either through natural or anthropogenic means, and approximately 170.000 km² of land within the region have burnt every year on average for the past ten years [22], making the Cerrado one of the regions most affected by wildfires globally [23].

Since the advent of satellite imagery, researchers have extensively attempted to map burnt areas as a critical step in understanding and preventing the social and environmental damage caused by fire [24]. Estimates show that fire consumes up to approximately 4 million km² of land yearly [25]. In the Cerrado biome, several studies analyze fire events based on remote sensing data, defining spatial patterns [26–28], temporal frequency [29,30], drivers of fire occurrences [31,32], and climatic effects [33].

Studies have found that commonly used global burnt area products such as the MCD64A1 MODIS dataset [34] often underestimate the real extent of the burnt areas [35–37]. Studies in the Brazilian territory [29,38] detected high values of errors in the MODIS-MCD45 product (commission error of 36.69% and omission error of 77.04%) and MODIS-MCD64 product (commission error of 45, 85% and omission error of 64.05%). Machine Learning (ML) algorithms have been shown to offer better results than such products [39]. Therefore, the current accuracy results of the global fire mapping products show the need for advances in the detection and classification of the area of fires.

Shallow ML algorithms such as Support Vector Machines (SVM) and Random Forest (RF) along with shallow fully connected neural networks such as the Multilayer Perceptron (MLP) have been used to classify and detect burnt areas [39–44]. However, recently these shallow algorithms were surpassed in most tasks by the deeper, more complex CNNs, which potential in classifying burnt area is still relatively unexplored. Studies that have applied DL algorithms to map burnt land have shown promising results [45,46], although several factors need further investigation, such as the use of different types of architectures and hyperparameter tuning. Several types of CNNs have been proposed and used for change detection and land cover classification [9,11,15,19,47]. The architectures known as

Autoencoders have shown consistently good results among the many types of CNNs used for image segmentation [48–51]. Autoencoders use the concept of downsampling and upsampling feature maps, which makes them very efficient memory-wise and helps detect both high-level semantic information and low-level spatial detail. Therefore, Autoencoders offer a good choice of architecture to classify RS data given how memory intensive it can be and how it is highly dependent on spatial information.

Within the scope of segmentation in Remote Sensing images, the sampling technique has been a topic of discussion. The sampling process is generally performed by either placing random sampling windows or by sliding a sampling window along the image to collect pixel data in the form of smaller patches [52]. The sliding window technique is much more common, although there is no consensus on the optimal window size, which seems to depend on the type of image used and the target analyzed. Varying window sizes have been investigated in literature for different objects, such as (a) 17 × 17 pixels to detect oil palm trees in a plantation area from QuickBird image [53]; (b) 50 × 50 pixels to detect vehicles in aerial images [54], (c) 224 × 224 pixels for the analysis of damaged buildings using aerial images of 0.5-m resolution [55]; (d) 256 × 256 pixels to classify urban buildings from an image with 0.075-m resolution [56]; and (e) 400 × 400 pixels resampled to 256 × 256 pixels to classify urban land cover using high-resolution aerial images [57]. Although studies consider that the adequate window size should cover the intended target, a window size sensitivity analysis allows for the detection of the optimal dimension. A study mapping land cover using RapidEye images [58] tested different window sizes (5, 10, 15, 20, 25, 30, 35, and 40) and determined the dimension of 30 x 30 pixels as the ideal size. Another study evaluated different window sizes (60, 80, 100, 120, 140, 160, 180, and 200) to locate cars in unmanned aerial vehicle (UAV) images [59], concluding that a patch size of 160 x 160 pixels provided the best total accuracy. Additionally, studies show that some degree of overlap between windows is beneficial to the classification as it reduces the loss of contextual information along image patch borders [60].

This study aimed to investigate the use of DL algorithms to map burnt area changes within the Cerrado region to provide an accurate automated classification method. This research evaluated three CNN models based on the concept of the Autoencoder architecture: (a) the basic Autoencoder, (b) the U-Net, and (c) the ResUnet architectures, which propose improvements over the basic Autoencoder. Furthermore, we tested four sampling strategies to find optimal sampling window sizes for this specific classification task. In the following

sections, we describe the study area; our dataset structure, how the models were built, how they were evaluated and lastly, we present the results found and a brief discussion over them.

2. Methodology

2.1. Landsat data

Our training and testing datasets were created by collecting Tier 1 atmospherically corrected reflectance data detected by the Landsat 8 OLI+ sensors and pre-processed by the United States Geological Survey (USGS) agency. This study used bands 2 to 7, which offer the majority of the spectral information relevant to the detection of burnt lands for the same 30-meter spatial resolution. The training used the Landsat scenes (path-row) 221-71 and 221-70 (sites A and B), while the validation used scene 221-69 (site C) (**Figure 1**). The areas fit into the Cerrado biome and offer detection dates on the same day. The overlapping region between scenes B and C was excluded from scene B to avoid sharing data between training and validation.

To detect changes, we selected four different dates in August and September 2017 (August 9 and 25, September 10 and 26). The choice of date was based on the more significant occurrence of fires in the region during the end of the dry season [37]. Additionally, these dates offered the least amount of cloud cover throughout the year.

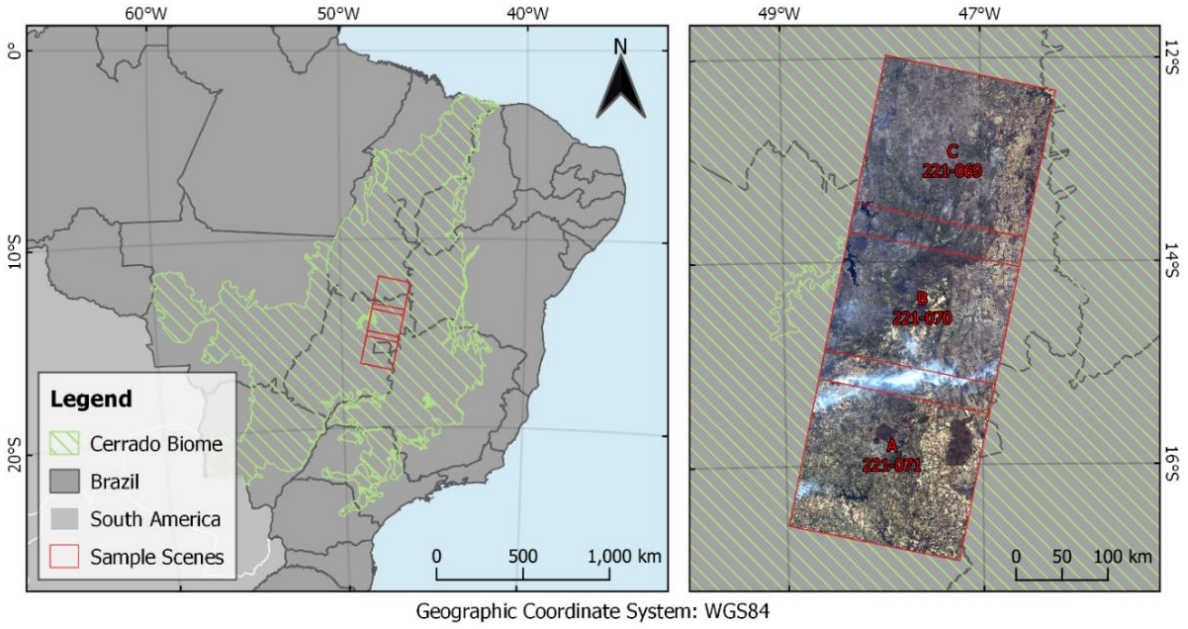

Figure 1. Delimitation of the (A, B) training and (C) validation sites within the Brazilian Cerrado region.

2.2. Burnt area change mask

The elaboration of the ground truth mask used the Normalized Burn Ratio (NBR) spectral index (Equation 1), which has been extensively used in research to highlight burnt areas and assess burn severity [38,61,62].

$$NBR = \frac{NIR-SWIR}{NIR+SWIR} \ , \quad (1)$$

where NIR and SWIR are the near and shortwave infrared bands, respectively. The NBR temporal difference (ΔNBR) can then be calculated to further highlight the burnt areas (Equation 2).

$$\Delta NBR = NBR_{T1} - NBR_{T2} \ , \quad (2)$$

where T1 and T2 are the pre-fire and post-fire images respectively. Specific ΔNBR threshold values allow assessing the severity of the burn. In this study, we classified pixels with ΔNBR values above 0.1 as burnt areas, regardless of the severity (**Figure 2**). Common false positives such as bodies of water and shadows, which are often also highlighted by this approach, were manually removed from the masks to guarantee that only burnt areas were present. This approach only detected new-burnt areas between two consecutive images without accounting for the accumulated burnt area.

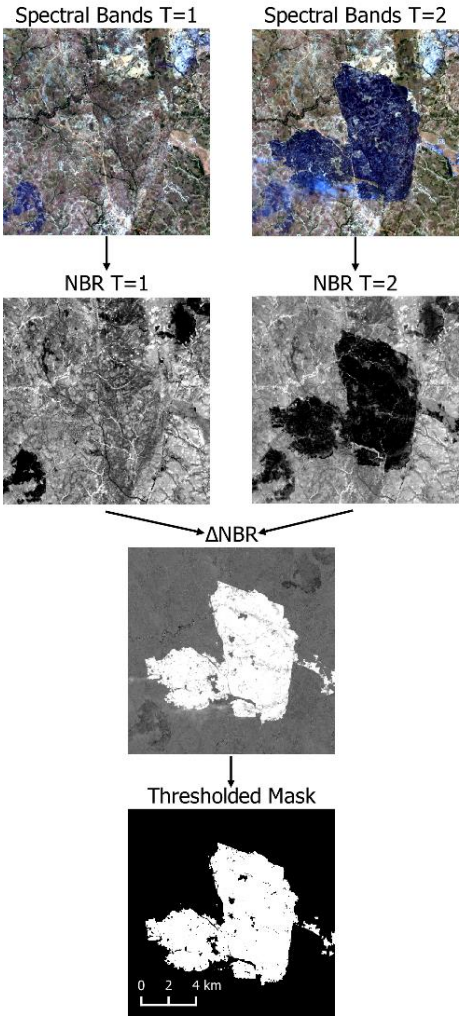

Figure 2. Visualization of the Normalized Burn Ratio thresholding process to generate the burnt area masks. From top to bottom, the true color satellite images, the calculated NBR values, the difference NBR between T1 and T2, and the thresholded burnt area mask.

2.3. Data structure

In this study, we used a bi-temporal approach in order to detect burnt area change. Therefore, images on two consecutive dates were paired and stacked depth-wise, generating a 12-band file associated with the respective change mask. Given our available images, this process generated three sets of bi-temporal images for each of the Landsat scenes.

Additionally, since data are structured in a batch by batch basis for deep learning models, our images had to be restructured and sampled as a 4D tensor containing multiple image patches and with shape [S x H x W x B] where S is the number of samples, H and W the height and width of the patches in number pixels and B the number of bands in the bi-temporal image pair. In this study, we sampled the images through a sliding window of four different sizes based on power of two (2^n) numbers: (a) 512 by 512, (b) 256 by 256, (c) 128 by 128, and (d) 64 by 64 pixels (**Figure 3**).

Furthermore, we used a 12.5% overlap between sampling windows to reduce the loss of predictive power near sample edges, an effect that is induced by the padding operation in convolutional layers and by the lack of contextual information near the patch edges. Incomplete windows, i.e., with empty pixel values, were discarded. The total number of samples per image generated through this process was 168, 747, 3140, and 12860, respectively.

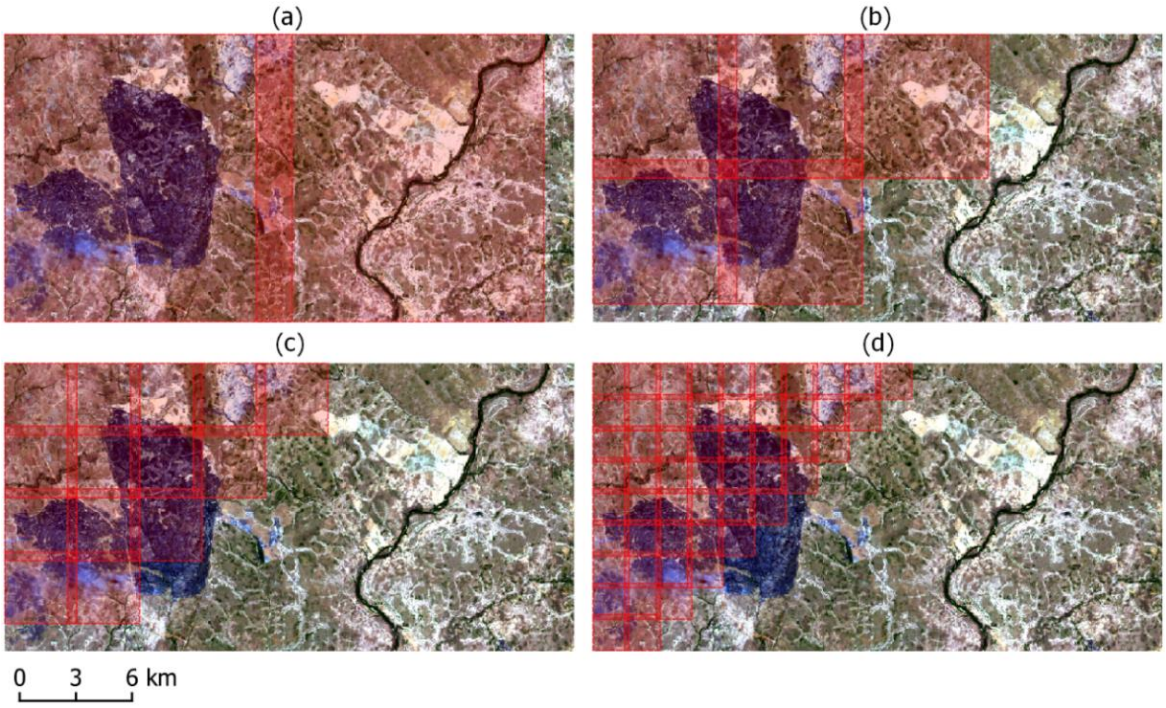

Figure 3. Example of the four different overlapping window sizes used to sample our images for the deep learning framework. Height and width in pixels of (a) 512, (b) 256, (c) 128 and (d) 64 respectively.

2.4. Deep learning models

The basis for the models used in this study was the Autoencoder model, which consists of an architecture that downsamples (decodes) the feature maps generated through convolutional layers to learn features compactly and then upsamples (encodes) them back to the desired output size. This process usually leads to the loss of spatial information as the feature maps are downsampled. The U-Net architecture[63] can be considered an evolution of the basic Autoencoder model, which tries to correct the loss of spatial information through the introduction of residual connections that propagate the information before being downsampled towards the upsampling layers. This allows the model to learn low-level detail while also keeping the high-level semantic information. A further enhancement of the architecture has been proposed through the insertion of residual connections within the architecture's blocks, resulting in what has been called ResUnet [51,64]. These three architectures have been used to classify remote sensing data before with good results [48–

51,64]. We adapted and evaluated these three architectures to describe possible differences when used to detect burnt area changes. **Figure 4** describes the general structure of the models used.

In this study, the architecture of the three models have the same number of layers and basic structure. Still, they differ in the way the residual connections are used: (a) the Autoencoder uses no connections at all, (b) the U-Net architecture uses connections only between blocks from both sides of the structure, and (c) the ResUnet uses connections between and within the blocks. This allowed us to evaluate the effect of the addition of the residual connections. The Keras [65] Python framework was used to build and train the models and to classify the images.

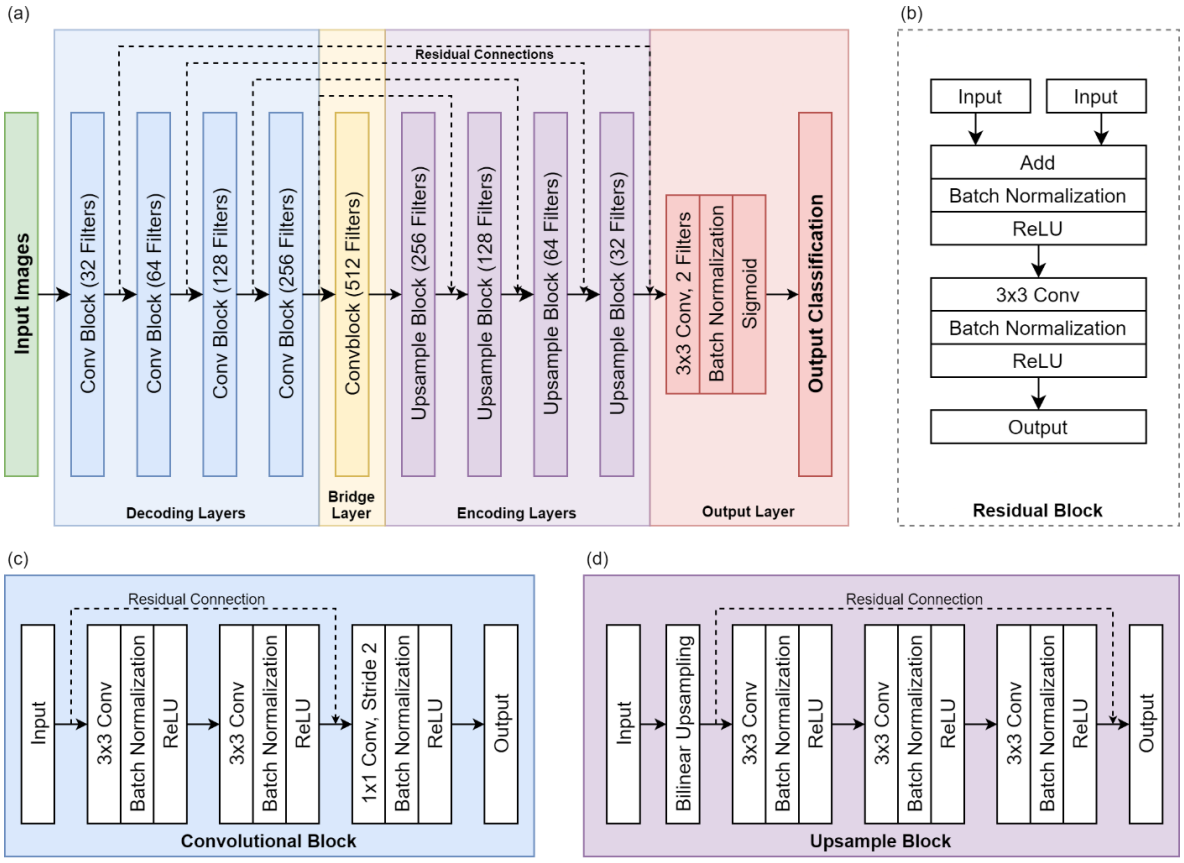

Figure 4. Illustration of the structures for the architectures used in this study. (a) The full general architecture and the (b) residual block, (c) convolutional block, and (d) the upsampling block structures. The use of residual connections determines the model, where the Autoencoder, U-Net, and ResUnet use: no connections, connections between blocks, and connections between and within blocks, respectively.

2.5. Model training

The three models shared most of the training parameters. To compute loss, we used the sum of the Binary Cross Entropy (BCE) loss and the Dice loss [66] functions (Equations 3, 4, and 5).

$$Final\ Loss = BCE\ loss + Dice\ loss\ , \tag{3}$$

where:

$$BCE\ loss = \frac{1}{m}\sum_{i=1}^{m} -(y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i))\ , \tag{4}$$

$$Dice\ loss = 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{2 * \sum(y_i * p_i)}{\sum y_i + \sum p_i} \tag{5}$$

where m is the number of mini-batches, y are the ground truth class values, $\hat{y}$ are the class scores from the sigmoid activation, and p are the predicted class values. The Dice loss comes from the Dice coefficient, also known as F1, which is especially useful for classifications with uneven class distributions (as in this study). This coefficient equally values positive and negative cases without the need to set arbitrary weights.

The gradient descent optimization used the RMSprop algorithm with a learning rate of 10^{-4} that automatically decreased by a magnitude of 10 every time the loss reached a plateau, to a minimum of 10^{-6}. The models were trained with the data from scenes A and B for a total of 200 epochs, which was enough to stabilize model loss and error for every model instance. The main differing parameter for model training was the batch size, which varied depending on the sample size as the available hardware memory constrained it. The batch sizes used were of 4, 8, 16, and 32 respectively for the window sizes of 512, 256, 128, and 64. Those were the largest possible batch sizes that allowed us to fit the samples into memory for their respective window sizes without reducing the number of samples. The model training used a computer equipped with Nvidia GeForce RTX 2080 TI graphics card with 11 GB of GPU memory, 16 GB of RAM memory, and an Intel Core i7-4770K CPU.

2.6. Model evaluation

As mentioned before, model validation used the data from scene C. The three main metrics used to evaluate the models were the F1 measure, the Kappa coefficient, and the mean Intersection Over Union (mIoU) value, represented by equations 6, 9, and 12, respectively.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}\ , \tag{6}$$

where:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}\ , \tag{7}$$

$$Recall = \frac{True\ Positives}{True\ Positives+False\ Negatives}\ , \quad (8)$$

$$Kappa = \frac{p_o-p_e}{1-p_e}\ , \quad (9)$$

where p_o is the rate of agreement between the ground truth and the classification, and p_e is the expected rate of random agreement (Equations 10 and 11):

$$p_o = \frac{True\ Positives+True\ Negatives}{True\ Positives+False\ Positives+True\ Negatives+False\ Negatives}\ , \quad (10)$$

$$p_e = \frac{(TP+FN)*(TP+FP)+(FP+TN)*(FN+TN)}{(TP+FN+TF+FP)^2}\ , \quad (11)$$

$$mIoU = \frac{IoU_1+IoU_2\cdots+IoU_n}{n}\ , \quad (12)$$

where IoU is the area of the intersection divided by the area of the union of the classification and ground truth for a class, and n is the total number of classes. All three of these measures range from 0 to 1, where a result of 1 would represent a perfect classification. In this study, they provide a better quantitative assessment over the traditional accuracy value, which tends to be misleadingly optimistic in classifications with an imbalanced number of observations and a large number of background (negative) cases relative to the foreground (positive) cases [67,68].

In addition, we employed McNemar's test [69] to evaluate whether the models were significantly different between each other. This test is a non-parametric test that mainly evaluates whether the error distribution between two classifications is similar. In this study, we used the variation of the test based on a chi-square distribution with a single degree of freedom and continuity correction [70] (Equation 13).

$$X^2 = \frac{(|f_{12}-f_{21}|-1)^2}{f_{12}+f_{21}}\ , \quad (13)$$

where f_{12} and f_{21} are the frequency of observations in disagreement between two classifications in a contingency table. A p-value of 0.05 was used as the threshold value, where lower values indicate that the distributions between two compared models are significantly different.

3. Results

Table 1 lists the detailed validation results for the F1, Kappa and mIoU measures for each model, while **Figure 5** shows a visual comparison of these measures. The basic Autoencoder architecture showed the worst results overall, although it showed good F1, Kappa, and mIoU values.

Table 1. Evaluation metrics for each instance of the models separated by the time sequence and window sizes. Best results in each column highlighted in bold text.

	08-09 to 08-25			08-25 to 09-10			09-10 to 09-26			Average		
Model	**Kappa**	**F1**	**mIoU**	**Kappa**	**F1**	**mIoU**	**Kappa**	**F1**	**mIoU**	**Kappa**	**F1**	**mIoU**
$Autoencoder_{64}$	0.823	0.825	0.849	0.844	0.845	0.865	0.849	0.851	0.868	0.839	0.840	0.861
$Autoencoder_{128}$	0.848	0.850	0.868	0.845	0.846	0.865	0.854	0.856	0.872	0.849	0.850	0.868
$Autoencoder_{256}$	0.863	0.865	0.879	0.865	0.866	0.881	0.868	0.870	0.883	0.865	0.867	0.881
$Autoencoder_{512}$	0.870	0.872	0.885	0.876	0.877	0.889	0.879	0.881	0.892	0.875	0.877	0.889
$U\text{-}Net_{64}$	0.889	0.890	0.900	0.920	0.920	0.926	0.922	0.923	0.927	0.910	0.911	0.918
$U\text{-}Net_{128}$	0.903	0.904	0.912	0.942	0.942	0.945	0.944	0.945	0.947	0.930	0.930	0.934
$U\text{-}Net_{256}$	**0.962**	**0.963**	**0.964**	0.959	0.959	0.960	0.960	0.961	0.962	**0.960**	**0.961**	**0.962**
$U\text{-}Net_{512}$	0.939	0.940	0.943	0.940	0.940	0.943	0.954	0.955	0.956	0.945	0.945	0.948
$ResUnet_{64}$	0.809	0.811	0.839	0.911	0.912	0.918	0.925	0.926	0.930	0.882	0.883	0.896
$ResUnet_{128}$	0.921	0.922	0.927	0.942	0.942	0.945	0.950	0.950	0.952	0.937	0.938	0.941
$ResUnet_{256}$	0.953	0.953	0.955	**0.963**	**0.964**	**0.965**	**0.962**	**0.963**	**0.964**	0.959	0.960	0.961
$ResUnet_{512}$	0.843	0.844	0.864	0.924	0.925	0.930	0.882	0.884	0.894	0.883	0.884	0.896

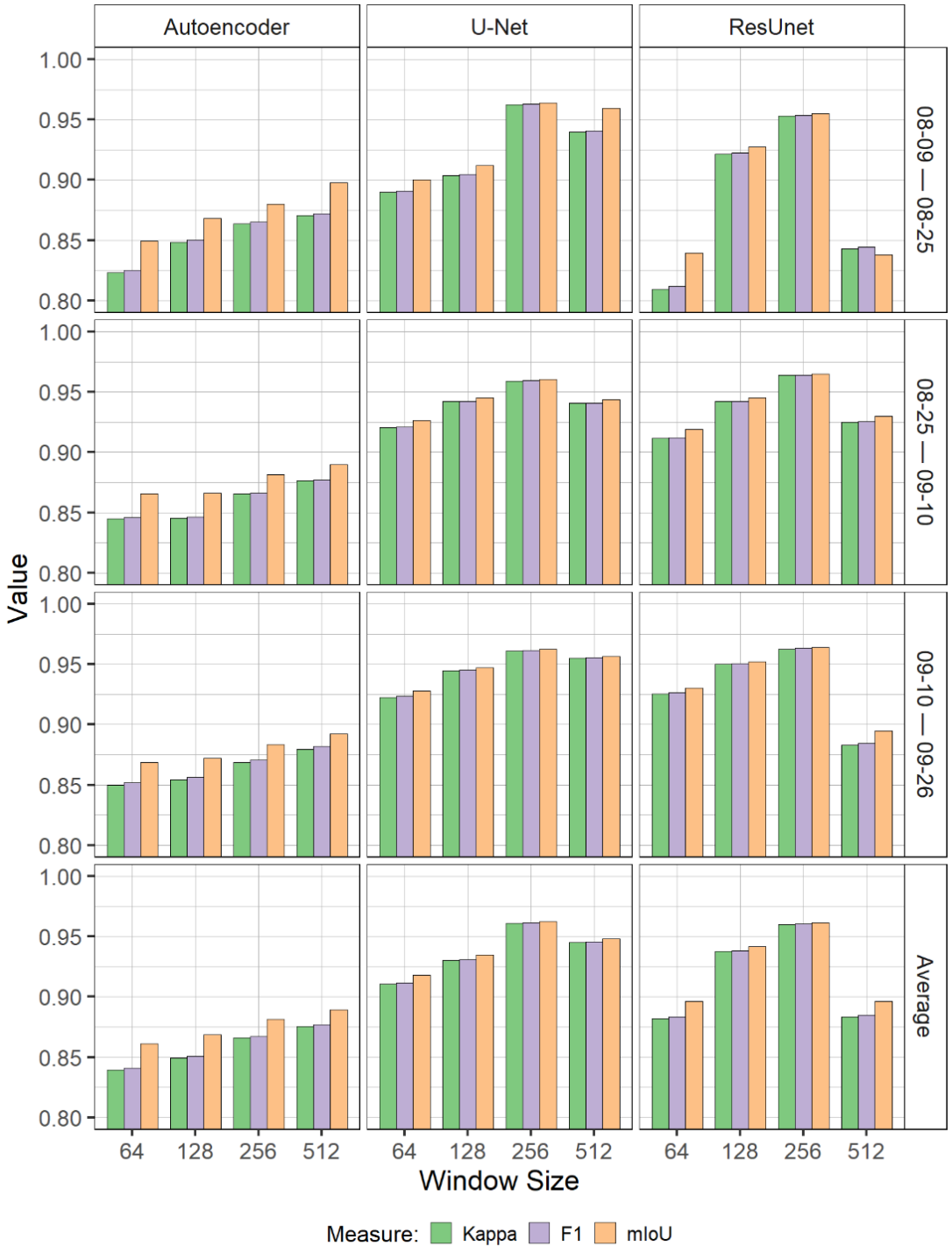

Figure 5. Visual comparison of the performance measures between the trained models for each time sequence and the average values with varying sample window sizes.

The U-Net and ResUnet architectures showed similar results. Despite performing better on average, the U-Net showed worse results in the time sequences of 08-25 to 09-10 and 09-10 to 09-26.

The models trained with samples with a size of 64 by 64 showed the worst results overall. The basic Autoencoder was the only model to show an improvement still when the window

size was increased to 512 by 512. The ResUnet model showed a more marked loss of performance, increasing the window size to 512 by 512. Using a window size of 256 by 256 resulted in the best F1, Kappa, and mIoU values for both the U-Net and ResUnet models.

In most cases, the models produced more false positives than false negatives (**Figure 6**). Improvements in the performance measures seemed to stem mainly from decreases in the number of false-positive predictions as the window sizes grew to 256 by 256. Comparatively, the number of false negatives varied little. However, the ResUnet model showed a noticeable increase in false negatives with the window size of 512 by 512. The time sequence between 08-25 and 09-10 showed the lowest amount of incorrectly classified pixels, which is explained by the fact that this sequence also showed the lowest extent of burnt areas overall.

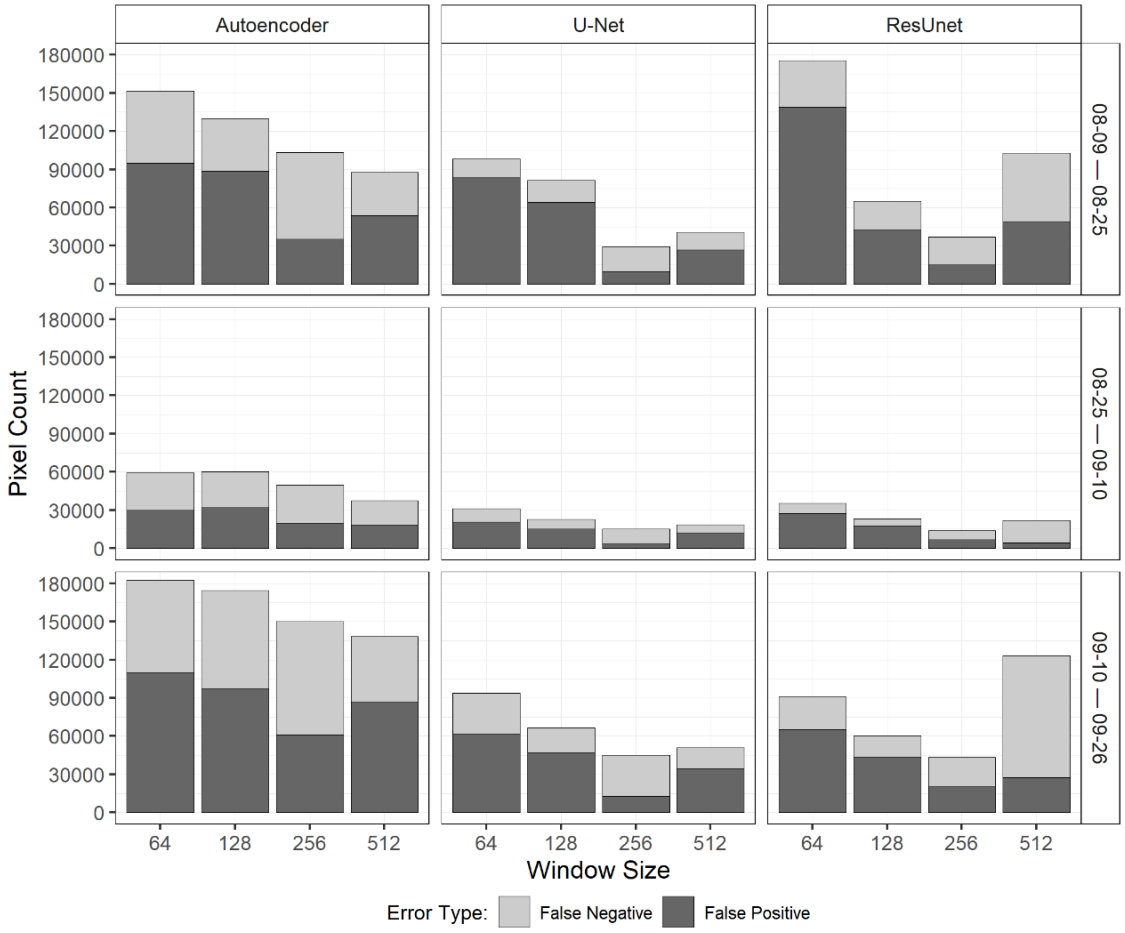

Figure 6. Error distributions in number of pixels for all model instances by time sequence and sample window size.

McNemar's test shows that when compared, most models have significantly different error distributions, which means the observed differences in the results did not occur at random (**Table 2**). The only models found to be statistically similar were the 256-window U-Net and ResUnet. The differences in error distributions are also visually noticeable in the

classification maps, particularly at the edges of burnt area patches. While the basic Autoencoder misclassified large groups of pixels, the U-Net and ResUnet models showed misclassifications mostly as very small groups or single pixels.

There were no noticeable differences between time sequences despite the physical and phenological changes (**Figures 7, 8, and 9**). The models automatically masked water bodies and most shadows, which are spectrally similar to burnt areas. However, burnt area patches with cloud cover in either image in a sequence were still sources of error in the classifications. Despite that, cloud shadows in unburnt lands were still correctly classified as negatives in most cases. Both the U-Net and ResUnet models were able to classify unclouded patches of burnt land with a low occurrence of errors.

Table 2. *p* values of the McNemar's test for comparing model classifications. Values under 0.05 indicate the error distribution from the two compared models are significantly different. (* indicates models that were statistically similar).

Model/Window		Autoencoder				U-Net				ResUnet			
		64	**128**	**256**	**512**	**64**	**128**	**256**	**512**	**64**	**128**	**256**	**512**
Autoencoder	**64**												
	128	<0.001											
	256	<0.001	<0.001										
	512	<0.001	<0.001	<0.001									
U-Net	**64**	<0.001	<0.001	<0.001	<0.001								
	128	<0.001	<0.001	<0.001	<0.001	<0.001							
	256	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001						
	512	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001					
ResUnet	**64**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001				
	128	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001			
	256	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.089*	<0.001	<0.001	<0.001		
	512	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	

4. Discussion

Results have shown that the DL models evaluated offer excellent classification results when used to detect burnt area changes. Even the worst model using the Autoencoder

architecture with a sample window of 64 by 64 resulted in F1, Kappa, and mIoU values over 0.8, which can already be considered a good result. The addition of residual connections between the decoding and encoding layers in the U-Net significantly improved the results. In contrast, the ResUnet's addition of connections within individual blocks gave marginal improvements and only in some cases. Visually, the Autoencoder's lack of connections translated in a noticeable loss of spatial information in the form of less detailed contours between the positive and negative classes. Overall, the U-Net architecture showed the best results, although not much higher than the ResUnet architecture, which was superior in the time sequences with greater extensions of burnt land. Regions with clouded patches of burnt land were among the primary sources of errors in the classifications as occasionally, the models misclassified the cloud shadows as extensions of the burnt land patches, generating false positives. Despite that, the models correctly classified regions without mixed burnt areas and cloud shadows and automatically masked objects commonly detected as false positives through ΔNBR thresholding, therefore reducing the need for human intervention. The presence of cloud cover is one of the main limitations when using Landsat data for change detection as it is a common occurrence that impacts both the creation of a ground truth mask and the training of the models. Radar data can be used instead but at the cost of a significant loss of spectral information and possibly accuracy [71]. Studies have been carried using CNNs and Synthetic Aperture Radar (SAR) data to detect burnt areas with results similar to those found in this study [46,72] although in much smaller extents. Our bi-temporal approach was similar to that used by the Brazilian Institute of Space Research (INPE) to produce official burnt area reports [73]. However, our use of DL architectures instead of a thresholding process produced a much lower rate of false positives (commission errors) and false negatives (omission errors). Furthermore, DL models can be trained incrementally with new training data and further improve results, although up to a specific limit.

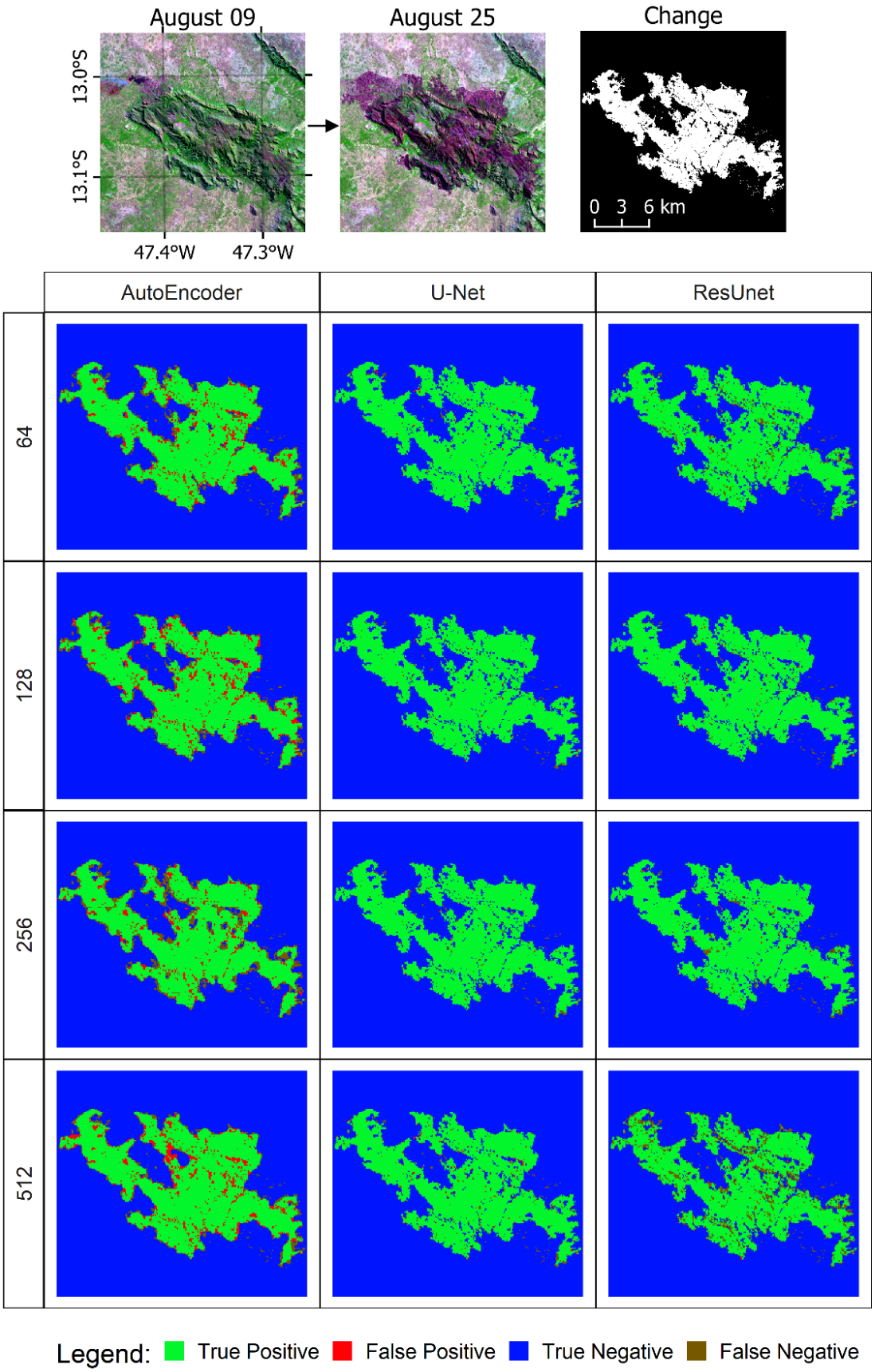

Figure 7. Example of a classified burnt area patch of the August 9 to August 25 sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.

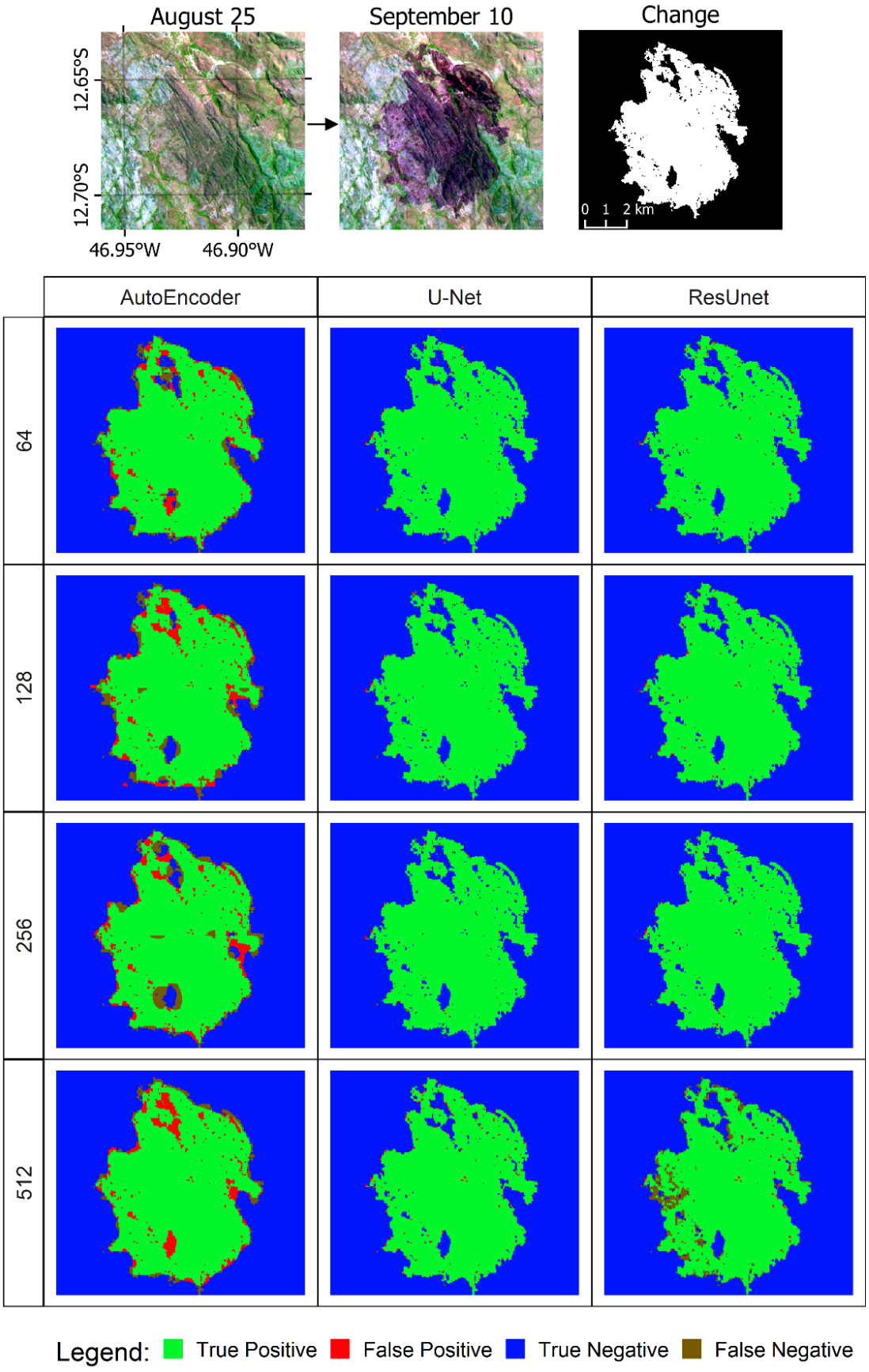

Figure 8. Example of a classified burnt area patch of the August 25 to September 10 sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.

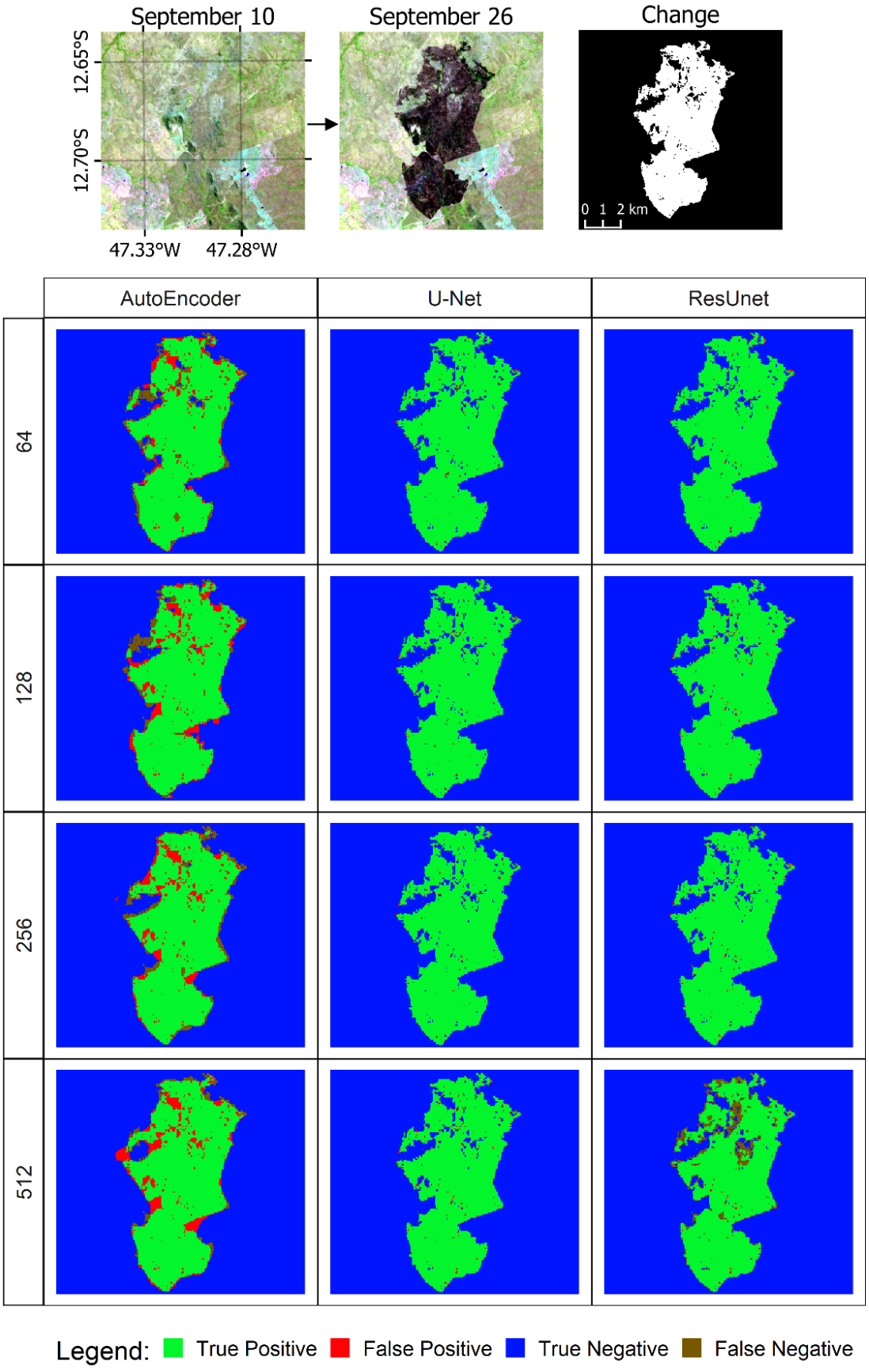

Figure 9. Example of a classified burnt area patch of the September 10 to September 26 sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.

Increasing the sample window size improved the results despite simultaneously decreasing the training batch sizes, although only up to 256 by 256 in the case of the U-Net and ResUnet models. The size of 512 by 512 worsened the results, particularly for the ResUnet model, which showed results close to the same model using the 64 by 64 samples. The cause can be attributed to the lower batch size used. Studies have shown that smaller batch sizes can introduce more noise in the training gradients, leading to a loss of generalizability and, therefore, less accuracy [74,75]. However, the window size of 64 by 64 showed the worst results overall, even using the largest batch size, which shows that there is possibly a balance between sample window size and batch size. This problem is ultimately limited by the quantity of memory available in the graphics card, which determines the number of samples, the size of samples, and the batch size that can be used in the same training process. In addition, increasing the model complexity (e.g., by increasing the number of layers or filters) can exponentially increase the memory required for training. Remote sensing data can be highly memory intensive, especially at higher spatial and spectral resolutions, making the process of optimizing the training parameters for DL models challenging with consumer-grade hardware.

The loss of performance with smaller window sizes is also related to the size of the object at hand. While small window sizes might cover the full extent of small objects, they cannot fully cover larger objects, leading to less information about the relationship between the object of study and its surroundings. Given the way convolutional networks function, the information within each image patch is highly important. In this study the extent of burnt areas ranged from single pixels (900m²) to several square kilometers, and, as seen in **Figure 3**, the smaller window sizes created several image patches with low or no background-foreground context, i.e., without enough information about the burnt area border dynamics. Despite that, the results show that the sampling window does not necessarily need to cover the full extent of the object of detection, corroborating with results found in other studies [59].

5. Conclusions

In this study, we evaluated three Deep Learning models to detect burnt area changes in 3 bi-temporal Landsat image pairs: a basic Autoencoder, U-Net, and ResUnet. All three networks were based on the same principles but with differences in the use of residual connections. The training and validation of the models used Landsat data from scenes within the region of the Brazilian Cerrado. The models were trained with four different sample

window sizes in pixels to verify performance differences: 64 by 64, 128 by 128, 256 by 256, and 512 by 512.

Results have shown that the architectures used are a reliable automated way to map burnt area changes between bi-temporal image pairs in the Cerrado. However, the U-Net and ResUnet models were superior to the basic Autoencoder as the introduction of residual connections significantly improved the results. The sample window size of 256 by 256 pixels showed the best results for the U-Net and ResUnet models, and further increasing it produced worse results for both of these models. The model evaluation considered the F1, Kappa, and mIoU measures, of which the 256 by 256 window U-Net model achieved the best overall results with average values of 0.960, 0.961, and 0.962 respectively. The ResUnet model had slightly worse results on average, but slightly better results in two of the three time sequences evaluated. McNemar's test verified the possibility that the differences between classifications were not statistically significant, and only the U-Net and ResUnet models using 256x256-pixel samples were found to be similar while every other model was statistically unique.

The Cerrado biome is an important region given its biodiversity, but it is constantly under the threat of destruction through fires and deforestation. We recommend that future studies investigate more uses of current Deep Learning techniques to provide better solutions for the detection and mitigation of these threats. In addition, certain spectral vegetation and burn indexes that have been shown to possibly improve the detection of burnt land [76,77] were not used in this study and could be investigated in future works. A few other suggestions for further studies of the theme arose from certain limitations found in this study. We recommend an investigation of the effect of training batch sizes along with sample window sizes. The batch size is an essential factor in the model performances but is limited by memory, along with several other parameters relevant to DL models. Secondly, the presence of cloud cover in Landsat images is a source of error. Radar data partly solves this problem at the expense of spectral information. Therefore, the possibility of the use of mixed sensor data for burnt area mapping should be investigated. Researchers might also find it interesting to compare the performance of DL algorithms to shallow ML algorithms in a direct approach to further highlight the difference in performances.

References

1. Zhang, X.; Zhao, J.; LeCun, Y. Character-level Convolutional Networks for Text Classification. In *Advances in Neural Information Processing Systems 28*; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc., 2015; pp. 649–657.

2. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* **2014**, *abs/1409.1*.
3. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 221–231, doi:10.1109/TPAMI.2012.59.
4. Abdel-Hamid, O.; Mohamed, A.; Jiang, H.; Penn, G. Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE, 2012; pp. 4277–4280.
5. Zhang, L.; Zhang, L.; Kumar, V. Deep learning for Remote Sensing Data. *IEEE Geoscience and Remote Sensing Magazine* **2016**, *4*, 22–40, doi:10.1155/2016/7954154.
6. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine* **2017**, *5*, 8–36, doi:10.1109/MGRS.2017.2762307.
7. Mou, L.; Ghamisi, P.; Zhu, X.X. Fully conv-deconv network for unsupervised spectral-spatial feature extraction of hyperspectral imagery via residual learning. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS); IEEE, 2017; Vol. 56, pp. 5181–5184.
8. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 778–782, doi:10.1109/LGRS.2017.2681128.
9. Scott, G.J.; England, M.R.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 549–553, doi:10.1109/LGRS.2017.2657778.
10. Imamoglu, N.; Kimura, M.; Miyamoto, H.; Fujita, A.; Nakamura, R. Solar Power Plant Detection on Multi-Spectral Satellite Imagery using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion. *CoRR* **2017**, *abs/1704.0*.
11. Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *International Journal of Computational Intelligence and Applications* **2017**, *16*, 1750001, doi:10.1142/S1469026817500018.
12. Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-Sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2018**, *11*, 978–989, doi:10.1109/JSTARS.2018.2794888.
13. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, 1–15, doi:10.1109/TGRS.2018.2817393.
14. Alcantarilla, P.F.; Stent, S.; Ros, G.; Arroyo, R.; Gherardi, R. Street-view change detection with deconvolutional networks. *Auton Robot* **2018**, *42*, 1301–1322, doi:10.1007/s10514-018-9734-5.
15. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Trans. Neural Netw. Learning Syst.* **2016**, *27*, 125–138, doi:10.1109/TNNLS.2015.2435783.
16. Zhang, P.; Gong, M.; Su, L.; Liu, J.; Li, Z. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *116*, 24–41, doi:10.1016/j.isprsjprs.2016.02.013.

17. Zhao, J.; Gong, M.; Liu, J.; Jiao, L. Deep learning to classify difference image for image change detection. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN); IEEE: Beijing, China, 2014; pp. 411–417.
18. Zeng, X.; Yang, J.; Deng, X.; An, W.; Li, J. Cloud detection of remote sensing images on Landsat-8 by deep learning. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018); Jiang, X., Hwang, J.-N., Eds.; SPIE: Shanghai, China, 2018; p. 173.
19. Zhan, Y.; Wang, J.; Shi, J.; Cheng, G.; Yao, L.; Sun, W. Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 1785–1789, doi:10.1109/LGRS.2017.2735801.
20. Chuvieco, E.; Aguado, I.; Yebra, M.; Nieto, H.; Salas, J.; Martín, M.P.; Vilar, L.; Martínez, J.; Martín, S.; Ibarra, P.; et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling* **2010**, *221*, 46–58, doi:10.1016/j.ecolmodel.2008.11.017.
21. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; da Fonseca, G.A.B.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858, doi:10.1038/35002501.
22. INPE - Instituto Nacional de Pesquisas Espaciais Monitoramento de Queimadas Available online: http://www.inpe.br/queimadas (accessed on Nov 6, 2017).
23. Costafreda-Aumedes, S.; Comas, C.; Vega-Garcia, C. Human-caused fire occurrence modelling in perspective: a review. *International Journal of Wildland Fire* **2017**, *26*, 983, doi:10.1071/WF17026.
24. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sensing of Environment* **2019**, *225*, 45–64, doi:10.1016/j.rse.2019.02.013.
25. Chuvieco, E.; Lizundia-Loiola, J.; Pettinari, M.L.; Ramo, R.; Padilla, M.; Tansey, K.; Mouillot, F.; Laurent, P.; Storm, T.; Heil, A.; et al. Generation and analysis of a new global burned area product based on MODIS 250m reflectance bands and thermal anomalies. *Earth System Science Data* **2018**, *10*, 2015–2031, doi:10.5194/essd-10-2015-2018.
26. Daldegan, G.A.; de Carvalho Júnior, O.A.; Guimarães, R.F.; Gomes, R.A.T.; Ribeiro, F. de F.; McManus, C. Spatial patterns of fire recurrence using remote sensing and GIS in the Brazilian savanna: Serra do Tombador Nature Reserve, Brazil. *Remote Sensing* **2014**, *6*, 9873–9894, doi:10.3390/rs6109873.
27. Pereira, J.M.C. Remote sensing of burned areas in tropical savannas. *Int. J. Wildland Fire* **2003**, *12*, 259, doi:10.1071/WF03028.
28. Sousa, I.M.P.; Carvalho, E.V. de; Batista, A.C.; Machado, I.E.S.; Tavares, M.E.F.; Giongo, M. Identification of Burned Areas by Special Index in a Cerrado Region of the State of Tocantins, Brazil. *RF* **2018**, *48*, 553, doi:10.5380/rf.v48i4.57362.
29. de Carvalho Júnior, O.A.; Guimarães, R.F.; Silva, C.; Gomes, R.A.T. Standardized Time-Series and Interannual Phenological Deviation: New Techniques for Burned-Area Detection Using Long-Term MODIS-NBR Dataset. *Remote Sensing* **2015**, *7*, 6950–6985, doi:10.3390/rs70606950.
30. Pereira Júnior, A.C.; Oliveira, S.L.J.; Pereira, J.M.C.; Turkman, M.A.A. Modelling Fire Frequency in a Cerrado Savanna Protected Area. *PLoS ONE* **2014**, *9*, e102380, doi:10.1371/journal.pone.0102380.

31. Alvarado, S.T.; Fornazari, T.; Cóstola, A.; Morellato, L.P.C.; Silva, T.S.F. Drivers of fire occurrence in a mountainous Brazilian cerrado savanna: Tracking long-term fire regimes using remote sensing. *Ecological Indicators* **2017**, *78*, 270–281, doi:10.1016/j.ecolind.2017.02.037.
32. de Bem, P.P.; de Carvalho Júnior, O.A.; Matricardi, E.A.T.; Guimarães, R.F.; Gomes, R.A.T. Predicting wildfire vulnerability using logistic regression and artificial neural networks: a case study in Brazil's Federal District. *Int. J. Wildland Fire* **2019**, *28*, 35, doi:10.1071/WF18018.
33. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* **2017**, *61*, 539–556, doi:10.1016/j.patcog.2016.07.001.
34. Giglio, L.C.J. MCD64A1 MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V006 2015.
35. Hall, J.V.; Loboda, T.V.; Giglio, L.; McCarty, G.W. A MODIS-based burned area assessment for Russian croplands: Mapping requirements and challenges. *Remote Sensing of Environment* **2016**, *184*, 506–521, doi:10.1016/j.rse.2016.07.022.
36. Hawbaker, T.J.; Vanderhoof, M.K.; Beal, Y.-J.; Takacs, J.D.; Schmidt, G.L.; Falgout, J.T.; Williams, B.; Fairaux, N.M.; Caldwell, M.K.; Picotte, J.J.; et al. Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment* **2017**, *198*, 504–522, doi:10.1016/j.rse.2017.06.027.
37. Moreira de Araújo, F.; Ferreira, L.G.; Arantes, A.E. Distribution Patterns of Burned Areas in the Brazilian Biomes: An Analysis Based on Satellite Data for the 2002–2010 Period. *Remote Sensing* **2012**, *4*, 1929–1946, doi:10.3390/rs4071929.
38. Santana, N.; de Carvalho Júnior, O.; Gomes, R.; Guimarães, R. Burned-Area Detection in Amazonian Environments Using Standardized Time Series Per Pixel in MODIS Data. *Remote Sensing* **2018**, *10*, 1904, doi:10.3390/rs10121904.
39. Pereira, A.; Pereira, J.; Libonati, R.; Oom, D.; Setzer, A.; Morelli, F.; Machado-Silva, F.; de Carvalho, L. Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires. *Remote Sensing* **2017**, *9*, 1161, doi:10.3390/rs9111161.
40. Ramo, R.; Chuvieco, E. Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing* **2017**, *9*, 1193, doi:10.3390/rs9111193.
41. Mithal, V.; Nayak, G.; Khandelwal, A.; Kumar, V.; Nemani, R.; Oza, N. Mapping Burned Areas in Tropical Forests Using a Novel Machine Learning Framework. *Remote Sensing* **2018**, *10*, 69, doi:10.3390/rs10010069.
42. Al-Rawi, K.R.; Casanova, J.L.; Calle, A. Burned area mapping system and fire detection system, based on neural networks and NOAA-AVHRR imagery. *International Journal of Remote Sensing* **2010**, *22*, 2015–2032, doi:10.1080/01431160117531.
43. Meng, R.; Zhao, F. Remote sensing of fire effects. A review for recent advances in burned area and burn severity mapping. In *Remote sensing of hydrometeorological hazards*; Petropoulos, G.P., Islam, T., Eds.; CRC Press: Boca Raton, FL, USA, 2017; pp. 261–276.
44. Shan, T.; Wang, C.; Chen, F.; Wu, Q.; Li, B.; Yu, B.; Shirazi, Z.; Lin, Z.; Wu, W. A Burned Area Mapping Algorithm for Chinese FengYun-3 MERSI Satellite Data. *Remote Sensing* **2017**, *9*, 736, doi:10.3390/rs9070736.

45. Langford, Z.; Kumar, J.; Hoffman, F. Wildfire Mapping in Interior Alaska Using Deep Neural Networks on Imbalanced Datasets. In Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW); IEEE: Singapore, Singapore, 2018; pp. 770–778.
46. Zhang, P.; Nascetti, A.; Ban, Y.; Gong, M. An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *158*, 50–62, doi:10.1016/j.isprsjprs.2019.09.013.
47. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention. *Remote Sensing* **2019**, *11*, 1702, doi:10.3390/rs11141702.
48. de Bem, P.; de Carvalho Junior, O.; Fontes Guimarães, R.; Trancoso Gomes, R. Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks. *Remote Sensing* **2020**, *12*, 901, doi:10.3390/rs12060901.
49. Li, L. Deep Residual Autoencoder with Multiscaling for Semantic Segmentation of Land-Use Images. *Remote Sensing* **2019**, *11*, 2142, doi:10.3390/rs11182142.
50. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sensing* **2019**, *11*, 68, doi:10.3390/rs11010068.
51. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sensing Lett.* **2018**, *15*, 749–753, doi:10.1109/LGRS.2018.2802944.
52. Bermudez, J.D.; Happ, P.N.; Oliveira, D.A.B.; Feitosa, R.Q. SAR TO OPTICAL IMAGE SYNTHESIS FOR CLOUD REMOVAL WITH GENERATIVE ADVERSARIAL NETWORKS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *IV–1*, 5–11, doi:10.5194/isprs-annals-IV-1-5-2018.
53. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sensing* **2016**, *9*, 22, doi:10.3390/rs9010022.
54. Yohei Koga; Hiroyuki Miyazaki; Ryosuke Shibasaki A CNN-Based Method of Vehicle Detection from Aerial Images Using Hard Example Mining. *Remote Sensing* **2018**, *10*, 124, doi:10.3390/rs10010124.
55. Ma, H.; Liu, Y.; Ren, Y.; Wang, D.; Yu, L.; Yu, J. Improved CNN Classification Method for Groups of Buildings Damaged by Earthquake, Based on High Resolution Remote Sensing Images. *Remote Sensing* **2020**, *12*, 260, doi:10.3390/rs12020260.
56. Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network. *Remote Sensing* **2019**, *11*, 1774, doi:10.3390/rs11151774.
57. Liu, C.; Zeng, D.; Wu, H.; Wang, Y.; Jia, S.; Xin, L. Urban Land Cover Classification of High-Resolution Aerial Imagery Using a Relation-Enhanced Multiscale Convolutional Network. *Remote Sensing* **2020**, *12*, 311, doi:10.3390/rs12020311.
58. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing* **2018**, *10*, 1119, doi:10.3390/rs10071119.
59. Ammour, N.; Alhichri, H.; Bazi, Y.; Benjdira, B.; Alajlan, N.; Zuair, M. Deep Learning Approach for Car Detection in UAV Imagery. *Remote Sensing* **2017**, *9*, 312, doi:10.3390/rs9040312.
60. de Albuquerque, A.O.; de Carvalho Júnior, O.A. de C.; de Carvalho, O.L.F.; de Bem, P.P.; Ferreira, P.H.G.; de Moura, R. dos S.; Silva, C.R.; Gomes, R.A.T.; Guimarães, R.F. Deep Semantic Segmentation

of Center Pivot Irrigation Systems from Remotely Sensed Data. *Remote Sensing* **2020**, *12*, 2159, doi:10.3390/rs12132159.

61. Escuin, S.; Navarro, R.; Fernández, P. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *International Journal of Remote Sensing* **2008**, *29*, 1053–1073, doi:10.1080/01431160701281072.
62. Miller, J.D.; Thode, A.E. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sensing of Environment* **2007**, *109*, 66–80, doi:10.1016/j.rse.2006.12.006.
63. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]* **2015**.
64. Cao, K.; Zhang, X. An Improved Res-UNet Model for Tree Species Classification Using Airborne High-Resolution Images. *Remote Sensing* **2020**, *12*, 1128, doi:10.3390/rs12071128.
65. Chollet, F.; others *Keras*; 2015;
66. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *arXiv:1606.04797 [cs]* **2016**.
67. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* **2002**, *80*, 185–201, doi:10.1016/S0034-4257(01)00295-4.
68. Maratea, A.; Petrosino, A.; Manzo, M. Adjusted F-measure and kernel scaling for imbalanced data learning. *Information Sciences* **2014**, *257*, 331–341, doi:10.1016/j.ins.2013.04.016.
69. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157, doi:10.1007/BF02295996.
70. Foody, G.M. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *photogramm eng remote sensing* **2004**, *70*, 627–633, doi:10.14358/PERS.70.5.627.
71. Tanase, M.A.; Belenguer-Plomer, M.A.; Roteta, E.; Bastarrika, A.; Wheeler, J.; Fernández-Carrillo, Á.; Tansey, K.; Wiedemann, W.; Navratil, P.; Lohberger, S.; et al. Burned Area Detection and Mapping: Intercomparison of Sentinel-1 and Sentinel-2 Based Algorithms over Tropical Africa. *Remote Sensing* **2020**, *12*, 334, doi:10.3390/rs12020334.
72. Ban, Y.; Zhang, P.; Nascetti, A.; Bevington, A.R.; Wulder, M.A. Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Sci Rep* **2020**, *10*, 1322, doi:10.1038/s41598-019-56967-x.
73. Melchiori, A.E.; W. Setzer, A.; Morelli, F.; Libonati, R.; Cândido, P. de A.; Jesús, S.C. de A Landsat-TM/OLI algorithm for burned areas in the Brazilian Cerrado: preliminary results. In *Advances in forest fire research*; Imprensa da Universidade de Coimbra, 2014; Vol. 4, pp. 1302–1311 ISBN 978-989-26-0884-6.
74. Kandel, I.; Castelli, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express* **2020**, S2405959519303455, doi:10.1016/j.icte.2020.04.010.
75. Radiuk, P.M. Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets. *Information Technology and Management Science* **2017**, *20*, doi:10.1515/itms-2017-0003.

76. Axel, A. Burned Area Mapping of an Escaped Fire into Tropical Dry Forest in Western Madagascar Using Multi-Season Landsat OLI Data. *Remote Sensing* **2018**, *10*, 371, doi:10.3390/rs10030371.
77. Saulino, L.; Rita, A.; Migliozzi, A.; Maffei, C.; Allevato, E.; Garonna, A.P.; Saracino, A. Detecting Burn Severity across Mediterranean Forest Types by Coupling Medium-Spatial Resolution Satellite Imagery and Field Data. *Remote Sensing* **2020**, *12*, 741, doi:10.3390/rs12040741.

Capítulo III

Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks

Resumo: O mapeamento do desmatamento é um passo essencial no processo de manejo de florestas tropicais, fornecendo base para um melhor conhecimento do processos de desmatamento e suas implicações. Considerando-se que existe amplo espaço para melhorias no processo de elaboração de mapas de áreas desmatadas por via de imagens de satélite, o objetivo deste estudo foi testar e avaliar o uso de algoritmos de Deep Learning (DL) para este fim, especificamente através de Redes Neurais Convolucionais (CNNs). Apesar da ampla aplicação de algoritmos de DL no campo do sensoriamento remoto, seu uso para mapear desmatamento ainda é relativamente inexplorado. Neste estudo os algoritmos foram utilizados para mapear áreas desmatadas em regiões com imagens em anos consecutivos, especificamente entre 2017 e 2018, e entre 2018 e 2019. Três arquiteturas CNN presentes na literatura foram utilizadas para classificar as mudanças: SharpMask, U-Net e ResUnet. Adicionalmente essas arquiteturas foram comparadas a dois algoritmos tradicionais de Machine Learning (ML): Random Forest (RF) e Multilayer Perceptron (MLP). Resultados após validação indicam que os modelos de DL foram melhores na maior parte das medidas de avaliação, incluindo os índices Kappa, F1 e mean Intersecion over Union (mIoU). O modelo ResUnet obteve os melhores resultados em geral, com um valor de 0.94 em todas as três medidas em ambas as sequencias anuais. Visualmente, os modelos de DL forneceram classificações com áreas desmatadas melhor definidas e não necessitou de nenhum tipo de pós processamento para remoção de ruído, diferentemente dos modelos tradicionais de ML.

Artigo publicado na revista *Remote Sensing*, 2020. DOI: 10.3390/rs12060901

Change Detection of Deforestation in The Brazilian Amazon Using Landsat Data and Convolutional Neural Networks

Abstract: Mapping deforestation is an essential step in the process of managing tropical rainforests. It lets us understand and monitor both legal and illegal deforestation and its implications, which include the effect deforestation may have on climate change through greenhouse gas emissions. Given that there is ample room for improvements when it comes to mapping deforestation using satellite imagery, in this study we aimed to test and evaluate the use of algorithms belonging to the growing field of Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), to this end. Although studies have been using DL algorithms for a variety remote sensing tasks for the past few years, they're still relatively unexplored for deforestation mapping. We attempted to map the deforestation between images approximately one year apart, specifically between 2017 and 2018 and between 2018 and 2019. Three CNN architectures available in literature — SharpMask, U-Net and ResUnet — were used to classify the change between years and were then compared to two classic Machine Learning (ML) algorithms — Random Forest (RF) and Multilayer Perceptron (MLP) —as points of reference. After validation, we have found that the DL models were better in most performance metrics, including the Kappa index, F1 score and mean Intersection over Union (mIoU) measure, while the ResUnet model achieved the best overall results with a value of 0.94 in all three measures in both time sequences. Visually the DL models also provided classifications with better defined deforestation patches and did not need any sort of post-processing to remove noise, unlike the ML models which needed some noise removal to improve results.

Keywords: Deep learning; CNN; classification; change detection; deforestation

1 Introduction

Deforestation is one of the primary sources of concern regarding climate change as it is one of the largest sources of greenhouse gas emissions in the world, second only to the burning of fossil fuels [1]. Within the region of the Brazilian Amazon, studies have shown that deforestation, in conjunction with forest fires, can make up for up to 48% of the total emissions [2]. It also bears substantial implications regarding the conservation of ecosystems and their biodiversity in the region, and it has been linked to the loss of species [3] and general loss of ecosystem stability

through fragmentation [4]. Locally, estimates also show that unchecked deforestation could lead to reductions in seasonal rainfall [5] and into the savanization of the environment [6].

Remote sensing imagery has been instrumental in the process of keeping track of deforestation in the Amazon. The Brazilian National Institute for Space Research (INPE) releases annual deforestation and land use information derived from satellite imagery data through their Program for Deforestation Monitoring (PRODES) and TerraClass projects [7,8], which have been widely used for monitoring, research and policymaking. Carbon emission estimates from deforestation are also dependent on land use and land-use change data [1]. However, they are likely to be underestimated due to the omission of illegal logging data in official reports [9].

Change detection is one of the most common goals within the field of remote sensing. It is defined as the process of analyzing and quantifying the state of an object or phenomenon at different times [10], and it is consequently an essential tool in the processes of understanding and tackling deforestation. The changes present in the images can be semantic (of the object under analysis) or noisy (variations in lighting, shadows, among others) [11]. Therefore, the challenge in change detection is to use a method that establishes features that minimize noisy changes and emphasize the semantic changes that are intertwined. Typically, the final map of change detection techniques is a binary classification that contains unchanged and changed regions.

Several literature reviews and classifications of digital change detection techniques have been proposed [10,12-18], evidencing a growing expansion of approaches and algorithms in this research area. Tewkesbury et al. [18] provide a synthesis of change detection methods, distinctly considering the unit of analysis (pixel, kernel, image-object overlay, image-object comparison, multi-temporal image-object, vector polygon, and hybrid) and the method used to identify the change (layer arithmetic, post-classification change, direct classification, transformation, change vector analysis, hybrid change detection).

Change detection methods based on Machine Learning (ML) algorithms typically use direct classification [18], which takes a set of stacked temporal images as input and uses complex nonlinear functions to determine changes. In this approach, it is not necessary to use pre-classification techniques that seek to define the best measures to detect changes (such as temporal subtraction, data transformation and change vector analysis). In long-term time series, the direct classification based on ML is predominant [19,20].

Deep learning (DL) has recently attracted increasing attention from remote sensing researchers because of its ability to automatically extract features from the image dataset; high-level semantic segmentation; nonlinear problem modeling; and mapping in complex environments [21]. DL has shown great potential in remote sensing, producing state-of-the-art results in different types of remote sensing data processing [22]: image registration [23-26], land-use and land-cover classification [27–30], object detection [31–34], image fusion [35-38], semantic segmentation [39-42], and precision evaluation [43]. DL has also been used for change detection techniques, showing superior performance with greater precision in comparison to classic ML methods [44]. The capacity for pattern recognition in the three dimensions of the image (special, spectral, and temporal) makes DL algorithms especially effective when used to change detection with common and recurring patterns [45]. DL-based change detection methods have been applied to different targets such as: urban [46-49], land use/land cover [50-52], landslides [53], among others. Among DL algorithms, the convolutional neural network (CNN) is the leading architecture [22]. CNNs differ from traditional ML algorithms by being able to identify patterns within an n-dimensional context with multiple abstraction levels through convolutional filters and use them for inference.

Peng et al. [54] propose a subdivision of DL-based change detection methods, considering three units of analysis: (1) feature [55-57]; (2) patch [58-61]; and (3) image [62,63]. In the case of image-based DL change detection, the learning for the segmentation of the changes comes directly from the bi-temporal image pairs, reducing the effect of the pixel patches [54]. In this approach, the U-Net architecture has been successfully employed [63,64].

The objective of this study was to investigate the use of CNNs for the detection of deforestation within the Brazilian Amazon to verify the hypotheses that DL algorithms are a viable and possibly better alternative in comparison to classic ML algorithms when it comes to mapping deforestation. Like many anthropogenic changes in the landscape, deforestation follows specific spatial patterns with support geometric or regular configurations and usually develop around official or unofficial roads, forming a dendritic or “fishbone” distribution [65]. Despite being a prime target for the application of DL algorithms, the number of studies related to deforestation is still small given the variety of types of algorithms available. In order to investigate the use of DL for deforestation detection, three different CNN architectures were used to classify deforested areas yearly and then compared to two classical ML algorithms as points of reference.

2 Material and methods

2.1 Training and test sites

In this study, we selected three regions within the Brazilian Amazon as study sites. These scenes encompass major deforestation centers that have developed along the "TransAmazon" (BR-230) [66,68] and "Cuiabá-Santarem" (BR-163) [69-71] highways (Figure 1). In the Amazon, roads are the driving forces for the spatial distribution of deforestation in the Amazon, where most deforestation occurs in the neighborhood of the main highway [72,73]. Widely discussed in the literature, the opening of roads in the Amazon forest favors the establishment of settlements, attracts migrants, facilitates the extraction of resources, increases the profitability of livestock and agriculture, and establishes access to wood [74-78].

The training used two scenes (Sites A and B), and validation utilized the remaining scene (Site C). We defined a bi-temporal approach for modelling and obtained Landsat 8/OLI imagery for each site for the years of 2017, 2018, and 2019, with approximately one year between each observation. Multitemporal images from the similar periods of the year reduce variations in the phenology and sun-terrain-sensor geometry. The images acquired were from the dry season to minimize cloud cover and reduce noise (Table 1). Tier 1 Landsat images were used as they offer consistent georregistration within prescribed image-to-image tolerances of less than 12-meter radial root mean square error (RMSE) and are therefore appropriate for time-series analysis (https://www.usgs.gov/land-resources/nli/landsat/landsat-collection-1).

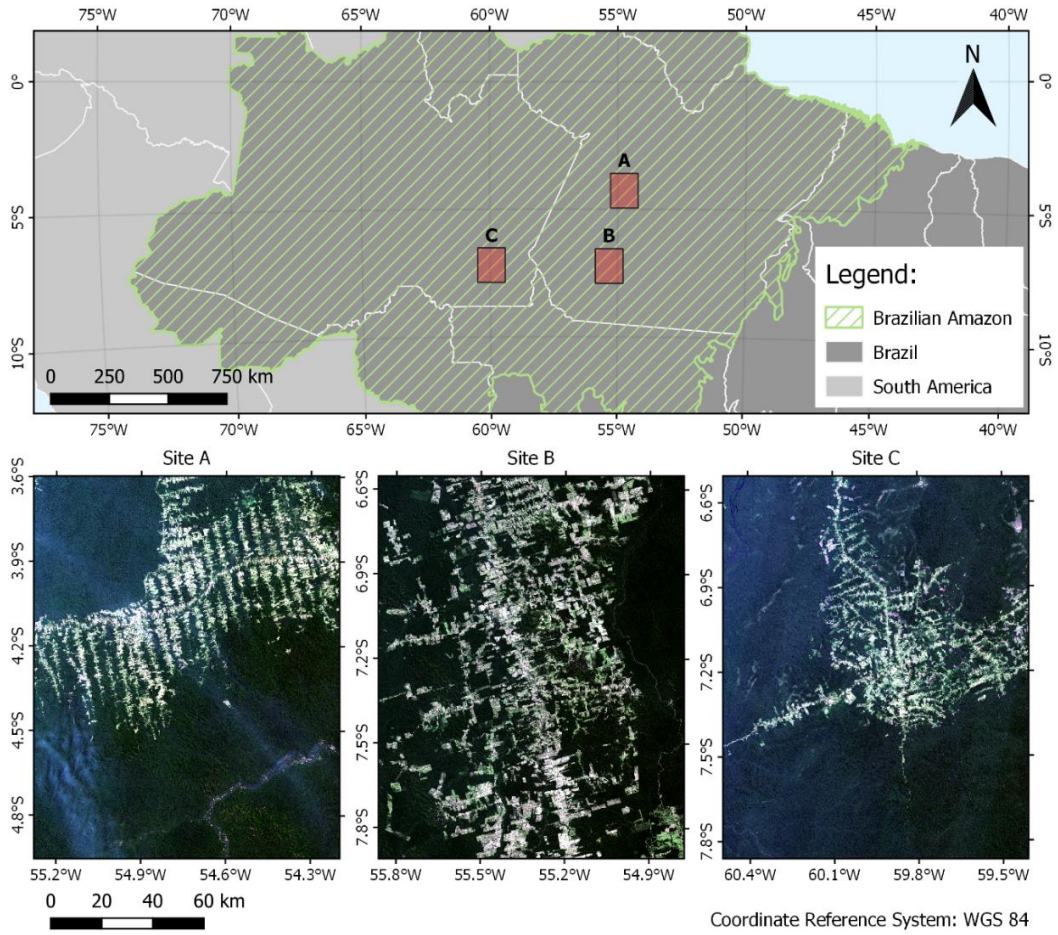

Figure 1. (a) Location of the study sites within the Amazon region with sites A, B and C as true color composite Landsat images taken from June (C) and July (A and B) 2018.

Table 1. Acquisition dates for each site and corresponding Landsat scenes.

Site	Landsat Scene	Acquisition date		
		2017	**2018**	**2019**
A	227_63	July 18	July 21	July 24
B	227_65	July 18	July 21	July 24
C	230_65	June 21	June 24	July 13

2.2 Deep learning models

This research used three different DL architectures available in the literature: U-Net [79], SharpMask [80], and ResUnet [81]. While the U-Net and SharpMask algorithms were not developed for classification with remote sensing data in mind, studies have found that they are not

only suitable but offer state-of-the-art results [82–84]. These algorithms share similarities, being based on architectures known as autoencoders with the addition of bridges or residual connections. Autoencoders downsample of the feature maps generated through convolutional filters while incrementally increasing their number to learn low-level features compactly, and then upsample them back to the original input shape for inference. This process can be further enhanced using connections bridging the downsampling and upsampling steps (Figure 2) to propagate information. These connections help speed up training and reduce the degradation of data by combining both low-level detail and high-level contextual information. Low-level spatial detail is essential for change detection and land cover classifications, and that is the main reason behind the choice of this specific type of architecture for this study.

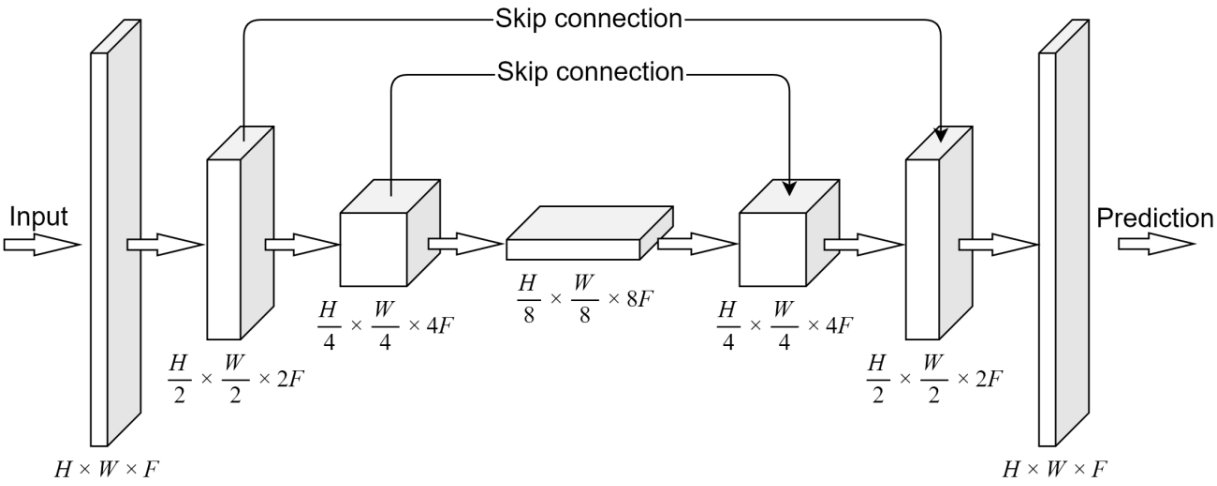

Figure 2. Simple representation of an autoencoder architecture with the addition of skip connections. H, W and F represent the height, width, and number of filtered feature maps respectively. In this study H and W are both 200 pixels, while F depends on the specific model architecture.

While similar in principle and structure, the chosen architectures differ in depth and complexity. Table 2 shows a summary of the number of layers and the total number of parameters in each model after adapting them for this study. Some of the inner workings of each model are also different. For example, the U-Net and SharpMask algorithms downsample the feature maps through a pooling operation, whereas the ResUnet architecture does downsampling by using a stride of 2 between convolutional filter windows. Another example is how the models use skip connections in different ways, where both U-Net and SharpMask use exclusively long connections

(linking the downsampling and upsampling sides of the architecture) while ResUnet makes use of long and short connections (between convolutional blocks).

Table 2. Total number of layers and parameters in each DL architecture used in this study.

Architecture	Layers	Parameters
U-Net	69	1,933,866
SharpMask	114	221,386
ResUnet	93	2,068,554

2.3 Data structure

The Landsat dataset consisted only of bands 1 through 7, as they share the same spatial resolution and contain most of the spectral information. Our initial training data was a bi-temporal cube stacking the base image and next year's image, constituting 14 bands. We maintained this data structure for the RF and MLP algorithms, where each pixel is an observation, and each band is a variable. The datasets had to be restructured for the DL algorithms due to the inner workings of CNNs and due to hardware memory constraints. To build and train the models in this study, we used the Keras [85] python library, a high-level wrapper for the well-known Tensorflow library [86]. When working with three-dimensional image data, Keras accepts inputs in the form of a four-dimensional array with shape (samples, sample rows, sample columns, channels). To convert our images to the correct format, we extracted patches through 200x200 pixel windows with a 10-pixel overlap on each side (Figure 3). This process generated a total of 844 samples per site per time sequence, with a total of 3376 training samples and 1688 test samples.

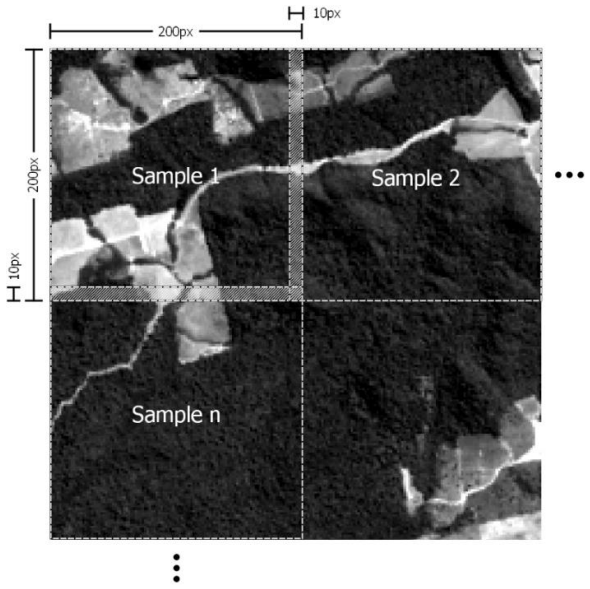

Figure 3. Example of the patch extraction method to prepare the datasets for the DL algorithms.

2.4 Ground truth

To create our ground truth masks, we used INPE's PRODES data (available at http://www.dpi.inpe.br/prodesdigital/dadosn/) for the years of 2018 and 2019 as a visual guide and then refined it by remapping the deforestation polygons on a smaller scale. PRODES data is commonly used for deforestation reports and studies have used it before when modelling and studying deforestation dynamics [87, 88]. The changes were mapped using digitizing tools from the QGIS software [89] at 1:30000 scale and subsequently transformed into binary raster files with 0 and 1 as absence–presence codes, respectively. In this process, we mapped changes exclusively to the natural forest, regardless of the land cover type in the following year (Figure 4).

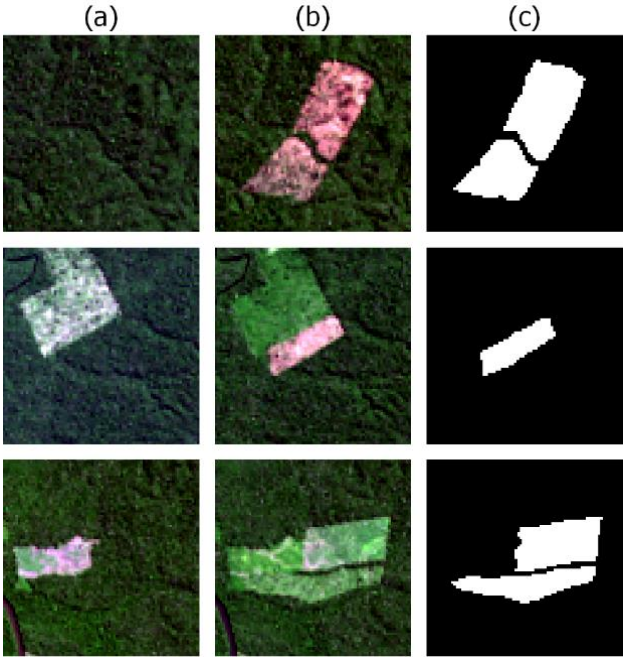

Figure 4. Example of the change mapping in three locations between (a) 2017 and (b) 2018 and the respective (c) rasterized deforestation mask.

2.5 Hyperparameters

The RF model only needed two hyperparameters set, the number of trees to build (ntree) and the number of variables randomly sampled as candidates at each split (mtry). Those were set to 500 trees, and 3 variables, respectively. The structure of the MLP algorithm consisted of a simple 3-layer network containing an input layer, a hidden layer with 256 nodes, and an output layer. The DL algorithms and MLP shared the same hyperparameters for training. Focal loss [84] was used as the loss function as it excels in classification problems with an uneven number of observations in each class, as is the case of our object of study. For gradient descent optimization, we used the Adaptive Moment Estimation (ADAM) algorithm [90] with incorporated Nesterov Momentum (NADAM) with a learning rate of 2e-3, β_1 of 0.9 and β_2 of 0.999. The number of epochs was set to 250 and the batch size to 16 to fit the training process into memory.

2.6 Modeling approach

Given the context of our main methodological steps described in the previous sections, a top-down view of our modeling approach is described in Figure 5. In addition to the DL algorithms, two classical ML algorithms — Random Forest (RF) and a simple Multilayer Perceptron (MLP)

architecture — were used as a reference point for the assessment of the DL models. Both models have been extensively researched for land cover classification, and change detection in remote sensing data with their performance is well documented [91,92].

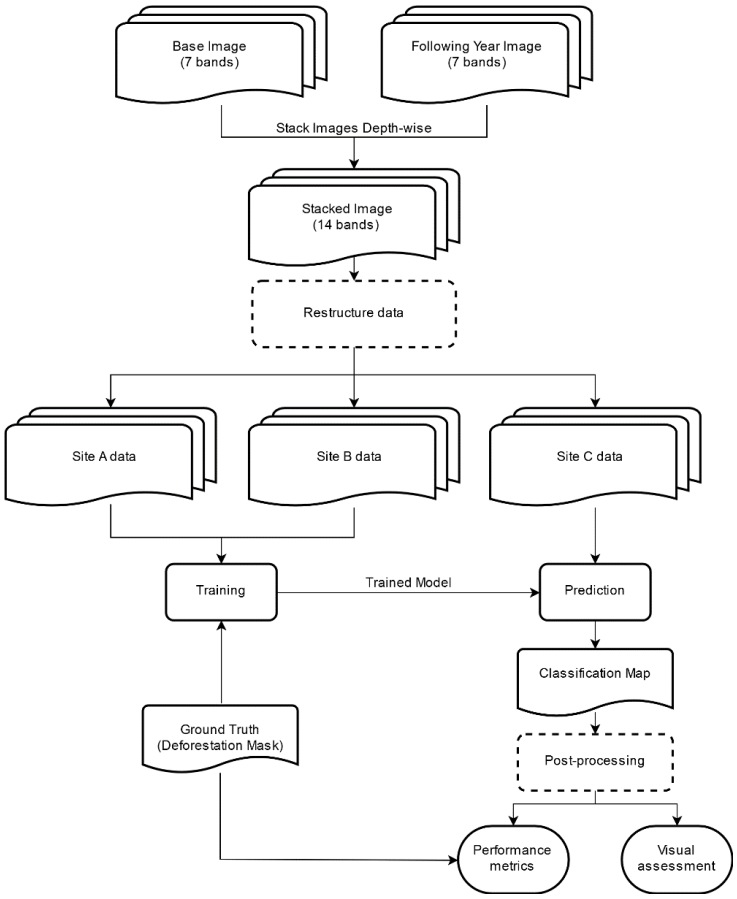

Figure 5. Flowchart of the modeling approach taken in this study.

2.7 Accuracy assessment

The accuracy metrics were calculated using exclusively the test site data to avoid the possibility of biased results due to overfitting. The classification results were compared to the ground truth mask for the test site in order to calculate the accuracy measures. Given that deforestation related change is typically a rare phenomenon, the change – no-change ratio is highly imbalanced [93, 94]. Therefore, change detection research usually shows a predominance of invariant areas, causing a bias in some accuracy metrics. For example, overall accuracy is relatively high on most change maps [95]. The Precision and Recall measures (Equations 1 and 2) were used to offer more insight in the distribution of errors in the classifications, along with three

other measures besides accuracy: F1 score (also known as Dice coefficient), Kappa index, and mean Intersection over Union (mIoU) measure (Equations 3,4 and 5 respectively). These measures are often used to evaluate DL and ML classifications and are better suited for classifications with imbalanced datasets than overall accuracy as they equally weight classe distributions.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

$$F1 = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \quad (3)$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

where p_o is the rate of agreement between the ground truth and the classification and p_e is the expected rate of agreement due to chance.

$$mIoU = \frac{IoU_1 + IoU_2 \cdots + IoU_n}{n} \quad (5)$$

where IoU is the area of intersection divided by the area of union between the classification and ground truth for a class and n is the total number of classes. Lastly, McNemar's test [96] was used to evaluate the statistical significance of differences between the classifications.

3 Results

Quantitatively the DL models showed a clear advantage over RF and MLP (Table 3). The ResUnet model had the best results in regards to every measure with the exception of Precision in the 2017–2018 time frame. The SharpMask and U-Net models, showed similar but slightly inferior results. In comparison, the RF model showed the worst results in most measures, although the performance measures still indicate a good classification. It should be noted that the RF and MLP classifications exhibited a considerable amount of impulse noise ("salt-and-pepper" type), and a majority filter was applied to reduce the noise and improve the classification both visually and quantitatively. The DL models did not require any post-processing steps as they produced classifications with virtually no noise. All models showed very high overall accuracy, but, as explained previously, this measure should be carefully considered as the ratio between the change

and no-change classes is highly imbalanced and it is mostly explained by the larger, no-change class. McNemar's test results indicate that despite the seemingly similar results, the model classifications were all significantly different from each other (Table 4).

Table 3. Performance measures for the model validation results for the 2017–2018 and 2018–2019 sequences. Best results in the column in bold text.

Model	2017 – 2018						2017 – 2019					
	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy
RF	0.8014	0.8003	0.8332	**0.9414**	0.6976	0.9979	0.8902	0.8892	0.9000	0.8877	0.8928	0.9979
MLP	0.8926	0.8920	0.9024	0.9282	0.8597	0.9987	0.9101	0.9093	0.9167	0.9314	0.8898	0.9983
Resunet	**0.9432**	**0.9428**	**0.9459**	0.9252	**0.9619**	**0.9993**	**0.9465**	**0.9460**	**0.9487**	**0.9358**	**0.9574**	**0.9990**
Unet	0.9112	0.9106	0.9179	0.9223	0.9003	0.9989	0.9339	0.9332	0.9373	0.9175	0.9508	0.9987
Sharpmask	0.9223	0.9218	0.9274	0.9173	0.9274	0.9990	0.9337	0.9331	0.9372	0.9218	0.9460	0.9987

Table 4. McNemar's test p-values between model classifications. Values bellow p=0.05 indicate the differences between classifications are statistically significant.

	2017 –2018					2018 –2019				
	MLP	**ResUnet**	**RF**	**SharpMask**	**U-Net**	**MLP**	**ResUnet**	**RF**	**SharpMask**	**U-Net**
MLP										
ResUnet	<0.001					<0.001				
RF	<0.001	<0.001				<0.001	<0.001			
SharpMask	<0.001	<0.001	<0.001			<0.001	<0.001	<0.001		
U-Net	<0.001	<0.001	<0.001	<0.001		<0.001	<0.001	<0.001	<0.001	

RF's higher precision in the 2017–2018 frame can be explained by the low amount of false positives produced. Conversely, however, it produced a very high number of false negatives within the same frame (Figure 10). The ResUnet model had the lowest number of misclassified pixels in both time sequences. It also produced the least number of false negatives out of all the models. When looking at the number of false-positive cases, the DL algorithms did not show a large difference over the ML models. In regards to false negatives, however, they showed a clear advantage. The reduction of false-negative classifications is a considerable advantage of the DL models over the classic ML algorithms, given that underestimating the extent of deforestation is a less desirable outcome than its overestimation.

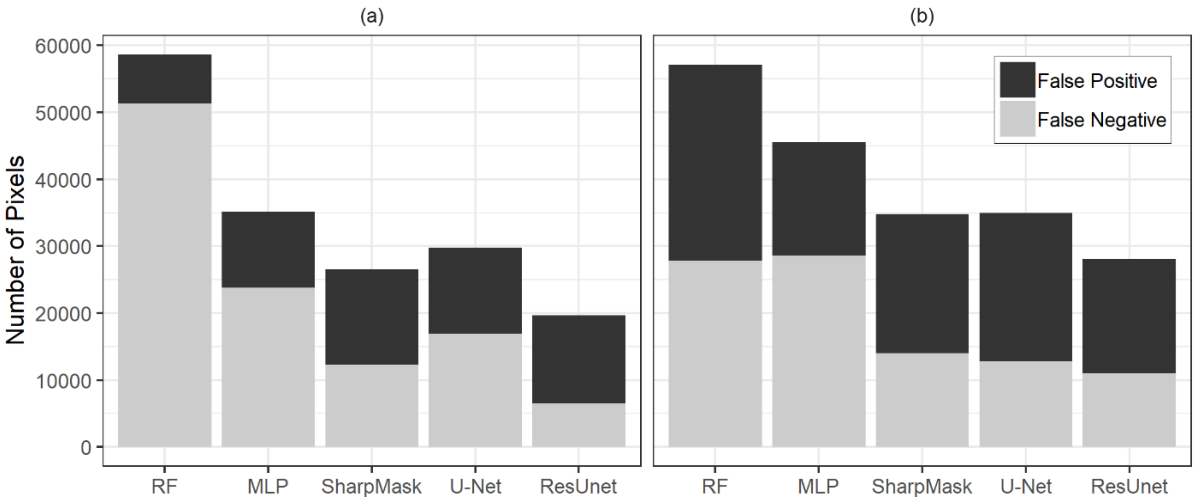

Figure 6. Error distributions in the (a) 2017–2018 and (b) 2018–2019 time sequences in total pixel numbers.

The models detected roughly the same deforestation sites at the validation site across both time sequences (Figures 6 and 7). However, the DL models provided more detailed classifications within smaller scales, particularly around feature edges. Moreover, all models were able to classify "easy" deforestation patches with less complex spectral mixtures (Figure 8), but the classification of the ML algorithms degraded as the spectral signatures within the patches increased in complexity (Figure 9). RF showed a higher tendency to produce false negatives both visually and quantitatively.

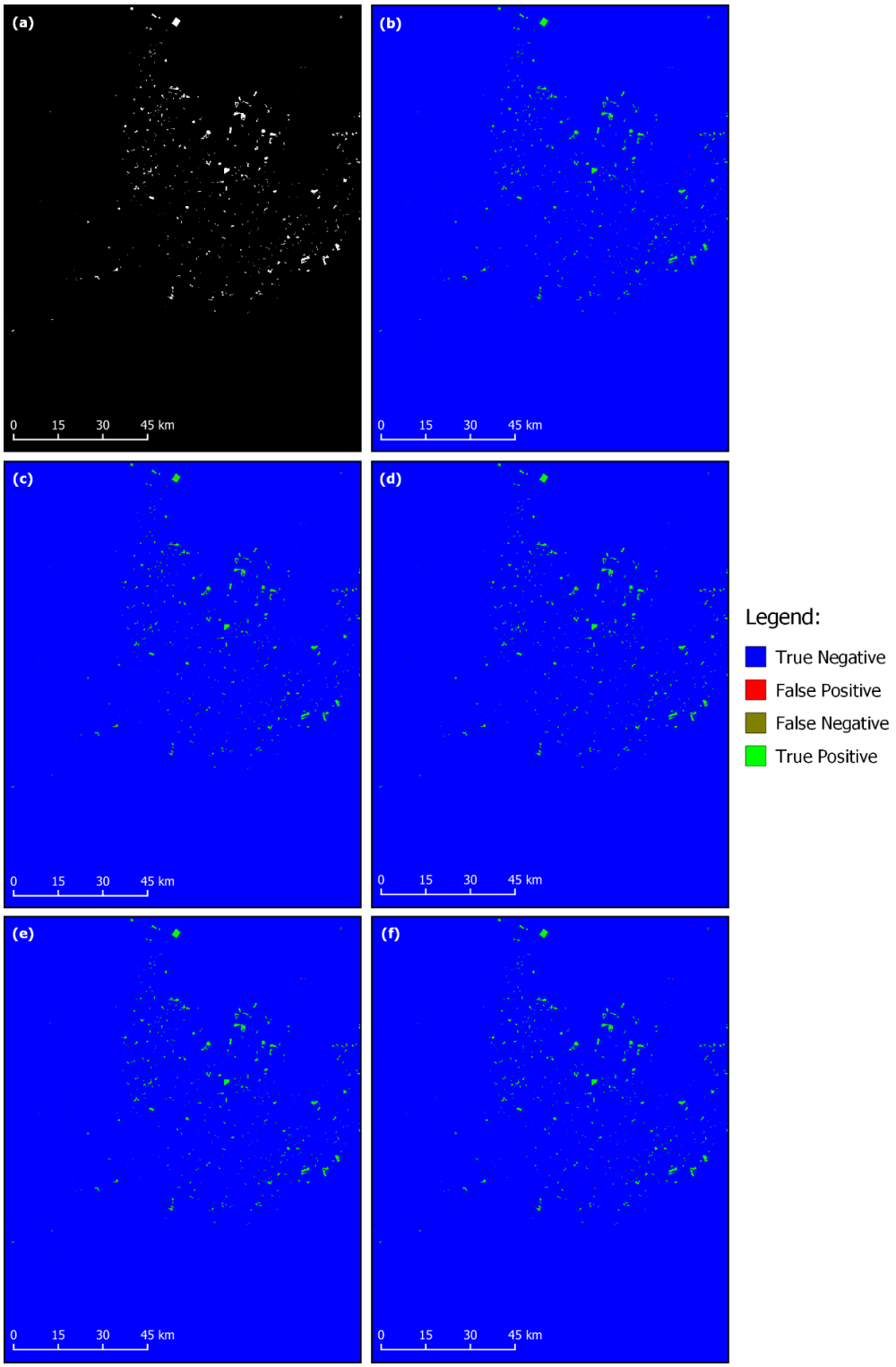

Figure 7. Deforestation masks according to the (**a**) ground truth and classifications produced by the (**b**) RF, (**c**) MLP, (**d**) SharpMask, (**e**) U-Net and (**f**) ResUnet models in the 2017–2018 sequence.

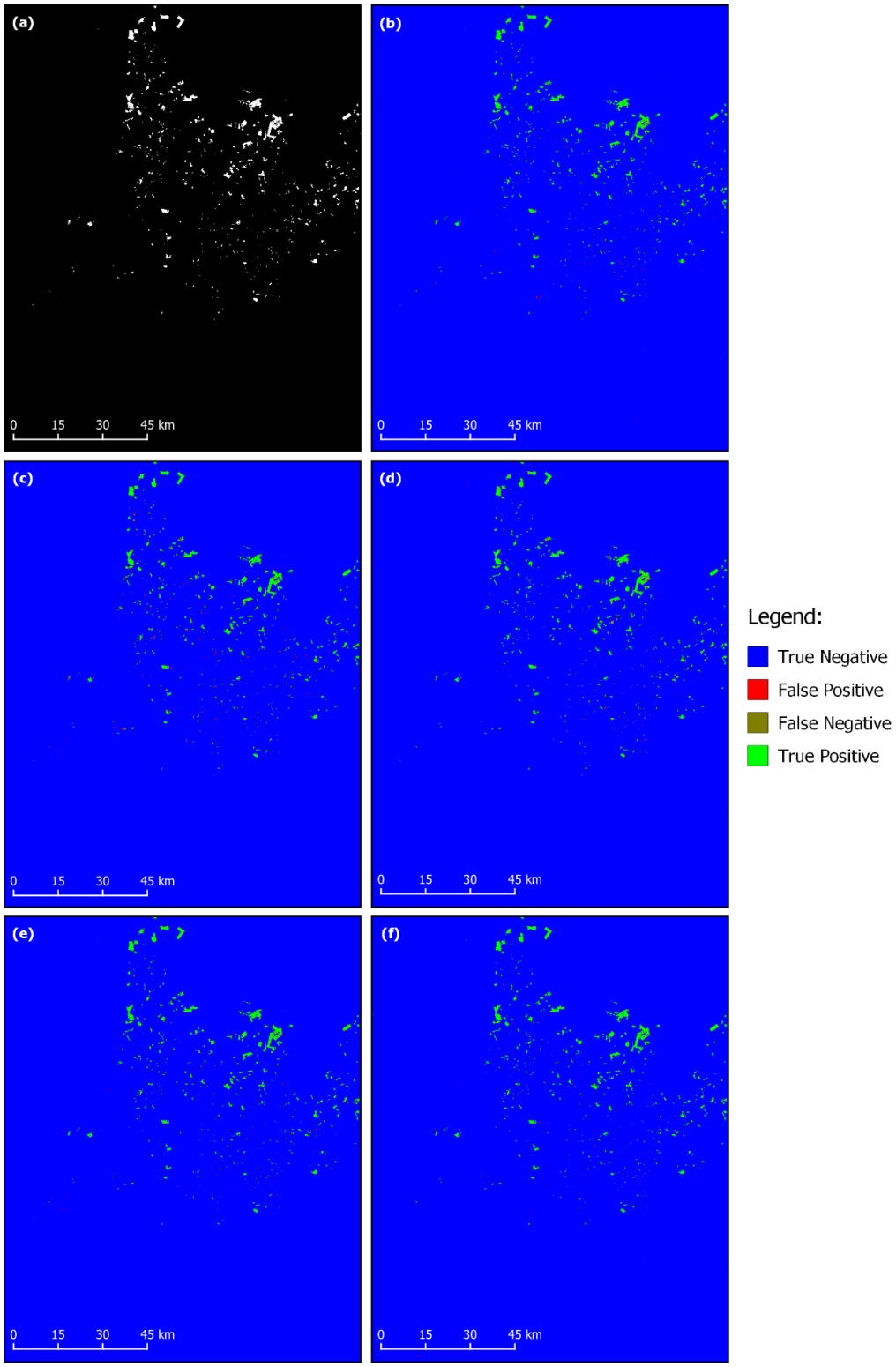

Figure 8. Deforestation masks according to the (**a**) ground truth and classifications produced by the (**b**) RF, (**c**) MLP, (**d**) SharpMask, (**e**) U-Net and (**f**) ResUnet models in the 2018–2019 sequence.

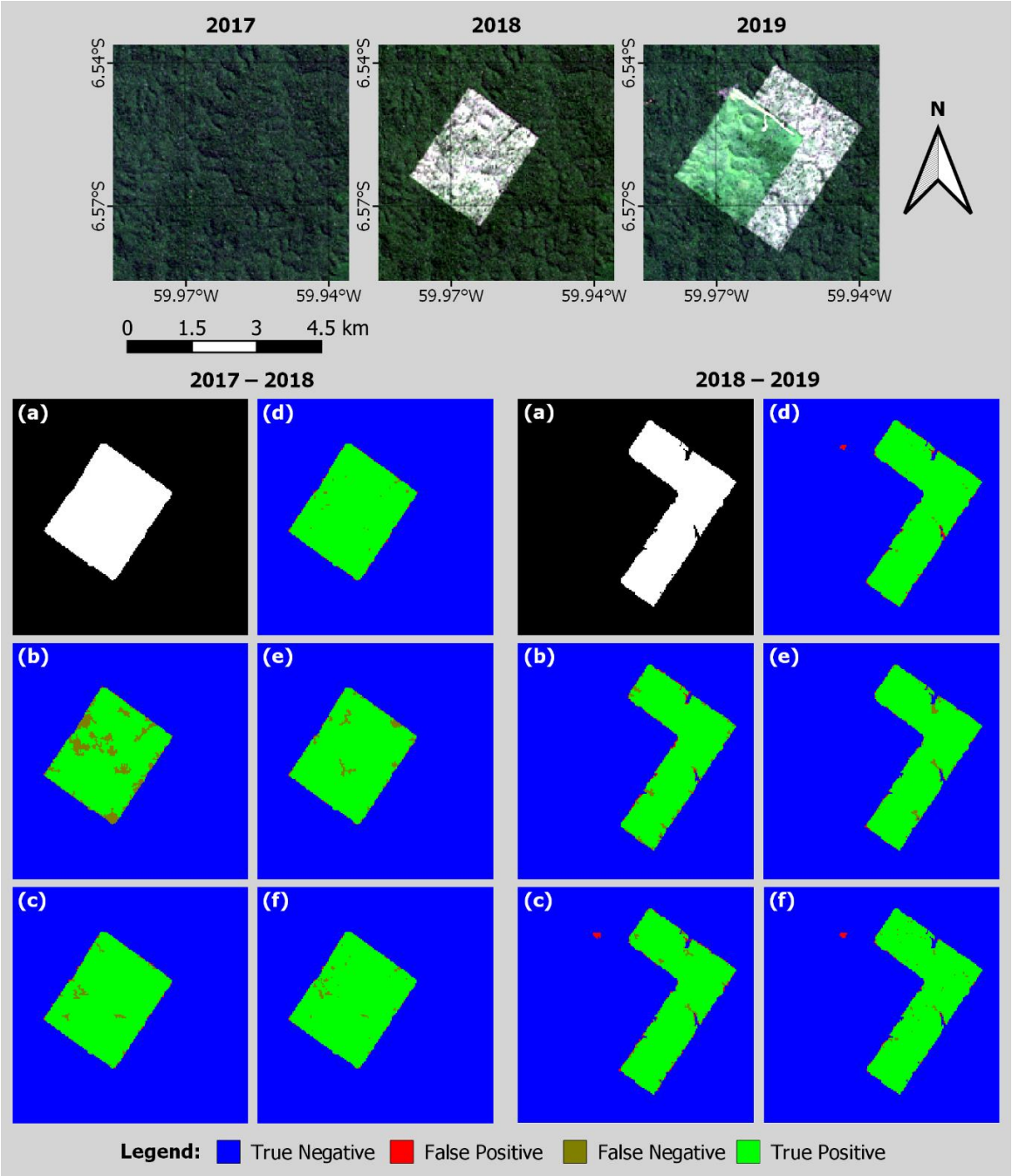

Figure 9. First example location within the test site with the (**a**) ground truth and classifications made by the (**b**) RF, (**c**) MLP, (**d**) SharpMask, (**e**) U-Net and (**f**) ResUnet models in each time sequence.

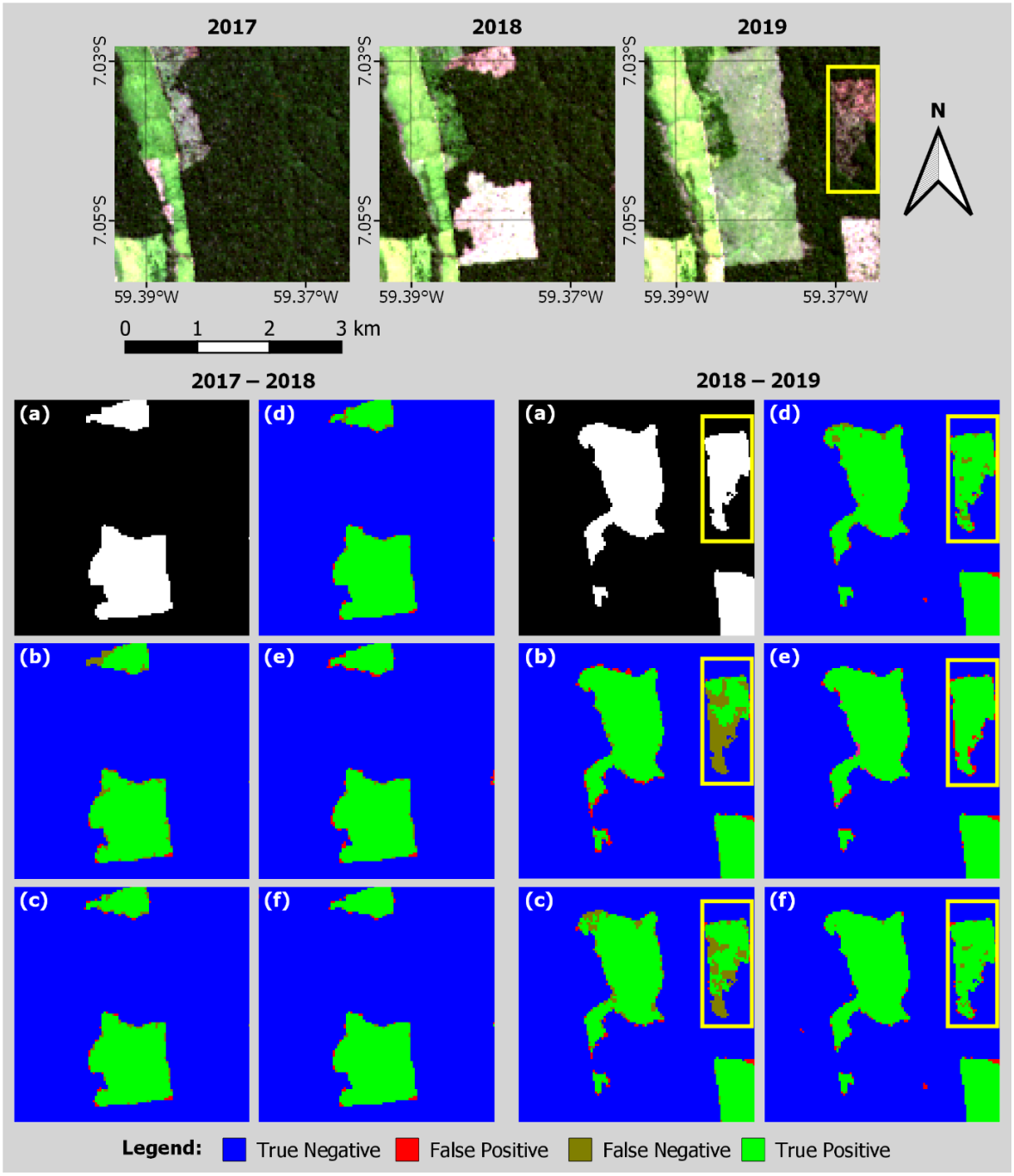

Figure 10. Second example location within the test site with the (a) ground truth and classifications made by the (b) RF, (c) MLP, (d) SharpMask, (e) U-Net and (f) ResUnet models in each time sequence. The yellow rectangle highlights an example of a "hard-to-classify" deforestation patch.

The total deforested area was slightly higher than the ground truth in the SharpMask and ResUnet predictions in both time sequences (Table 5). The opposite was true for the MLP

prediction, which slightly underestimated the total area in both time spans. The RF model underestimated the total deforested area by a very large portion (almost 40km² or a 26% decrease in area) in the 2017–2018 sequence due to a large amount of false negative predictions, but despite that it came closest to the ground truth area in the 2018–2019 sequence, although, that does not necessarily mean the predicted areas were the same as ground truth.

Table 5. Total deforested area according to the ground truth and each model's prediction.

Reference	Deforested Area (km²)		Difference from Ground Truth (%)	
	2017 – 2018	2018 – 2019	2017 – 2018	2018 – 2019
Ground Truth	152.73	233.44	---	---
RF	113.17	234.79	-25.90	+0.58
MLP	141.45	223.01	-7.39	-4.47
SharpMask	154.40	239.56	+1.10	+2.62
U-Net	149.10	241.90	-2.38	+3.62
ResUnet	158.78	238.84	+3.96	+2.31

Processing times varied from model to model but MLP and the DL models offered faster training and prediction times than RF, mainly due to the fact that the Tensorflow framework uses the computer's Graphical Processing Unit (GPU) for parallel processing instead of the Central Processing Unit (CPU) which is traditionally used for ML. Using an NVIDIA GTX 1070 GPU and a batch size of 16, the total training time ranged from approximately 40 minutes for the simpler MLP model (around 10 seconds per epoch) to almost 3 hours for the more complex ResUnet model (approximately 40 seconds per epoch). Given the size of the datasets, RF took approximately 6 hours to train using parallel processing with an Intel Core i5-4690k processor. The difference in processing times was particularly considerable when using the models to classify the images after training. The DL models and MLP classified the test scene within seconds, whereas RF took almost an hour to complete the task.

4 Discussion

The CNN architectures used in this study showed a clear advantage to the classic ML algorithms, both quantitatively and visually, regarding deforestation mapping. Similarly, a comparative study of methods developed by [97] for wetland mapping found that deep learning

methods (Completely Convolutional Networks and Patch-based Deep CNN) obtained better accuracy than RF and Support Vector Machine. The authors found that CNN may produce inferior performance when the training sample size is small, but it tends to show substantially higher accuracy than conventional classifiers using a larger training sample size. We assume the difference in performance between the DL and traditional ML methods stems from the former's capability to understand both the spatial and spectral context, whereas the regular ML models inherently only see the spectral information.

Although the current methodologies to detect deforestation with DL architectures vary widely, studies are in agreement that they produce excellent classification results [98–100]. Other analogous studies that investigated the use of DL for single class classifications seem to corroborate with this trend, although there is a large variation between the choice of targets and architectures [34,101–103]. While choice and development of architectures for certain targets is a relevant topic for future research, we have found that autoencoder networks with residual connections seem to be a good starting point for classifications in remote sensing imagery as they can take advantage of spatial and spectral information in a very efficient manner.

Despite their advantages, DL algorithms are still not as accessible or as easy to use as classic ML models. Besides needing specific hardware for training, they require a relatively large quantity of samples, and developing ground truth masks for specific targets can be challenging and time-consuming in large extents as both spatial and spectral context is strictly needed, whereas the traditional ML algorithms work with simpler sampling schemes and can produce reasonably good results with a much smaller sample size. Therefore, the process of building a model for broader use (i.e., country-wide monitoring) can be complicated. However, these models have another advantage in the fact that they can be incrementally trained, meaning they could be gradually provided with new samples to update the model weights and improve their classifications with time. With that said, the "black box" nature of these networks can make them undesirable for those who might wish to know and disclose their internal workings such as public and governmental entities. Despite that, through our findings, we believe that with enough development, DL algorithms can provide a viable automatic solution for mapping deforestation in the Amazon alongside projects such as INPE's PRODES and TerraClass.

It should be noted that while the models showed good capability for generalization within our region of study, we cannot assert that they would achieve the same results in different areas in

which deforestation is a common occurrence. A broader reaching model would necessarily need samples from different regions to account for the possible spatial and spectral variability from one region to another. Further research should be carried to study the applicability of the models for similar targets in different areas. In addition, while Landsat data is enough for annual deforestation mapping between dry seasons, more frequent monitoring is virtually impossible as clouds are present above the forest canopy during most of the year, and the ground reflectance cannot reach the satellite's optical sensors. One solution would be the use of radar data to be able to cross the cloud cover. As such, we also recommend the investigation of the use of radar data and DL algorithms to detect deforestation within a shorter time-frame.

5 Conclusions

In this study, we proposed the use of existing DL architectures to detect yearly changes in the vegetation cover in the region of the Brazilian Amazon, successfully achieving our goal. Results show that these algorithms are a viable alternative to classical ML algorithms, with the improvement of all performance measures and clear advantages, such as faster prediction times and lack of noise in the classifications. The SharpMask, U-Net, and ResUnet models showed similar results. However, ResUnet achieved the best values of accuracy, Kappa, F1, and mIoU, and the least amount of errors overall. Visually the DL algorithms also produced classification masks with well-defined deforestation patches, while the ML models showed an evident loss of quality in harder to classify patches, with a tendency to produce false negatives and salt-and-pepper noise that needed to be filtered. One of the main shortcomings of CNNs seems to be the necessity of a 1:1 ground truth regarding the extent of the area of study as the spatial context is critical. In contrast, simpler ML models can be trained on a point-by-point basis (e.g., random sampling points within the extent). Developing a whole ground truth can be an extensive process. However, we have achieved very good results with a relatively small sample size with very little augmentation in the form of overlapping sample patches. The additional bands in remote sensing data may facilitate the detection of targets with less samples but that supposition needs further research. Furthermore, considering the models were validated by being applied to an independent dataset, the performance measures show that they have very good potential for generalization. DL is still a growing technology, particularly in the remote sensing field as not even popular libraries such as Keras and Tensorflow have built-in tools for dealing with multi-band satellite imagery, but researchers are slowly adapting and developing better architectures specific for remote sensing

data. The architectures used in this study performed well in our specific task, although they were developed for entirely different targets. Therefore, these algorithms do not necessarily need to be tailored for specific cases and can even work interchangeably between fields of research.

References

1. Le Quéré, C.; Andrew, R.M.; Friedlingstein, P.; Sitch, S.; Hauck, J.; Pongratz, J.; Pickers, P.A.; Korsbakken, J.I.; Peters, G.P.; Canadell, J.G.; et al. Global carbon budget 2018. *Earth Syst. Sci. Data* **2018**, *10*, 2141–2194.
2. Aragão, L.E.O.C.; Poulter, B.; Barlow, J.B.; Anderson, L.O.; Malhi, Y.; Saatchi, S.; Phillips, O.L.; Gloor, E. Environmental change and the carbon balance of Amazonian forests: Environmental change in Amazonia. *Biol Rev* **2014**, *89*, 913–931.
3. Rosa, I.M.D.; Smith, M.J.; Wearn, O.R.; Purves, D.; Ewers, R.M. The environmental legacy of modern tropical deforestation. *Current Biology* **2016**, *26*, 2161–2166.
4. Vedovato, L.B.; Fonseca, M.G.; Arai, E.; Anderson, L.O.; Aragão, L.E.O.C. The extent of 2014 forest fragmentation in the Brazilian Amazon. *Reg Environ Change* **2016**, *16*, 2485–2490.
5. Spracklen, D.V.; Garcia-Carreras, L. The impact of Amazonian deforestation on Amazon basin rainfall: AMAZONIAN DEFORESTATION AND RAINFALL. *Geophys. Res. Lett.* **2015**, *42*, 9546–9552.
6. Boisier, J.P.; Ciais, P.; Ducharne, A.; Guimberteau, M. Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Clim Change* **2015**, *5*, 656–660.
7. INPE Projeto PRODES: Monitoramento da Floresta Amazônica Brasileira por satélite Available online: http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes (accessed on Oct 7, 2019).
8. INPE Projeto TerraClass Available online: http://www.inpe.br/cra/projetos_pesquisas/dados_terraclass.php (accessed on Oct 7, 2019).
9. Pearson, T.R.H.; Brown, S.; Murray, L.; Sidman, G. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance Manage* **2017**, *12*, 3.
10. Singh, A. Review article digital change detection techniques using remotely-sensed data. . *Int. J. Remote Sens*. **1989**, *10*, 6, 989-1003.

11. Guo, E.; Fu, X.; Zhu, J.; Deng, M.; Liu, Y.; Zhu, Q.; Li, H. Learning to Measure Change: Fully Convolutional Siamese Metric Networks for Scene Change Detection. arXiv **2018**, arXiv:1810.09111.
12. Coppin, P.; Jonckheere, I.; Nackaerts, K.; Muys, B.; Lambin, E. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.* **2004**, *25*, 9, 1565-1596.
13. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*,12, 2365-2407.
14. Radke, R.J.; Andra, S.; Al-Kofahi, O.; Roysam, B. Image Change Detection Algorithms: A Systematic Survey. *IEEE Trans. Image Process*. **2005**, *14*, 3, 294-307.
15. Warner, T.; Almutairi, A.; Lee, J.Y. Remote sensing of land cover change. In *The SAGE handbook of remote sensing;* Warner, T.A., Nellis, D.M., Foody G.M., Eds.; SAGE Publications: London, UK, **2009**; pp. 459–472.
16. Hecheltjen A.; Thonfeld F.; Menz G. Recent Advances in Remote Sensing Change Detection – A Review. In: *Land Use and Land Cover Mapping in Europe*; Manakos, I., Braun, M., Eds.; Springer: Dordrecht, The Netherlands, **2014**. Volume 18, pp. 145–178.
17. Zhu, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens*. **2017**, *130*, 370-384.
18. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ*. **2015**, *160*, 1-14.
19 Ghosh, S.; Roy, M.; Ghosh, A. Semi-supervised change detection using modified self-organizing feature map neural network. *Appl. Soft Comput.* 2014, **15**, 1–20.
20 Schneider, A. Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sens. Environ*. **2012**, *124*, 689-704.
21. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosc. Remote Sens. M.* **2016**, *4*, 2, 22-40.
22. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens*. **2019**, *152*, 166-177.

23. Hughes, L.; Schmitt, M.; Zhu, X. Mining hard negative samples for SAR-optical image matching using generative adversarial networks. *Remote Sens*. 2018, 10, 10, 1552.
24. Ma, W.; Zhang, J.; Wu, Y.; Jiao, L.; Zhu, H.; Zhao, W. A Novel Two-Step Registration Method for Remote Sensing Images Based on Deep and Local Features. *IEEE Trans. Geosci. Remote Sens*. 2, 2019.
25. Merkle, N.; Auer, S.; Müller, R.; Reinartz, P. Exploring the potential of conditional adversarial networks for optical and SAR image matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 2018,*11*(6), 1811–1820.
26. Wang, S.; Quan, D.; Liang, X.; Ning, M.; Guo, Y.; Jiao, L. A deep learning framework for remote sensing image registration. *ISPRS J. Photogramm. Remote Sens*. 2018. 145, 148–164.
27. Carranza-García, M.; García-Gutiérrez, J.; Riquelme, J. A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks. *Remote Sens*. **2019**, *11*, 274.
28. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782.
29. Li, M.; Wang, L.; Wang, J.; Li, X.; She, J. Comparison of land use classification based on convolutional neural network. *J. Appl. Remote Sens.* **2020**, *14*, 1, 016501.
30. Scott, G.J.; England, M.R.; Starms, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553.
31. Chen, F.; Ren, R.; Van de Voorde, T.; Xu, W.; Zhou, G.; Zhou, Y. Fast Automatic Airport Detection in Remote Sensing Images Using Convolutional Neural Networks. *Remote Sens.* **2018**, *10*, 443.
32. Kang, M; Ji, K; Leng, X; Lin, Z Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection. Remote Sensing **2017**, 9, 860.
33. Qian, X.; Lin, S.; Cheng, G.; Yao, X.; Ren, H.; Wang, W. Object Detection in Remote Sensing Images Based on Improved Bounding Box Regression and Multi-Level Features Fusion. *Remote Sens.* **2020**,*12*, 143.
34. Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *Int. J. Comput. Int. Sys.* **2017**, *16*, 1750001.

35. Liu, X.; Liu, Q.; Wang, Y. Remote sensing image fusion based on two-stream fusion network. *Inf. Fusion*, **2020**, *55*, 1-15.
36. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018.** *42*, 158–173.
37. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sens*. **2018**, 1–15.
38. Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-Sharpening. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. **2018**, *11*, 978–989.
39. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogramm. Remote Sens*. **2018.** *145*, 60–77.
40. Malambo, L.; Popescu, S.; Ku, N.-W.; Rooney, W.; Zhou, T.; Moore, S. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sens.* **2019**, *11*, 2939.
41. Xiao, X.; Zhou, Z.; Wang, B.; Li, L.; Miao, L. Ship Detection under Complex Backgrounds Based on Accurate Rotated Anchor Boxes from Paired Semantic Segmentation. *Remote Sens.* **2019**, *11*, 2506.
42. Zhuo, X.; Fraundorfer, F.; Kurz, F.; Reinartz, P. Optimization of openstreetmap building footprints based on semantic information of oblique UAV images. *Remote Sens*. **2018**. *10*, 4, 624.
43. Xing, H.; Meng, Y.; Wang, Z.; Fan, K.; Hou, D. Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS J. Photogramm. Remote Sens*. **2018**. *141*, 237–251.
44. Khan, S.H.; He, X.; Porikli, F.; Bennamoun, M. Forest change detection in incomplete satellite images with deep neural networks. *IEEE Trans. Geosci. Remote Sens*. **2017**, 55, 5407–5423.
45. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 924–935.
46. Ajami, A.; Ku_er, M.; Persello, C.; Pfeffer, K. Identifying a slums' degree of deprivation from VHR images using convolutional neural networks. *Remote Sens*. **2019**, 11, 1282.

47. Cao, G.; Li, Y.; Liu, Y.; Shang, Y. Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification. *Int. J. Remote Sens*. **2014**, 35, 6255–6270.
48. Mboga, N.; Persello, C.; Bergado, J.R.; Stein, A. Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sens*. **2017**, *9*, 1106.
49. Liu, R.; Kuffer, M.; Persello, C. The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach. *Remote Sens.* **2019**, *11*, 2844.
50. Cao, C.; Dragićević, S.; Li, S. Land-use change detection with convolutional neural network methods. *Environments* **2019**, *6*, 25.
51. Zhang, X.; Shi, W.; Lv, Z.; Peng, F. Land cover change detection from high-resolution remote sensing imagery using multitemporal deep feature collaborative learning and a semi-supervised chan–vese model. *Remote Sens.* **2019**, *11*, 2787.
52. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70.
53. Liu, Y.; Wu, L. Geological disaster recognition on optical remote sensing images using deep learning. *Procedia Comput. Sci.* **2016**, *91*, 566-575.
54. Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sens.* **2019**, *11*, 1382.
55. Hou, B.; Wang, Y.; Liu, Q. Change Detection Based on Deep Features and Low Rank. *IEEE Geosci. Remote Sens. Lett*. **2017**, 14, 2418–2422.
56. Niu, X.; Gong, M.; Zhan, T.; Yang, Y. A Conditional Adversarial Network for Change Detection in Heterogeneous Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 45–49.
57. Zhang, M.; Xu, G.; Chen, K.; Yan, M.; Sun, X. Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. *IEEE Geosci. Remote Sens. Lett*. **2019**, *16*, 266–270.
58. Gong, M.; Zhan, T.; Zhang, P.; Miao, Q. Superpixel-based difference representation learning for change detection in multispectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2658–2673.

59. Ma, W.; Xiong, Y.; Wu, Y.; Yang, H.; Zhang, X.; Jiao, L. Change Detection in Remote Sensing Images Based on Image Mapping and a Deep Capsule Network. *Remote Sens.* **2019**, *11*, 626.
60. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyper- spectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 3–13.
61. Zhang, W.; Lu, X. The Spectral-Spatial Joint Learning for Change Detection in Multispectral Imagery. *Remote Sens*. **2019**, *11*, 240.
62. Lebedev, M.; Vizilter, Y.V.; Vygolov, O.; Knyaz, V.; Rubis, A.Y. Change detection in remote sensing images using conditional adversarial networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*. **2018**, *42*, 565–571.
63. Lei, T.; Zhang, Y.; Lv, Z.; Li, S.; Liu, S.; Nandi, A.K. Landslide Inventory Mapping from Bi-temporal Images Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 982–986.
64. Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. High Resolution Semantic Change Detection. arXiv **2018**, arXiv:1810.08452v1.
65. Arima, E.Y.; Walker, R.T.; Perz, S.; Souza, C. Explaining the fragmentation in the Brazilian Amazonian forest. *J. Land Use Sci.* **2015**, 1–21.
66. Godar, J.; Tizado, E.J.; Pokorny, B. Who is responsible for deforestation in the Amazon? A spatially explicit analysis along the Transamazon Highway in Brazil. *Forest Ecol. Manag.* **2012** *267*, 58-73.
67. Carrero, G.C.; Fearnside, P.M. Forest clearing dynamics and the expansion of landholdings in Apuí, a deforestation hotspot on Brazil's Transamazon Highway. *Ecol. Soc*., **2011**, *16*, 2, 26.
68. Li, G.; Lu, D.; Moran, E.; Calvi, M.F.; Dutra, L.V.; Batistella, M. Examining deforestation and agropasture dynamics along the Brazilian TransAmazon Highway using multitemporal Landsat imagery. *Gisci. Remote Sens.* **2019**, *56*, 2, 161-183.
69. Cerqueira, G.; Diaz, M.D.C.V.; Rivero, S. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: The Santarem–Cuiaba corridor. *Global Change Biol.* 2003, *10*, 745-764.

70. Müller, H.; Griffiths, P.; Hostert, P. Long-term deforestation dynamics in the Brazilian Amazon—Uncovering historic frontier development along the Cuiabá–Santarém highway. *Int. J. Appl. Earth Obs.* **2016**, *44*, 61-69.
71. Soares-Filho, B.; Alencar, A.; Nepstad, D.; Cerqueira, G.; Vera Diaz, M.D.C.; Rivero, S.; Solórzano, L.; Voll, E. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarém–Cuiabá corridor. *Global Change Biol*. **2004**, *10*, 5, 745-764.
72. Barber, C.P.; Cochrane, M.A.; Souza Jr, C.M.; Laurance, W.F. Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biol. Conserv.* **2014**, 177, 203-209.
73. Fearnside, P.M. Highway construction as a force in destruction of the Amazon forest. In *Handbook of road ecology*; van der Ree, R., Smith, D.J., Grilo C.; Eds.; John Wiley & Sons Publishers: Oxford, UK, 2015; pp. 414-424.
74. Alves, D.S. Space-time dynamics of deforestation in Brazilian Amazônia. *Int. J. Remote Sens.* **2002**, *23*, 2903–2908.
75. Arima, E.; Walker, R.T.; Perz, S.G.; Caldas, M. Loggers and forest fragmentation: behavioral models of road building in the Amazon basin. *Ann. Assoc. Am. Geogr.* **2005**, *95*, 525–541.
76. Arima, E.Y.; Walker, R.T.; Sales, M.; Souza Jr, C.; Perz, S.G. The fragmentation of space in the Amazon basin: emergent road networks. *Photogramm. Eng. Rem. Sens.* **2008**. *74*, 699–709.
77. Asner, G.P.; Broadbent, E.N.; Oliveira, P.J.C.; Keller, M.; Knapp, D.E.; Silva, J.N.M.; Condition and fate of logged forests in the Brazilian Amazon. *Proc. Natl. Acad. Sci. U.S.A.* **2006.** *103*, 12947–12950.
78. Pfaff, A.; Robalino, J.; Walker, R.; Aldrich, S.; Caldas, M.; Reis, E.; Perz, S.; Bohrer C.; Arima, E.; Laurance, W.; Kirby, K. Road investments, spatial spillovers, and deforestation in the Brazilian Amazon. *J. Reg. Sci.* **2007**, 47: 109–123.
79. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs] 2015.
80. Pinheiro, P.O.; Lin, T.-Y.; Collobert, R.; Dollàr, P. Learning to Refine Object Segments. arXiv:1603.08695 [cs] 2016.
81. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote Sensing Lett. 2018, 15, 749–753.

82. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 2018, 145, 60–77.
83. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. Remote Sensing 2019, 11, 68.
84. Li, L.; Liang, J.; Weng, M.; Zhu, H. A Multiple-Feature Reuse Network to Extract Buildings from Remote Sensing Imagery. Remote Sensing 2018, 10, 1350.
85. Chollet, F. *Keras*; https://github.com/fchollet/keras, 2015.
86. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. arXiv **2016**, arXiv:1603.04467
87. Shimabukuro, Y.E.; Arai, E.; Duarte, V.; Jorge, A.; Santos, E.G. dos; Gasparini, K.A.C.; Dutra, A.C. Monitoring deforestation and forest degradation using multi-temporal fraction images derived from Landsat sensor data in the Brazilian Amazon. International Journal of Remote Sensing **2019**, 40, 5475–5496
88. Cabral, A.I.R.; Saito, C.; Pereira, H.; Laques, A.E. Deforestation pattern dynamics in protected areas of the Brazilian Legal Amazon using remote sensing data. Applied Geography **2018**, 100, 101–115.
89. QGIS Development Team QGIS Geographic Information System. Open Source Geospatial Foundation Project.; Open Source Geospatial Foundation, **2020**;
90. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In *Proceedings of the 2017 IEEE International Conference on Computer Vision* (ICCV); IEEE: Venice, Italy, October **2017**; pp. 2999–3007.
91. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv **2014**, arXiv:1412.6980.
92. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens*. **2016**, *114*, 24–31.
93. Mahmon, N.A.; Ya'acob, N. A review on classification of satellite image using Artificial Neural Network (ANN). In *Proceedings of the 2014 IEEE 5th Control and System Graduate Research Colloquium*; IEEE: Shah Alam, Malaysia, **2014**; pp. 153–157.

94. Stehman, S.V. Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing* **2009**, 30, 5243–5272.
95. Foody, G.M. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering and Remote Sensing* **2004**, 70, 627–633.
96. Stehman, S. Comparing estimators of gross change derived from complete coverage mapping versus statistical sampling of remotely sensed data. *Remote Sensing of Environment* **2005**, 96, 466–474.
97. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, 12, 153–157.
98. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *Gisci. Remote Sens.* **2018**, *55*, 2, 243-264.
99. Rakshit, S.; Debnath, S.; Mondal, D. Identifying Land Patterns from Satellite Imagery in Amazon Rainforest using Deep Learning. *arXiv:1809.00340 [cs]* **2018**.
100. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *arXiv:1709.00029 [cs]* **2019**.
101. Ortega, M.X.; Bermudez, J.D.; Happ, P.N.; Gomes, A.; Feitosa, R.Q. Evaluation of Deep Learning Techniques for Deforestation Detection the Amazon Forest. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *IV-2/W7*, 121–128.
102. Liu, C.-C.; Zhang, Y.-C.; Chen, P.-Y.; Lai, C.-C.; Chen, Y.-H.; Cheng, J.-H.; Ko, M.-H. Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation. *Remote Sens.* **2019**, *11*, 119.
103. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.; Tiede, D.; Aryal, J. Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sens.* **2019**, *11*, 196.

Capítulo IV

Irrigated Rice Crop Identification in Southern Brazil Using Convolutional Neural Networks and Sentinel-1 Time Series

Resumo: O arroz é uma das fontes básicas de alimentação, com milhões de toneladas produzidas e consumidas todos os anos. Portanto, o mapeamento de plantios de arroz é essencial para o processo do seu manejo agricultural. Esse estudo teve como objetivo a classificação de plantios de arroz no sul do Brasil utilizando séries temporais de radar SENTINEL-1 e modelos de Deep Learning (DL), comparando duas arquiteturas (U-Net e LinkNet) e quatro estruturas básicas de construção dessas arquiteturas (ResNet-34, ResNeXt-50, DenseNet-121 e VGG16). As séries temporais incluíram um total de dez imagens mensais cobrindo todo o ciclo de plantio do arroz na região. As redes utilizadas foram adaptadas para uso com as séries temporais, permitindo a extração de padrões em todos os tempos ao longo da série. Este estudo também avaliou três conjuntos de dados utilizando três combinações de polarizações de radar: (a) somente VV, (b) somente VH e (c) VV e VH em conjunto. Além da acurácia geral, foram utilizadas medidas para avaliar os modelos: o índice F1, a área abaixo da curva Precision–Recall (AUPRC) e a medida mean Intersection over Union (mIoU). Os resultados mostram que o conjunto de dados usando ambas as polarizações produziu os melhores resultados, seguido dos dados com somente VH e por ultimo os dados com somente VV. Modelos utilizando somente a polarização VV mostraram resultados consideravelmente piores que os demais (aproximadamente 10% menos mIoU que somente VH e 15% menos comparado aos dados VV+VH). Através desta avaliação, a arquitetura LinkNet utilizando a estrutura ResNeXt-50 mostrou os melhores resultados com acurácia de 0.98, F1 de 0.93, AUPRC de 0.93 e mIoU de 0.91, revelando que modelos de DL podem ser utilizados para o mapeamento de plantios de arroz utilizando imagens de radar.

Artigo publicado na revista *Remote Sensing Applications: Society and Environment*, 2021. DOI: 10.1016/j.rsase.2021.100627

Irrigated Rice Crop Identification in Southern Brazil Using Convolutional Neural Networks and Sentinel-1 Time Series

Abstract

Rice is one of the world's staple food sources, with millions of tonnes produced and consumed every year. Therefore, mapping rice paddies is essential for agricultural management and ensuring food security. This study aimed to classify rice crops in southern Brazil using the SENTINEL-1 SAR time series and deep learning models, comparing two architectures (U-net and LinkNet) and four backbones (ResNet-34, ResNeXt-50, DenseNet-121, and VGG16). The time series construction considered ten images, each for a month, covering the rice planting cycle. The Convolutional Neural Network architectures were adapted to use multi-band data, allowing the extraction of features from all-temporal images. This approach provides capturing spatiotemporal information from rice plantations, which favors its detection. Besides, the research evaluated three data sets considering the polarizations: (a) VV-only, (b) VH-only, and (c) both VV and VH (VV+VH). The classification accuracies used to measure the performance of the models were the overall accuracy, F1-measure, area under the precision-recall curve (AUPRC), and the intersection over union (IoU). Results show that the VH+VV polarization combination yielded the best results, followed by VH-only and VV-only. The VV-only polarization had significantly worst results (nearly 10% less IoU than VH-only and nearly 15% less IoU compared to VV+VH). The results show that rice fields can be successfully classified with deep learning models and through our evaluation the LinkNet architecture with the ResNeXt-50 backbone showed the best results with an accuracy of 0.98, F1 of 0.93, AUPRC of 0.93, and IoU of 0.91.

Keywords: deep learning, semantic segmentation, land cover classification, agriculture, SAR

1. Introduction

Rice is one of the world's staple food sources. In 2020, global rice production was approximately 514 million tonnes (FAO, 2021a), with a dominance of Asian countries such as China and India. Among the world's top 10 producers, Brazil is the only non-Asian country, producing an average of approximately 12 million tonnes of rice per year (FAO, 2021b). Most of the Brazilian rice production comes from irrigated rice produced in the state of Rio Grande do Sul, encompassing 80% of the country's production (CONAB, 2020).

The spatial analysis and dynamics of rice production are fundamental for defining long-term development strategies and decision-making aimed at food security and sustainable agricultural systems. In this context, remote sensing data acquired by orbital sensors became an indispensable tool for offering low-cost, efficient, and quick solutions for crop mapping, particularly through free public data obtained by platforms such as the Landsat and Sentinel satellites (Chaves et al., 2020). In the mapping of rice culture, one of the primary approaches is the phenological analysis based on optical or radar time series, distinguishing the crops by its main planting stages (Dong et al., 2016; Kuenzer & Knauer, 2013; Mosleh et al., 2015; Van Niel & McVicar, 2004; Zhao et al., 2021). Rice cultivation through the lowland method typically has agronomic flooding phase in which a continuous or periodic layer of water covers the soil to control weeds and pests. This chronological variation of the water-soil-rice

components is a critical factor for identifying the crops due to the differences in spectral and textural responses received by satellite sensors.

Many propositions have considered optical time series in rice paddy mapping, such as multispectral data from the LANDSAT-8 and SENTINEL-2 satellites (Qin et al., 2015; Dong et al., 2016; Zhou et al., 2016). However, a permanent challenge in studies with optical sensor data is the geographical and temporal discontinuity due to cloud cover that impacts the detection of the rice paddy's diagnostic phases (Chen et al., 2020). Agronomic flooding that lasts only 1–2 months in the rainy season is difficult to appear in optical images due to weather conditions. Therefore, the radar data has the advantages of being cloudless, independent of sunlight, day and night data acquisition, and uninterrupted monitoring of the entire crop cycle. Although multitemporal radar images have been used since the 1990s to map rice (Le Toan et al., 1997; Liew et al., 1998; Kurosu et al., 1997), it had historically fewer applications than optical images. One of the main reasons was the acquisition cost of dense Synthetic Aperture Radar (SAR) time series covering large rice regions. This condition changed with the advent of Sentinel-1 satellite images with global coverage, free access to data, and high temporal resolution from the two-satellite constellation with a 6-day revisit cycle at the equator.

Different algorithms have been used to SAR time series-based rice classifications considering the different growing stages demarcated by distinct SAR backscatter values: (a) low values in the period of flooding/transplantation due to the presence of water, (b) significant increase during the plant growth period, and (c) decrease after the harvesting phase due to the land surface. Initially, rice mapping used simpler methods such as threshold-based segmentation (Bouvet & Le Toan, 2011; Chen et al., 2007; Yang et al., 2008) and maximum likelihood technique (Kurosu et al., 1997; Choudhury & Chakraborty, 2004). A second phase establishes the machine learning (ML) domain, producing greater precision in higher dimensional feature space than traditional parametric classifiers. Among the main ML methods used in the rice paddy classification are decision tree (DT) classifier (He et al., 2018; Nguyen et al., 2015, 2016), rule-based classifier (Nelson et al., 2014), Random Forest (RF) (Chen et al., 2020; Lasko et al., 2018; Singha et al., 2019; Subbarao et al., 2020), and Support Vector Machines(SVM) (Gao et al., 2018; Hoang et al., 2016; Li et al., 2012; Minh et al., 2019). Some studies compare the different ML techniques such as SVM and RF (Mansaray et al., 2020; Son et al., 2018; Park et al., 2018); DT and RF (Bazzi et al., 2019); DT, SVM, and K-nearest neighbor (k-NN) (Küçük et al., 2016); and DT, SVM, k-NN, and quadratic discriminant analysis (Chang et al., 2021).

Most recently, deep learning (DL) techniques bring significant improvements in digital image processing due to the high ability to learn and represent data at various abstraction levels, being the state-of-the-art for computer vision problems (Liu et al., 2020; Guo et al., 2018). In remote sensing, DL gained high popularity demonstrated by recent reviews (Ball et al., 2018; Ma et al., 2019; Yuan

et al., 2020; Zhu et al., 2017). The DL algorithms overcame the ML methods, presenting better crop classification results, mostly due to their ability to extract and recognize textures and shapes in multi-dimensional data (Kussul et al., 2017; Zhong et al., 2019). The few DL studies on rice mapping based on SAR time series highlights temporal profile analysis, in which the most used methods are the unidimensional convolutional neural networks (1D CNN) and different models of Recurrent Neural Networks (RNN): Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM) fully connected recurrent neural networks (FCRNNs), and Domain-adapted RNNs (DARNNs) (Crisóstomo de Castro Filho et al., 2020; Jo et al., 2020; Ndikumana et al., 2018; Zhao et al., 2019).

However, these DL approaches using sequential data analysis consider only temporal information, disregarding rice plantations' spatial pattern demarcated by homogeneous morphology and irrigation channels that descend through the terraces. In this context, the Convolutional Neural Networks (CNNs) over spatial dimension allows an efficient semantic segmentation based on textural attributes. The CNNs use convolutional layers to aggregate neighboring pixels' values through matrix operations, being well-suited for remote sensing data's spatial nature (Li, 2019; Wei et al., 2019; Cao & Zhang, 2020). These architectures use a downsampling-up sampling process that learned both high-level semantic information and low-level detail in a very memory-efficient way. A variety of recent studies have successfully classified and detected changes in land use using CNN algorithms and SAR data (Liu et al., 2017; Mehra et al., 2020; Tiwari et al., 2020; Zhang et al., 2020). Nevertheless, the number of similar rice detection studies is still small, and aspects of the classification process remain relatively unexplored given the ample number of choices, such as optimal architectures, variables, and data structure.

The objective of this study was to evaluate CNN models for irrigated rice crop mapping in a southern region of Brazil using the SENTINEL-1 time series. The research compared two architectures (U-net and LinkNet), and 4 backbones (ResNeXt-50, ResNet-34, DenseNet-121, and VGG16). We also compared three datasets: vertical–vertical (VV) polarized bands, vertical–horizontal (VH) polarized bands, and both polarizations (VV+VH). Lastly, we also offer a brief analysis of the backscatter behavior of the rice crops along its agricultural cycle. The main contribution of this study lies in the proposal of a novel method of automatic, high accuracy classification of lowland rice crops that is free of cloud cover using publicly available radar data.

2. Material and methods

2.1. Study Area

Our area of study encompassed two rectangular areas around the central region of the state of Rio Grande do Sul, Brazil's southernmost state (Figure 1). The first area was reserved as the training site and is situated between longitudes 53°53.287' and 53°03.837' west, and latitudes 29°99.327' and 29°51.544' south. The second site was used exclusively to test the models, and is situated between longitudes 53°54.491' and 54°03.941' west, and latitudes 29°81.219' and 30°29.001' south. Both areas covered approximately 2500 km² of land.

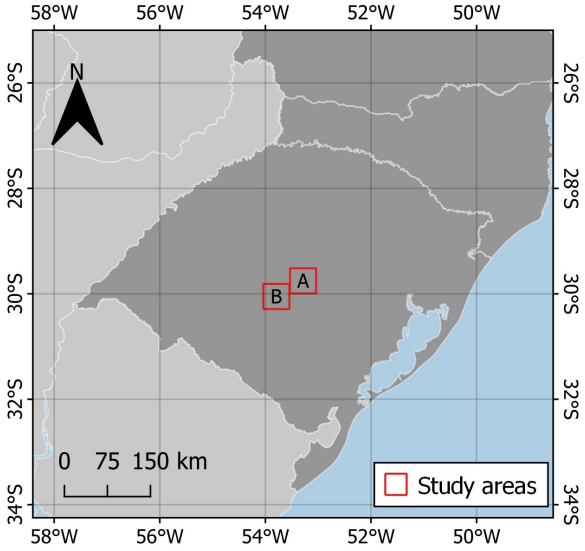

Figure 1. Location of the (A) training and (B) test sites in the state of Rio Grande do Sul.

The region is among the top rice-producing spots in the country and according to Brazil's *Instituto Rio Grandense do Arroz* (IRGA, available at https://irga.rs.gov.br/safras-2), which is responsible for accounting rice production within the state, rice planting typically starts in September and ends in the first two weeks of December, while the harvest of the plants happens between mid-December and May of the following year.

2.2. Ground truth data

The ground truth (GT) data of the irrigated rice crops that we used was produced by the Brazilian National Supply Company (CONAB, available at https://www.conab.gov.br/info-agro/safras). This GT map was created by experts through visual interpretation and analysis of the phenological and spectral behavior of crops in Sentinel-2 images obtained between August 2019 and May 2020 (Figure 2), supported by the use of high-resolution Google Earth Pro imagery. While the region is also

known to produce soybean and corn crops, the rice fields offer a significantly distinct phenological behavior that helps manually map this particular crop. We obtained this map as a shapefile, which was subsequently converted into binary raster data for compatibility with the DL models.

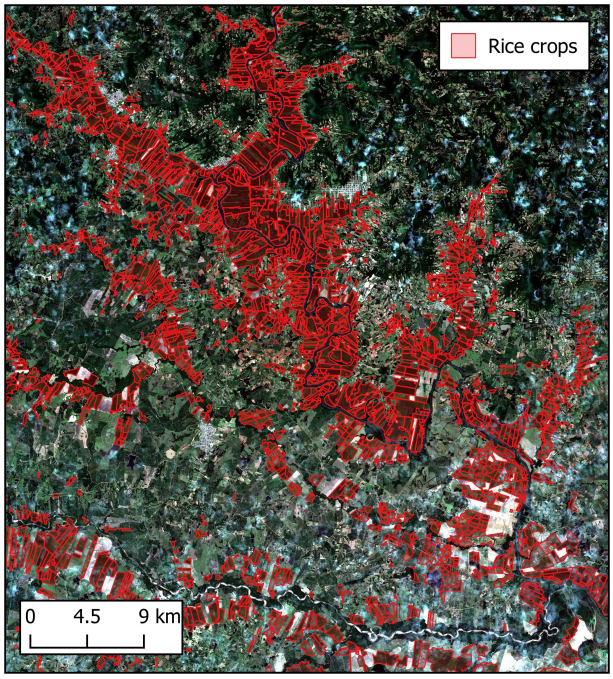

Figure 2. Ground truth map for the training area over a true color SENTINEL-2 image.

2.3. Sentinel-1B data

In this study, we used Sentinel-1B SAR data in interferometric wide swath (IW) mode, ground range detected (GRD). The acquisition of images was done through the Google Earth Engine (GEE) platform (Gorelick et al., 2017), which offers pre-processed 10 meter resolution dual-polarization (VV and VH) C-Band imagery with the thermal noise removal, radiometric calibration, and terrain correction steps already applied. We selected the first monthly image between August 2019 and May 2020 to coincide with the time window of our ground truth data, totaling ten sets of images (Table 1) given the typical revisit time of 12 days for Sentinel-1 images. We evaluated three data sets composed of single and dual-band polarizations (VV-only, VH-only, and VV+VH). Furthermore, we used the complete time series between August 2019 and June 2020 to briefly analyze the backscatter evolution in rice crops and their planting and harvest seasons.

2.4. Data structure

The training of the DL models considered spatiotemporal samples composed of smaller patches with vertical and horizontal dimensions of 256 $\times$ 256 pixels and depth of 10 for the single polar-

Table 1. Detailed sensing dates of the SENTINEL-1B images used in this study.

Image Set	Month	Year	Day	Unique ID
1	August	2019	05	6A0A
2	September	2019	10	B326
3	October	2019	04	9FD9
4	November	2019	09	C69C
5	December	2019	03	5EDD
6	January	2020	08	7CD8
7	February	2020	01	8EDD
8	March	2020	08	E363
9	April	2020	01	C51A
10	May	2020	07	9A39

ization datasets and 20 for the dual-polarization dataset. The preparation of samples used a sliding window approach with a slight overlap (32 pixels) between patches. This process generated a total of 528 training patches, considering an image with 4,888x5,393 spatial dimensions (approximately 2636km²). The same process was applied to generate the testing samples, yielding the same number of images. We randomly selected 20% of the test samples (105 patches) to be used for the validation of the models during training. Therefore, the total number of samples for the model training, validation, and testing were 423 (40%), 105 (10%), and 528 (50%), respectively, with individual each patch containing a total of 655,360 data pixels in the case of the single polarization datasets and 1,310,720 in the case of the dual-polarization dataset.

2.5. Deep learning models

In this study, the models considered two basic architectures: U-Net (Ronneberger et al., 2015) and LinkNet (Chaurasia & Culurciello, 2017). Both architectures have encoder and decoder paths. The encoder path is usually a CNN for extracting features, and the decoder is responsible for recovering the original image dimensions. Despite their similarity (Figure 3), these architectures present two main differences: (a) the LinkNet has extra convolutional layers in the upsampling path; and (b) LinkNet weight propagation uses the addition operation, while the U-Net architecture concatenates the feature maps.

We considered four backbone types for each architecture: VGG16 (Simonyan & Zisserman, 2014), ResNet-34 (He et al., 2016), DenseNet-121 (Huang et al., 2018), and ResNeXt-50 (Xie et al., 2017), totaling eight combinations. The models were build using the Segmentation Models python package (Yakubovskiy, 2019), which offers pre-built model architectures and backbone structures using the Keras and Tensorflow frameworks.

The backbone structures yield different results since they vary in complexity (number of parameters), depth (e.g., ResNet-50 and ResNet-101), and the type of connections. The most suitable struc-

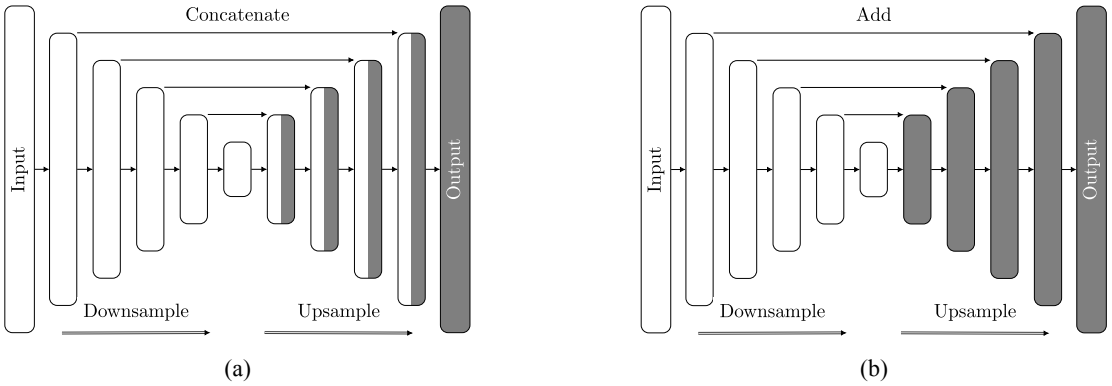

Figure 3. General structure of the (a) U-Net and (b) LinkNet architectures. Adapted from Yakubovskiy (2019)

ture is often a tradeoff between computational and predictive power. Regarding the total number of parameters, the VGG16 is the simplest, followed by ResNet, DenseNet, and, lastly, ResNeXt. The backbones present different characteristics: VGG16 is mostly composed of MaxPooling, and convolutional layers (Figure 4a); the ResNet and DenseNet modules use residual connections, which improves the propagation of information and reduces problems such as vanishing gradients (Figures 4b and 4c); and the ResNeXt structure (Figure 4d) introduces the concept of cardinality to its convolutional blocks which further increases the complexity of the models.

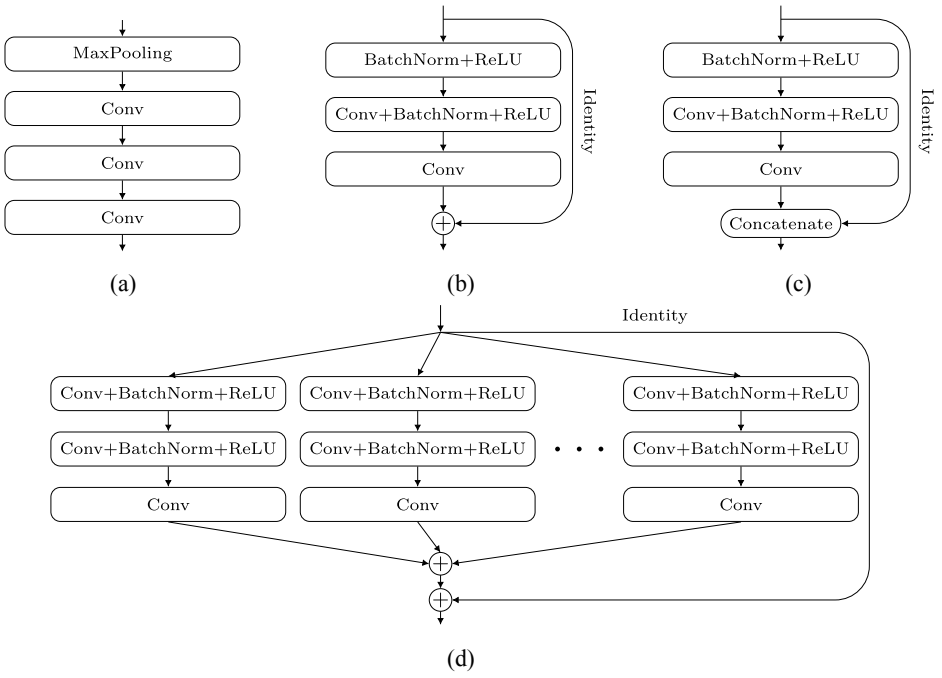

Figure 4. Modified (a) VGG16, (b) ResNet, (c) DenseNet and (d) ResNeXt modules used to build the downsampling path of the networks in this study.

We applied random horizontal and vertical flip augmentation strategies to increase the number of

samples synthetically and avoid overfitting. Moreover, the present study used a computer equipped with an NVIDIA GTX 1070 graphics card and an 8-core AMD 3700X processor.

Regarding hyperparameters, we used: (a) the RMSprop optimizer; (b) 100 epochs; (c) batch size of 16; and (d) learning rate starting at 10^{-4}, configured to automatically decrease by 10^{-1} every time the loss plateaued to stabilize the final weights. Lastly, the loss function used was the Dice loss (Equation 1).

$$DiceLoss = \frac{(1+\beta^2) \times TP}{(1+\beta^2) \times FP + \beta^2 \times FN + FP} \tag{1}$$

where TP, FP, and FN are the number of true positive, false positive, and false-negative observations within the batch and β is a balancing coefficient. When $\beta = 1$, the Dice loss equates 1 minus the F1 measure.

2.6. Model evaluation

Model evaluation on the test dataset considered the following metrics: (a) overall accuracy; (b) the F1 measure, (c) the value of the area under the Precision-Recall curve (AUPRC), and (d) the intersection over union (IoU) value, also known as the Jaccard-index. The Precision, Recall and F1 values are formally given by Equation 2, while the IoU value is given by Equation 3. The AUPRC value is the area under the curve given by plotting Precision-Recall pairs at different classification probability thresholds, measured through the trapezoidal rule.

$$Precision = \frac{TP}{TP + FP}\,, \tag{2a}$$

$$Recall = \frac{TP}{TP + FN}\,, \tag{2b}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN}\,, \tag{2c}$$

$$IoU = \frac{TP}{TP + FN + FP}, \tag{3}$$

where TP, TN, FP and FN are the number of true positive, true negative, false positive and false negative observations respectively in a classification with a probability threshold of 0.5.

These metrics are commonly used to assess land cover classification tasks with DL algorithms and usually offer a better measure of classification correctness as they consider class distribution imbalances, unlike the overall accuracy measure. In this study, the observed class ratio of rice crop

area to background area was approximately 1:5. All evaluation metrics range from 0 to 1, where values closer to 1 indicate the best result.

3. Results

3.1. Backscatter analysis

The backscattering behavior within the area of study follows the established planting and harvest seasons (Figure 5). As the crop fields were prepared and sown, the backscatter values decreased, increasing again after the crops' initial growing period. However, the rice sowing in the study area was slightly delayed compared to the usual agricultural calendar and other state regions due to excessive rainfall in October and November, and the sown area reached 50% on November 25, with some producers finishing sowing only in December (IRGA, 2021). The lowest average value of σ_0 corresponds to the point where the rice sowing in the region was in the middle (Figure 6) according to the data provided by IRGA (2021) (Figure 6). There was no evident relationship between backscattering behavior and harvest completion rates in the region, possibly because the grown crops and post-harvest fields show similar σ_0 values due to the adoption of the minimum tillage technique within the region. This technique minimizes the amount of tilling done after harvest, leaving the seed banks and the water sheet relatively untouched (Elias, 1969; Chauhan & Johnson, 2009), resulting in less noticeable textural differences pre and post harvest.

3.2. Training results

Training for 100 epochs was enough for all models to reach a point where they stopped improving considerably, and all models learned most of the information in the first few epochs. The training curves of the models were similar, but the models using VGG16 modules displayed a lower learning limit than the other models (Figure 7). The speed at which the models reached stable validation error values varied depending on the combination of architecture, backbone modules, and training dataset (Figure 8). Models using the ResNet and ResNeXt modules generally reached stable validation values faster than the others and improved smoothly. Conversely, models with the VGG16 and DenseNet modules took longer to find optimal weights and stabilize errors.

Despite offering the best results, models with ResNeXt modules took approximately twice as long to train as the other models (Figure 9). Conversely, models using the simpler ResNet modules showed the fastest training times on average. The processing times for generating the predictions followed the same trend. To classify the test set, the models using ResNet, VGG16 and DenseNet modules took an average of 5 seconds, whereas the models with ResNeXt modules showed an average of 12 seconds.

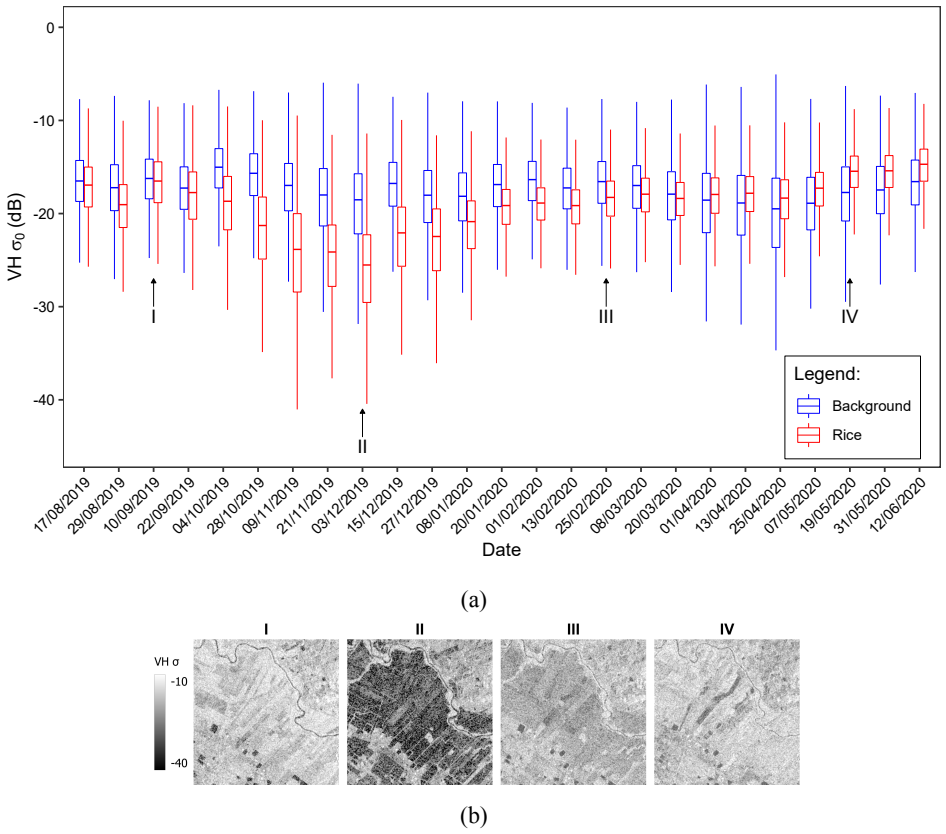

Figure 5. (a) Monthly boxplot distributions of VH σ_0 values within our area of study between the months of August 2019 and June 2020 and (b) example fields at specific crop stages: (I) unprepared, at the start of the planting season, (II) sown area close to 50%, (III) grown crops mid harvest season and (IV) at the end of harvest season.

3.3. Classification results

Table 2 lists the evaluation metrics for all combinations (architecture and backbones). The LinkNet architecture with the ResNeXt backbone trained with the VH+VV dataset presented the best results. However, results were similar to the U-Net architecture under the same conditions. There was no clear advantage when using one architecture or the other (Table 3). The differences in performance are mostly related to the polarization data used to train the model and backbone modules.

Using both polarizations to train the models yielded better results overall compared to using a single polarization (Table 4). Alternatively, using only the VH polarization also showed better results compared to using the VV polarization. Models using the VGG16 modules showed the worst results for each polarization category (Table 5), which is expected due to the lower complexity and lack of residual connections in their structure.

Figure 10 shows the precision-recall curves of the models. For the most part, all curves were similar, except the VGG16 curves, which showed noticeably smaller areas than the other backbone types in all cases. The curves for the ResNeXt backbones showed a slight edge over the others, as

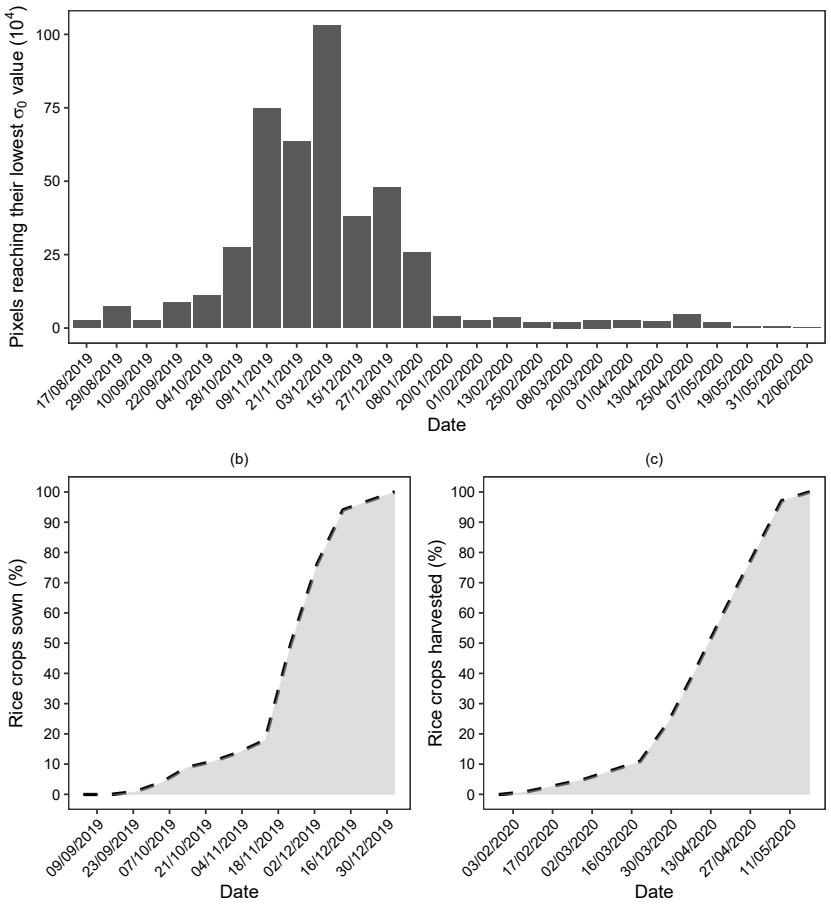

Figure 6. (a) Number of pixels reaching their lowest σ_0 values by date and the completion rates provided by the IRGA for the (b) sowing and (c) harvesting of rice crops along its agricultural cycle.

seen from the AUPRC values in Table 2.

Visually, the classifications using the dataset with both polarizations also clearly showed more accurate and complete results when compared to the results using single-polarization datasets (Figures 11 and 12). A significant portion of the misclassifications happened in the thin division between the individual plots of irrigated rice, as the models tended to group multiple plots into a single patch.

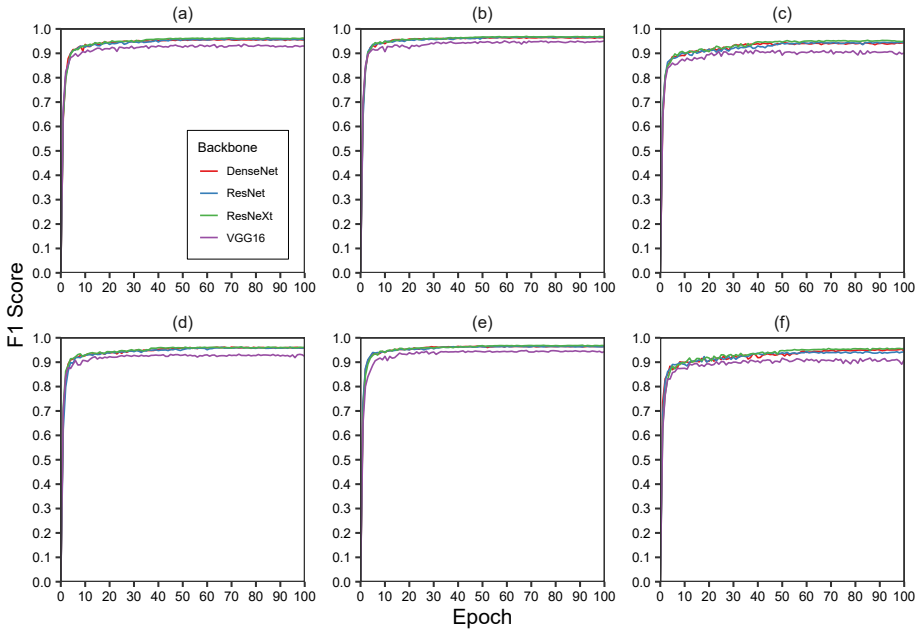

Figure 7. Training F1 scores at each epoch for each architecture and polarization combination. (a) LinkNet+VH, (b) LinkNet+VHVV, (c) LinkNet+VV, (d) U-Net+VH, (e) U-Net+VHVV and (f) U-Net+VV.

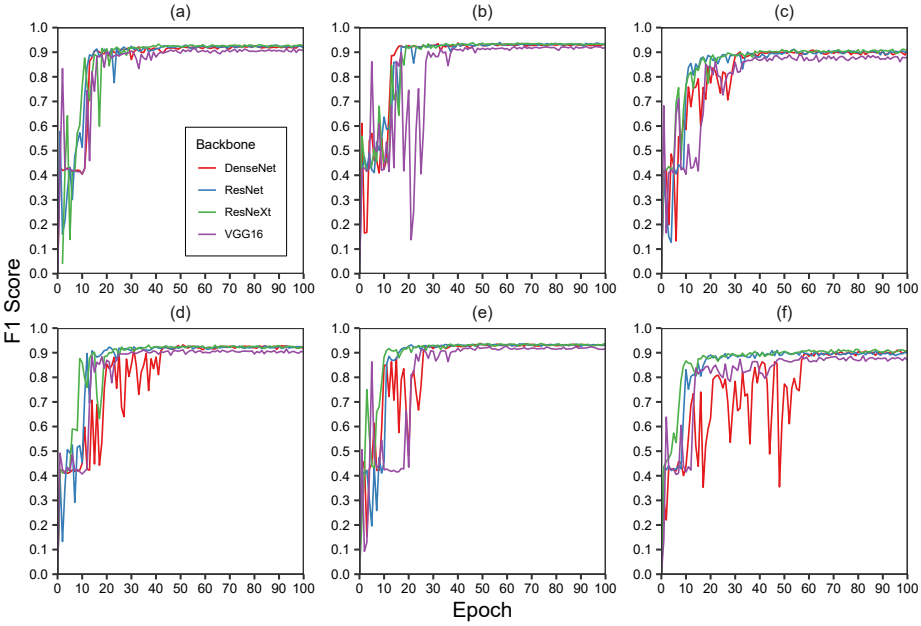

Figure 8. Validation F1 scores at each epoch for each architecture and polarization combination. (a) LinkNet+VH, (b) LinkNet+VHVV, (c) LinkNet+VV, (d) U-Net+VH, (e) U-Net+VHVV and (f) U-Net+VV.

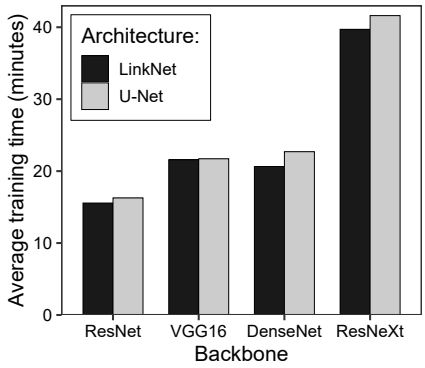

Figure 9. Average training time in minutes for all architecture and backbone combinations using an NVIDIA GTX 1070 video card and AMD 3700X 8-core processor.

Table 2. Classification performance metric results for all model combinations using the test dataset, sorted from best to worst accuracy.

Polarization	Architecture	Backbone	Accuracy	Precision	Recall	F1	AUPRC	IoU
VH+VV	LinkNet	ResNeXt	0.9871	0.8925	0.9325	0.9120	0.9368	0.9122
VH+VV	U-Net	ResNet	0.9867	0.8950	0.9227	0.9087	0.9114	0.9092
VH+VV	U-Net	ResNeXt	0.9866	0.8966	0.9181	0.9072	0.9213	0.9079
VH+VV	LinkNet	ResNet	0.9863	0.8830	0.9326	0.9071	0.9306	0.9077
VH+VV	U-Net	DenseNet	0.9862	0.8928	0.9174	0.9050	0.9054	0.9058
VH+VV	LinkNet	DenseNet	0.9852	0.8942	0.8996	0.8969	0.9135	0.8986
VH	LinkNet	ResNet	0.9793	0.8229	0.9060	0.8625	0.8886	0.8680
VH	LinkNet	ResNeXt	0.9793	0.8154	0.9191	0.8641	0.8756	0.8693
VH	U-Net	ResNeXt	0.9784	0.8083	0.9158	0.8587	0.8771	0.8647
VH	LinkNet	DenseNet	0.9781	0.8068	0.9136	0.8569	0.8651	0.8631
VH	U-Net	DenseNet	0.9780	0.8108	0.9038	0.8548	0.8769	0.8614
VH+VV	LinkNet	VGG16	0.9778	0.8195	0.8856	0.8513	0.9057	0.8587
VH+VV	U-Net	VGG16	0.9775	0.8445	0.8411	0.8428	0.8856	0.8522
VH	U-Net	ResNet	0.9774	0.8050	0.9030	0.8512	0.8418	0.8584
VH	U-Net	VGG16	0.9726	0.7715	0.8774	0.8210	0.8795	0.8336
VH	LinkNet	VGG16	0.9710	0.7744	0.8405	0.8061	0.8552	0.8222
VV	U-Net	ResNeXt	0.9642	0.7356	0.7806	0.7574	0.7427	0.7858
VV	LinkNet	ResNeXt	0.9625	0.7174	0.7868	0.7505	0.7327	0.7805
VV	LinkNet	DenseNet	0.9596	0.7006	0.7610	0.7295	0.6834	0.7658
VV	LinkNet	ResNet	0.9594	0.6902	0.7853	0.7347	0.7112	0.7688
VV	U-Net	DenseNet	0.9578	0.6687	0.8133	0.7340	0.7374	0.7674
VV	U-Net	ResNet	0.9542	0.6426	0.8139	0.7181	0.7166	0.7558
VV	U-Net	VGG16	0.9518	0.6487	0.7140	0.6798	0.6729	0.7321
VV	LinkNet	VGG16	0.9479	0.6194	0.7062	0.6600	0.6515	0.7188

Table 3. Mean values of the performance measures per architecture type, sorted by accuracy.

Architecture	Accuracy	Precision	Recall	F1	AUPRC	IoU
LinkNet	0.9728	0.7864	0.8557	0.8193	0.8292	0.8361
U-Net	0.9726	0.7850	0.8601	0.8199	0.8307	0.8362

Table 4. Mean values of the performance measures per polarization, sorted by accuracy.

Polarization	Accuracy	Precision	Recall	F1	AUPRC	IoU
VH+VV	0.9842	0.8773	0.9062	0.8914	0.9138	0.8940
VH	0.9768	0.8019	0.8974	0.8469	0.8700	0.8551
VV	0.9572	0.6779	0.7701	0.7205	0.7061	0.7594

Table 5. Mean values of the performance measures per backbone structure, sorted by accuracy.

Backbone	Accuracy	Precision	Recall	F1	AUPRC	IoU
ResNeXt	0.9764	0.8110	0.8755	0.8417	0.8477	0.8534
DenseNet	0.9741	0.7956	0.8681	0.8295	0.8303	0.8437
ResNet	0.9739	0.7898	0.8772	0.8304	0.8334	0.8447
VGG16	0.9664	0.7463	0.8108	0.7768	0.8084	0.8029

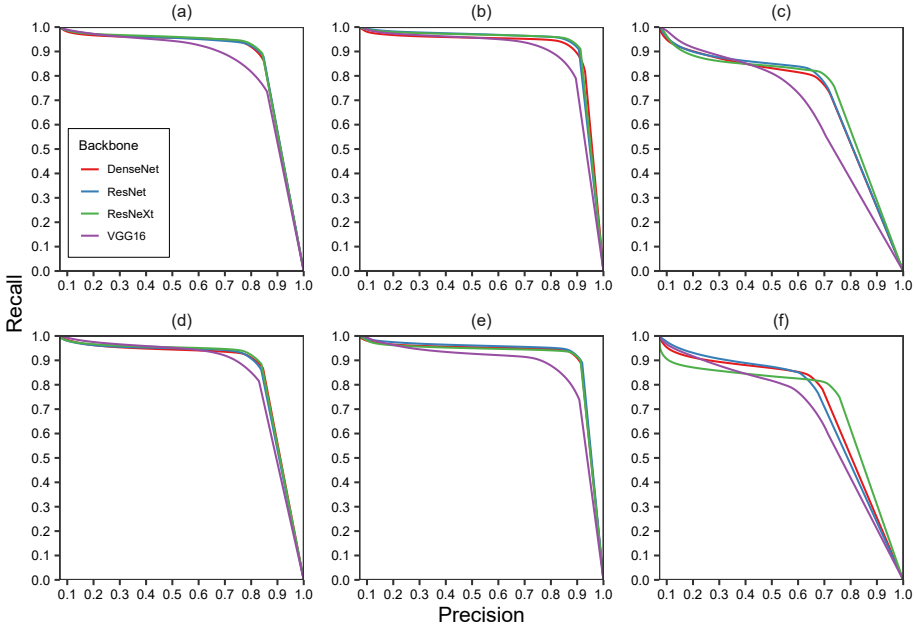

Figure 10. Precision–Recall curves for each architecture and polarization combination. (a) LinkNet+VH, (b) LinkNet+VHVV, (c) LinkNet+VV, (d) U-Net+VH, (e) U-Net+VHVV and (f) U-Net+VV.

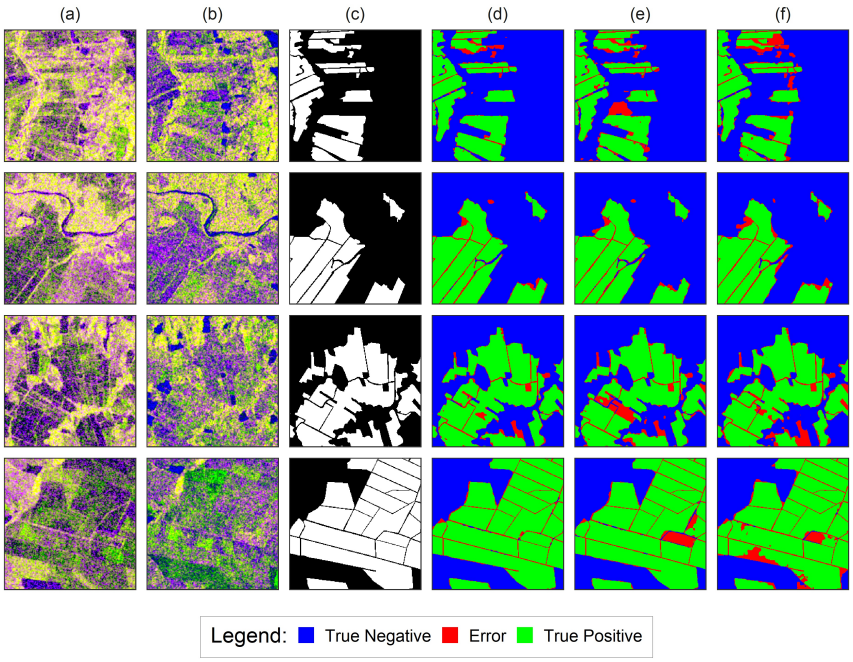

Figure 11. Example validation sample sites with false color RGB C-SAR composites using VV, VH and VV/VH ratio in the months of (a) December and (b) February, along with (c) the ground truth mask and LinkNet+ResNeXt classifications using the (d) VV+VH, (e) VH and (f) VV training datasets.

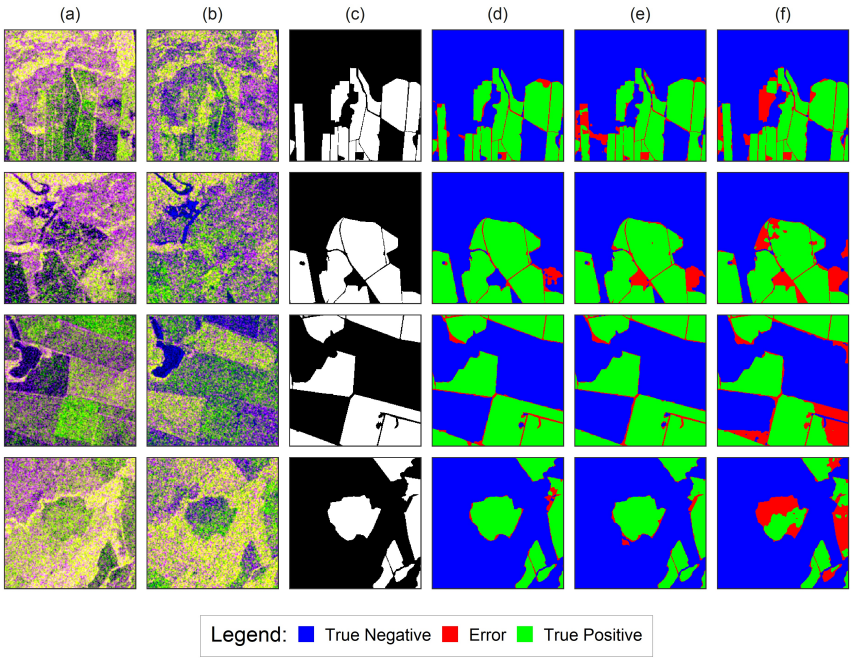

Figure 12. Example test sample sites with false color RGB C-SAR composites using VV, VH and VV/VH ratio in the months of (a) December and (b) February, along with (c) the ground truth mask and LinkNet+ResNeXt classifications using the (d) VV+VH, (e) VH and (f) VV training datasets.

4. Discussion

The results show that CNNs on radar images are an efficient way to classify irrigated rice crops. The temporal data sets using different polarization combinations demonstrate an accuracy performance where VH + VV $>$ VH-only $>$ VV-only, which is in line with other studies involving the classification of rice with Sentinel data. Comparisons between the two Sentinel-1 polarizations for rice growth stages establish that VH backscattering is more efficient than VV (Nguyen et al., 2016; Chang et al., 2021; Minh et al., 2019; Ndikumana et al., 2018). In turn, the classification using the two polarizations together shows better results than a single polarization (Lasko et al., 2018). Studies combining optical and radar images also show the best performance by including the two polarizations in the classification procedures (Chen et al., 2020; Wu et al., 2019).

One of the main aspects of applying CNN to rice detection is to incorporate spatial information, which is a limitation in pixel-based time series analysis. We introduced the temporal information as additional bands in our datasets, obtaining phenology and textural-based classification with relevant results. The spatial pattern of rice fields has characteristics different from other land uses, partly due to the homogeneous distribution demarcated by irrigation channels. However, few SAR rice mappings have explored the spatial pattern. One of the main approaches to extract textural information in remote sensing images is through the use of Gray Level Co-occurrence Matrixes (GLCM) on different land uses or land covers (Mishra et al., 2017; Zakeri et al., 2017). However, Kim & Yeom (2014) assess that GLCM textures using RapidEye multi-spectral satellite image contributed little or even caused a slight decrease in the rice crop precision. Some SAR studies for rice detection consider different sets of input images, including GLGM texture images (Ngo et al., 2020). An alternative approach to pixel-based analysis is the Object-Based Image Analysis (OBIA) (Blaschke, 2010), which has been applied in tasks such as mapping paddy rice, considering spatial information, and minimizing the speckle effect (Cai et al., 2019; Clauss et al., 2018a,b; Singha & Sarmah, 2019). However, methods based on CNN-based methods have significant advantages over OBIA-based methods, such as greater accuracy, less human supervision, direct transferability to other regions or scenes with different characteristics, and systematic reproduction based on free software (Guirado et al., 2017).

Besides, our results show better performance measures when compared to studies with simpler pixel-based time series analysis and 1D DL models (Zhao et al., 2019; Liao et al., 2020). In our study, a training sample of 423 patches with a height$\times$width of 256$\times$256 pixels (27721728 total data points) proved to be satisfactory. Zhang et al. (2018) also obtained similar accuracy and kappa index results coupling optical data and vegetation indexes with a simple CNN model. The results obtained show that the textural information over a time series using CNNs is also adequate for mapping rice, with

the advantage of being free of cloud cover.

Performance metrics show that models with ResNeXt modules offered the best results, although they need more time to train than others. Models with ResNet modules showed comparable results in less than half the training time. While time is not necessarily a constraint, training times also reflect how complex and computationally intensive the models can be. Given the general scope of remote sensing data, available computer memory can often be a constraint, making efficient memory models more desirable. In the mapping of rice fields in Brazil, future studies can evaluate the integration of optical and radar images. Much research on rice detection uses this combination of data (Mansaray et al., 2020; Onojeghuo et al., 2018; Park et al., 2018; Iyyappan & Ramakrishnan, 2020). The limitation is the optical data gap due to cloud coverage, making it challenging to apply trained networks in one year in the others.

5. Conclusion

The results show that the CNNs can classify irrigated rice croplands with high accuracy using SAR time-series datasets. The image classification used eight models composed of different combinations of architecture and backbone modules, considering single (VV and VH) and dual-polarization (VV+VH) C-SAR images between August 2019 and May 2020. There was no considerable difference in performance between the LinkNet and U-Net architectures. Conversely, the different backbone modules yielded varying results, obtaining the best results with ResNeXt and the worst with VGG16. Furthermore, the choice of SAR polarizations for the training dataset also greatly influenced the classification results. The dual VV+VH polarization dataset offered better results than the single VH polarization, which offered considerably better results than the VV polarization. The best-performing model using the LinkNet architecture and ResNeXt modules trained with the VH+VV dataset yielded accuracy, F1, AUPRC, and IoU values of 0.98, 0.93, 0.93, and 0.91, respectively.

We have provided an extensive look into the usage of CNN architectures to classify rice crops using SAR data. However, some aspects of the process can be subject to further studies. The applicability of this methodology is currently valid for irrigated rice crops in southern Brazil. Therefore, we encourage further studies in other regions. Besides, we recommend investigating other SAR products such as L and X-Band data with deep learning algorithms since the different SAR wavelengths might provide different responses on this type of crop.

References

Ball, J. E., Anderson, D. T., & Chan, C. S. (2018). Special Section Guest Editorial: Feature and Deep Learning in Remote Sensing Applications. *Journal of Applied Remote Sensing*, *11*, 1.

doi:10.1117/1.JRS.11.042601.

Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., Minh, D. H. T., Ndikumana, E., Courault, D., & Belhouchette, H. (2019). Mapping Paddy Rice Using Sentinel-1 SAR Time Series in Camargue, France. *Remote Sensing*, *11*, 887. doi:10.3390/rs11070887.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, *65*, 2–16. doi:10.1016/j.isprsjprs.2009.06.004.

Bouvet, A., & Le Toan, T. (2011). Use of ENVISAT/ASAR wide-swath data for timely rice fields mapping in the Mekong River Delta. *Remote Sensing of Environment*, *115*, 1090–1101. doi:10.1016/j.rse.2010.12.014.

Cai, Y., Lin, H., & Zhang, M. (2019). Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. *Advances in Space Research*, *64*, 2233–2244. doi:10.1016/j.asr.2019.08.042.

Cao, K., & Zhang, X. (2020). An Improved Res-UNet Model for Tree Species Classification Using Airborne High-Resolution Images. *Remote Sensing*, *12*, 1128. doi:10.3390/rs12071128.

Chang, L., Chen, Y.-T., Wang, J.-H., & Chang, Y.-L. (2021). Rice-field mapping with sentinel-1a SAR time-series data. *Remote Sensing*, *13*, 103. doi:10.3390/rs13010103.

Chauhan, B. S., & Johnson, D. E. (2009). Influence of tillage systems on weed seedling emergence pattern in rainfed rice. *Soil and Tillage Research*, *106*, 15–21. doi:10.1016/j.still.2009.10.004.

Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. *2017 IEEE Visual Communications and Image Processing (VCIP)*, (pp. 1–4). doi:10.1109/VCIP.2017.8305148. arXiv:1707.03718.

Chaves, M., Picoli, M., & Sanches, I. (2020). Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sensing*, *12*, 3062. doi:10.3390/rs12183062.

Chen, J., Lin, H., & Pei, Z. (2007). Application of ENVISAT ASAR data in mapping rice crop growth in Southern China. *IEEE Geoscience and Remote Sensing Letters*, *4*, 431–435. doi:10.1109/lgrs.2007.896996.

Chen, N., Yu, L., Zhang, X., Shen, Y., Zeng, L., Hu, Q., & Niyogi, D. (2020). Mapping paddy rice fields by combining multi-temporal vegetation index and synthetic aperture radar remote

sensing data using google earth engine machine learning platform. *Remote Sensing*, *12*, 2992. doi:10.3390/rs12182992.

Choudhury, I., & Chakraborty, M. (2004). Analysis of temporal SAR and optical data for rice mapping. *Journal of the Indian Society of Remote Sensing*, *32*, 373–385. doi:10.1007/bf03030862.

Clauss, K., Ottinger, M., & Künzer, C. (2018a). Mapping rice areas with sentinel-1 time series and superpixel segmentation. *International Journal of Remote Sensing*, *39*, 1399–1420. doi:10.1080/01431161.2017.1404162.

Clauss, K., Ottinger, M., Leinenkugel, P., & Kuenzer, C. (2018b). Estimating rice production in the mekong delta, vietnam, utilizing time series of sentinel-1 sar data. *International journal of applied earth observation and geoinformation*, *73*, 574–585. doi:10.1016/j.jag.2018.07.022.

CONAB (2020). Arroz. In *Acompanhamento Da Safra Brasileira: Grãos*. Brasília, Brazil: Companhia Nacional de Abastecimento volume 8.

Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O. L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., Rosa Silva, C., Guimarães Ferreira, P. H., Fontes Guimarães, R., & Trancoso Gomes, R. A. (2020). Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series. *Remote Sensing*, *12*, 2655. doi:10.3390/rs12162655.

Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., Biradar, C., & Moore, B. (2016). Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sensing of Environment*, *185*, 142–154. doi:10.1016/j.rse.2016.02.016.

Elias, R. S. (1969). Rice production and minimum tillage. *Outlook on Agriculture*, *6*, 67–71. doi:10.1177/003072706900600204.

FAO (2021a). Food and Agriculture Organization of the United Nations - FAOSTAT Platform. URL: http://www.fao.org/faostat/en/#data/QC/.

FAO (2021b). Food and Agriculture Organization of the United Nations - World Food Situation: Cereal Supply and Demand Brief. URL: http://www.fao.org/worldfoodsituation/csdb/en/.

Gao, Q., Zribi, M., Escorihuela, M. J., Baghdadi, N., & Segui, P. Q. (2018). Irrigation mapping using Sentinel-1 time series at field scale. *Remote Sensing*, *10*, 1495. doi:10.3390/rs10091495.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. doi:10.1016/j.rse.2017.06.031.

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., & Herrera, F. (2017). Deep-learning versus obia for scattered shrub detection with google earth imagery: Ziziphus lotus as case study. *Remote Sensing*, *9*, 1220. doi:10.3390/rs9121220.

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, *7*, 87–93. doi:10.1007/s13735-017-0141-z.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). Las Vegas, NV, USA: IEEE. doi:10.1109/CVPR.2016.90. arXiv:1512.03385.

He, Z., Li, S., Wang, Y., Dai, L., & Lin, S. (2018). Monitoring rice phenology based on backscattering characteristics of multi-temporal RADARSAT-2 datasets. *Remote Sensing*, *10*, 340. doi:10.3390/rs10020340.

Hoang, H. K., Bernier, M., Duchesne, S., & Tran, Y. M. (2016). Rice mapping using RADARSAT-2 dual-and quad-pol data in a complex land-use Watershed: Cau River Basin (Vietnam). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*, 3082–3096. doi:10.1109/jstars.2016.2586102.

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2018). Densely Connected Convolutional Networks, . arXiv:1608.06993.

IRGA (2021). Instituto Rio Grandense do Arroz - Boletim de Resultados da Lavoura de Arroz - Safra 2019/2020. URL: https://irga.rs.gov.br/safras-2.

Iyyappan, M., & Ramakrishnan, S. S. (2020). Enhancing land cover classification for multispectral images using hybrid polarimetry sar data. *International Journal of Remote Sensing*, *41*, 6718–6754. doi:10.1080/01431161.2020.1750730.

Jo, H.-W., Lee, S., Park, E., Lim, C.-H., Song, C., Lee, H., Ko, Y., Cha, S., Yoon, H., & Lee, W.-K. (2020). Deep Learning Applications on Multitemporal SAR (Sentinel-1) Image Classification Using Confined Labeled Data: The Case of Detecting Rice Paddy in South Korea. *IEEE Trans. Geosci. Remote Sensing*, *58*, 7589–7601. doi:10.1109/TGRS.2020.2981671.

Kim, H.-O., & Yeom, J.-M. (2014). Effect of red-edge and texture features for object-based paddy rice crop classification using rapideye multi-spectral satellite image data. *International Journal of Remote Sensing*, *35*, 7046–7068. doi:10.1080/01431161.2014.965285.

Küçük, Ç., Taşkın, G., & Erten, E. (2016). Paddy-rice phenology classification based on machine-learning methods using multitemporal co-polar X-band SAR images. *IEEE Journal of selected topics in applied earth observations and remote sensing*, *9*, 2509–2519. doi:10.1109/jstars.2016.2547843.

Kuenzer, C., & Knauer, K. (2013). Remote sensing of rice crop areas. *International Journal of Remote Sensing*, *34*, 2101–2139. doi:10.1080/01431161.2012.738946.

Kurosu, T., Fujita, M., & Chiba, K. (1997). The identification of rice fields using multi-temporal ERS-1 C band SAR data. *International Journal of Remote Sensing*, *18*, 2953–2965. doi:10.1080/014311697217143.

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sensing Lett.*, *14*, 778–782. doi:10.1109/LGRS.2017.2681128.

Lasko, K., Vadrevu, K. P., Tran, V. T., & Justice, C. (2018). Mapping double and single crop paddy rice with Sentinel-1A at varying spatial scales and polarizations in Hanoi, Vietnam. *IEEE journal of selected topics in applied earth observations and remote sensing*, *11*, 498–512. doi:10.1109/jstars.2017.2784784.

Le Toan, T., Ribbes, F., Wang, L.-F., Floury, N., Ding, K.-H., Kong, J. A., Fujita, M., & Kurosu, T. (1997). Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 41–56. doi:10.1109/36.551933.

Li, K., Brisco, B., Yun, S., & Touzi, R. (2012). Polarimetric decomposition with RADARSAT-2 for rice mapping and monitoring. *Canadian Journal of Remote Sensing*, *38*, 169–179. doi:10.5589/m12-024.

Li, L. (2019). Deep Residual Autoencoder with Multiscaling for Semantic Segmentation of Land-Use Images. *Remote Sensing*, *11*, 2142. doi:10.3390/rs11182142.

Liao, C., Wang, J., Xie, Q., Baz, A. A., Huang, X., Shang, J., & He, Y. (2020). Synergistic Use of Multi-Temporal RADARSAT-2 and VENμS Data for Crop Classification Based on 1D Convolutional Neural Network. *Remote Sensing*, *12*, 832. doi:10.3390/rs12050832.

Liew, S. C., Kam, S.-P., Tuong, T.-P., Chen, P., Minh, V. Q., & Lim, H. (1998). Application of multitemporal ERS-2 synthetic aperture radar in delineating rice cropping systems in the Mekong River Delta, Vietnam. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1412–1420. doi:10.1109/36.718845.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, *128*, 261–318. doi:10.1007/s11263-019-01247-4.

Liu, T., Li, Y., Cao, Y., & Shen, Q. (2017). Change detection in multitemporal synthetic aperture radar images using dual-channel convolutional neural network. *J. Appl. Remote Sens*, *11*, 1. doi:10.1117/1.JRS.11.042615.

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. doi:10.1016/j.isprsjprs.2019.04.015.

Mansaray, L. R., Wang, F., Huang, J., Yang, L., & Kanu, A. S. (2020). Accuracies of support vector machine and random forest in rice mapping with Sentinel-1A, Landsat-8 and Sentinel-2A datasets. *Geocarto International*, *35*, 1088–1108. doi:10.1080/10106049.2019.1568586.

Mehra, A., Jain, N., & Srivastava, H. S. (2020). A novel approach to use semantic segmentation based deep learning networks to classify multi-temporal SAR data. *Geocarto International*, (pp. 1–16). doi:10.1080/10106049.2019.1704072.

Minh, H. V. T., Avtar, R., Mohan, G., Misra, P., & Kurasaki, M. (2019). Monitoring and mapping of rice cropping pattern in flooding area in the Vietnamese Mekong Delta using Sentinel-1A data: A case of an Giang Province. *ISPRS International Journal of Geo-Information*, *8*, 211. doi:10.3390/ijgi8050211.

Mishra, V. N., Prasad, R., Kumar, P., Gupta, D. K., & Srivastava, P. K. (2017). Dual-polarimetric c-band sar data for land use/land cover classification by incorporating textural information. *Environmental Earth Sciences*, *76*, 1–16.

Mosleh, M. K., Hassan, Q. K., & Chowdhury, E. H. (2015). Application of remote sensors in mapping rice area and forecasting its production: A review. *Sensors*, *15*, 769–791. doi:10.3390/s150100769.

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep recurrent

neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*, 1217. doi:10.3390/rs10081217.

Nelson, A., Setiyono, T., Rala, A. B., Quicho, E. D., Raviz, J. V., Abonete, P. J., Maunahan, A. A., Garcia, C. A., Bhatti, H. Z. M., Villano, L. S. et al. (2014). Towards an operational SAR-based rice monitoring system in Asia: Examples from 13 demonstration sites across Asia in the RIICE project. *Remote Sensing*, *6*, 10773–10812. doi:10.3390/rs61110773.

Ngo, K. D., Lechner, A. M., & Vu, T. T. (2020). Land cover mapping of the mekong delta to support natural resource management with multi-temporal sentinel-1a synthetic aperture radar imagery. *Remote Sensing Applications: Society and Environment*, *17*, 100272. doi:10.1016/j.rsase.2019.100272.

Nguyen, D. B., Clauss, K., Cao, S., Naeimi, V., Kuenzer, C., & Wagner, W. (2015). Mapping rice seasonality in the Mekong Delta with multi-year Envisat ASAR WSM data. *Remote Sensing*, *7*, 15868–15893. doi:10.3390/rs71215808.

Nguyen, D. B., Gruber, A., & Wagner, W. (2016). Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data. *Remote Sensing Letters*, *7*, 1209–1218. doi:10.1080/2150704x.2016.1225172.

Onojeghuo, A. O., Blackburn, G. A., Wang, Q., Atkinson, P. M., Kindred, D., & Miao, Y. (2018). Mapping paddy rice fields by applying machine learning algorithms to multi-temporal sentinel-1a and landsat data. *International journal of remote sensing*, *39*, 1042–1067. doi:10.1080/01431161.2017.1395969.

Park, S., Im, J., Park, S., Yoo, C., Han, H., & Rhee, J. (2018). Classification and mapping of paddy rice by combining Landsat and SAR time series data. *Remote Sensing*, *10*, 447. doi:10.3390/rs10030447.

Qin, Y., Xiao, X., Dong, J., Zhou, Y., Zhu, Z., Zhang, G., Du, G., Jin, C., Kou, W., Wang, J., & Li, X. (2015). Mapping paddy rice planting area in cold temperate climate region through analysis of time series Landsat 8 (OLI), Landsat 7 (ETM+) and MODIS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 220–233. doi:10.1016/j.isprsjprs.2015.04.008.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, . arXiv:1505.04597.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, *abs/1409.1*. arXiv:1409.1556.

Singha, M., Dong, J., Zhang, G., & Xiao, X. (2019). High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Scientific data*, *6*, 1–10. doi:10.1038/s41597-019-0036-3.

Singha, M., & Sarmah, S. (2019). Incorporating crop phenological trajectory and texture for paddy rice detection with time series modis, hj-1a and alos palsar imagery. *European Journal of Remote Sensing*, *52*, 73–87. doi:10.1080/22797254.2018.1556568.

Son, N.-T., Chen, C.-F., Chen, C.-R., & Minh, V.-Q. (2018). Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto international*, *33*, 587–601. doi:10.1080/10106049.2017.1289555.

Subbarao, N. V. T., Mani, J. K., Shrivastava, A., Srinivas, K., & Varghese, A. (2020). Acreage estimation of kharif rice crop using Sentinel-1 temporal SAR data. *Spatial Information Research*, (pp. 1–11). doi:10.1007/s41324-020-00374-2.

Tiwari, A., Narayan, A. B., & Dikshit, O. (2020). Deep learning networks for selection of measurement pixels in multi-temporal SAR interferometric processing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *166*, 169–182. doi:10.1016/j.isprsjprs.2020.06.005.

Van Niel, T., & McVicar, T. (2004). Current and potential uses of optical remote sensing in rice-based irrigation systems: A review. *Australian Journal of Agricultural Research*, *55*, 155–185. doi:10.1071/ar03149.

Wei, S., Zhang, H., Wang, C., Wang, Y., & Xu, L. (2019). Multi-Temporal SAR Data Large-Scale Crop Mapping Based on U-Net Model. *Remote Sensing*, *11*, 68. doi:10.3390/rs11010068.

Wu, W., Wang, W., Meadows, M. E., Yao, X., & Peng, W. (2019). Cloud-based typhoon-derived paddy rice flooding and lodging detection using multi-temporal sentinel-1&2. *Frontiers of Earth Science*, *13*, 682–694.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks, . doi:10.1109/CVPR.2017.634. arXiv:1611.05431.

Yakubovskiy, P. (2019). Segmentation models. https://github.com/qubvel/segmentation_models.

Yang, S., Shen, S., Li, B., Le Toan, T., & He, W. (2008). Rice mapping and monitoring using ENVISAT ASAR data. *IEEE Geoscience and Remote Sensing Letters*, *5*, 108–112. doi:10.1109/LGRS.2007.912089.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J. et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716. doi:10.1016/j.rse.2020.111716.

Zakeri, H., Yamazaki, F., & Liu, W. (2017). Texture analysis and land cover classification of tehran using polarimetric synthetic aperture radar imagery. *Applied Sciences*, *7*, 452. doi:10.3390/app7050452.

Zhang, M., Lin, H., Wang, G., Sun, H., & Fu, J. (2018). Mapping Paddy Rice Using a Convolutional Neural Network (CNN) with Landsat 8 Datasets in the Dongting Lake Area, China. *Remote Sensing*, *10*, 1840. doi:10.3390/rs10111840.

Zhang, X., Liu, G., Zhang, C., Atkinson, P. M., Tan, X., Jian, X., Zhou, X., & Li, Y. (2020). Two-Phase Object-Based Deep Learning for Multi-Temporal SAR Image Change Detection. *Remote Sensing*, *12*, 548. doi:10.3390/rs12030548.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., & Feng, M. (2019). Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China. *Remote Sensing*, *11*, 2673. doi:10.3390/rs11222673.

Zhao, R., Li, Y., & Ma, M. (2021). Mapping paddy rice with satellite remote sensing: A review. *Sustainability*, *13*, 503. doi:10.3390/su13020503.

Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, 430–443. doi:10.1016/j.rse.2018.11.032.

Zhou, Y., Xiao, X., Qin, Y., Dong, J., Zhang, G., Kou, W., Jin, C., Wang, J., & Li, X. (2016). Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. *International Journal of Applied Earth Observation and Geoinformation*, *46*, 1–12. doi:10.1016/j.jag.2015.11.001.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*, 8–36. doi:10.1109/MGRS.2017.2762307. arXiv:1710.03959.

Capítulo V

Discussão

1. Contribuições

As contribuições oferecidas por esta tese podem ser separadas sob duas perspectivas. Sob uma ótica da ciência internacional, procurou-se expandir o espaço metodológico conhecido sobre a construção e otimização de parâmetros de algoritmos de Deep Learning. Nos três artigos apresentados foram explorados diversos modelos e suas variações a fim de fornecer uma base comparativa para futuras pesquisas. Já sob a ótica do desenvolvimento científico brasileiro, a contribuição residiu no estudo do uso destas metodologias para o aprimoramento de produtos nacionais de sensoriamento remoto críticos para o desenvolvimento ambiental e agrícola do país.

Apesar da escala relativamente pequena coberta pelos estudos apresentados nesta tese, acredita-se que estes algoritmos são aplicáveis em maior escala a fim de cobrir regiões criticas como a Amazônia legal e o Cerrado brasileiro considerando-se um número suficiente de amostras. Uma das maiores vantagens destes modelos reside no fato de que podem ser treinados indefinidamente conforme o fornecimento de amostras adicionais através das quais os parâmetros dos modelos são atualizados. No entanto, é importante notar que introduzir os modelos a regiões consideravelmente diferentes das quais eles foram treinados pode acarretar em perdas de desempenho, a depender da capacidade que os modelos possuem de generalizar a informação conhecida.

As arquiteturas do tipo *ConvDeconv*, ou *Autoencoders*, se mostraram altamente eficientes pare realizar todos os tipos de classificações realizadas no desenvolvimento desta tese, e portanto são recomendadas como uma base para o desenvolvimento de algoritmos em pesquisas futuras. Em geral, modelos mais complexos ofereceram melhores resultados, mas é importante considerar que maiores graus de complexidade acarretam na necessidade de um maior poder de processamento.

2. Revisitando os artigos

Ao longo do desenvolvimento deste trabalho, muitas mudanças aconteceram no campo de Deep Learning e muito se aprendeu e esclareceu sobre aspectos do fundamentais e metodo-

lógicos. Com isto em mente, é importante refletir sobre certos pontos que mereceram maior consideração e aprofundamento nos artigos apresentados.

Uma das maiores diferenças entre as metodologias dos artigos produzidos nesta tese foi quanto ao número de amostras utilizadas nos modelos. Acreditava-se, inicialmente, que os modelos necessitavam de uma quantidade consideravelmente maior de amostras dedicadas para seu treinamento relativo ao número de amostras para validação para que os modelos fossem suficientemente treinados. Nos primeiros dois artigos (Capítulos II e III) aproximadamente 66% das amostras foram utilizadas para treinar os modelos. No terceiro artigo (Capítulo IV), apesar da redução dessa proporção para 40% de amostras de treinamento os modelos se mostraram igualmente eficientes. A determinação da divisão entre amostras de treinamento e teste irá depender do tamanho das amostras e da quantidade de bandas devido a limitações em função do hardware utilizado, porém, a recomendação que segue esta conclusão é a de que estudos verifiquem diferentes proporções treinamento–validação–teste a fim de esclarecer a questão de suficiência amostral com algoritmos de DL. Apesar da possibilidade da perda de desempenho, uma proporção maior de amostras de teste representaria melhor a capacidade generalizadora dos modelos.

Nos artigos produzidos, pouco se discutiu sobre o padrão de distribuição dos erros de classificação. Nota-se, ao observar as classificações realizadas nos artigos, que a maior parte dos erros de classificação se encontram nas bordas ou faixas de transição entre classes. Isso possivelmente se atribui aos diversos gradientes de mudança de classe que podem ser observados na realidade. Em todos os artigos as classificações tiveram resultados binários onde somente foram indicadas a classe de interesse e uma única classe representando todas as demais classes. É possível que mapeamentos onde todas as classes mais distintas sejam representadas apresente gradientes mais claros para o aprendizado dos modelos, porem é importante lembrar que a elaboração de uma máscara contendo todas as classes é consideravelmente mais dispendioso em relação a uma máscara binária.

Nos trabalhos onde foram utilizadas imagens de sensores ópticos Landsat, houve uma falta da avaliação do uso de índices espectrais para complementar as bandas ópticas apesar de sua conhecida utilidade. Não se sabe quão grande seria a contribuição destes índices para o resultado das classificações, porém é possível levantar a hipótese de que sua utilização poderia ao menos acelerar o processo de treinamento por criarem uma síntese da relação entre bandas as quais fossem de maior interesse para o objeto de estudo.

3. Próximos passos

Embora os estudos apresentados tenham explorado um amplo espaço metodológico, ao longo do seu desenvolvimento e conclusão se identificaram certos pontos sujeitos a possíveis futuras investigações:

- Os três estudos se contiveram a um único tipo de dado de imagem de satélite. Os estudos nos capítulos II e III utilizaram somente imagens de sensores ópticos, enquanto o estudo

do capítulo IV fez uso de imagens de radar. Existem, na literatura, estudos onde diferentes tipos de imagens são utilizados em conjunto ou em paralelo (He e Yokoya 2018; Bermudez et al. 2018), porém seu uso para detecção de alvos similares aos apresentados nesta tese ainda é relativamente inexplorado. Acredita-se que o uso de diferentes tipos de imagem em um ambiente de DL permitiria o aproveitamento dos aspectos positivos de cada tipo como, por exemplo, a informação textural de imagens de radar e a informação espectral contida em imagens ópticas.

- Apesar da utilização de séries temporais, os estudos se limitaram à utilização de redes compostas em maior parte por camadas convolucionais 2D, limitando a modo como os dados puderam ser estruturados (informação espacial e canais ou bandas). A utilização de camadas 3D disponíveis nos frameworks de DL possivelmente ampliaria o nível de complexidade de dados, porém com um custo adicional de processamento. Adicionalmente, existem atualmente camadas especializadas na extração de padrões temporais, tal como as camadas *Long-short term memory* (LSTM), que têm sido aplicadas em um número crescente de estudos de sensoriamento remoto (Teimouri et al. 2019; Crisóstomo de Castro Filho et al. 2020).

- Há uma distinta falta de estudos avaliando a quantidade de amostras necessárias para atingir bons resultados de classificação. O estudo no capítulo II avaliou diferentes tamanhos de amostra, resultando também numa variação na quantidade de amostras, porém idealmente diferentes quantidades deveriam ser comparadas em função de um único tamanho de amostra.

Referências

Bermudez, J. D. et al. (set. de 2018). "SAR to Optical Image Synthesis for Cloud Removal with Generative Adversarial Networks". Em: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1, pp. 5–11. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-IV-1-5-2018.

Crisóstomo de Castro Filho, Hugo et al. (ago. de 2020). "Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series". Em: *Remote Sensing* 12.16, p. 2655. ISSN: 2072-4292. DOI: 10.3390/rs12162655.

He, Wei e Naoto Yokoya (2018). *Multi-temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation*. arXiv: 1807.09954 [cs.CV].

Teimouri, Nima, Mads Dyrmann e Rasmus Nyholm Jørgensen (abr. de 2019). "A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images". Em: *Remote Sensing* 11.8, p. 990. ISSN: 2072-4292. DOI: 10.3390/rs11080990.

Capítulo VI

Conclusões e considerações finais

Este trabalho teve sucesso em seu objetivo inicial de evidenciar o potencial do uso de algoritmos de Deep Learning para o monitoramento e da classificação de alvos críticos utilizando imagens de satélite no Brasil. Estes algoritmos demonstram diversas vantagens em relação a métodos tradicionais do campo do aprendizado de máquina, tais como melhores medidas de desempenho, maior velocidade de classificação e a possibilidade de aprimora-los de maneira incremental com a adição de novas amostras. No entanto, eles ainda apresentam certas considerações que dificultam seu uso, tais como a necessidade de uma grande quantidade de dados e de hardware específico para processamento.

Os três estudos apresentados previamente chegaram a conclusões satisfatórias em seus objetivos primários. No primeiro estudo, apresentado no Capítulo II, áreas queimadas no bioma Cerrado foram mapeadas com sucesso e evidenciou-se que as arquiteturas conhecidas como *Autoencoders* são excelentes candidatos para detecção de mudança em imagens ópticas de satélite. Adicionalmente, a avaliação de quatro tamanhos de amostras revelou que este é um fator a ser considerado em futuros estudos e que dimensões maiores de amostras não acarretam necessariamente em melhorias de desempenho.

Em seguida, no Capitulo III, demonstrou-se sucesso na a utilização de modelos similares porém com o objetivo de detectar áreas desmatadas na floresta amazônica. Neste caso, os algoritmos de DL foram comparados a modelos clássicos e mais simples encontrados na literatura, e observou-se que seu desempenho foi consideravelmente melhor, com classificações divergindo menos em relação à realidade.

Por fim, no estudo apresentado no Capítulo IV, houve um maior distanciamento metodológico em relação aos demais com o uso de imagens de radar para classificar plantios de arroz no estado do Rio Grande do Sul. Porém, de modo similar, chegou-se à conclusão de que estes modelos oferecem resultados excelentes de classificação. A principal comparação realizada neste estudo foi em relação ao tipo de arquitetura e os tipos de estruturas básicas dos quais elas podem ser compostas, evidenciando ainda mais o espaço e a complexidade metodológica ao qual estes estudos podem estar sujeitos.

Ao se observar as metodologias dos três estudos em paralelo, nota-se que há uma ampla variedade de escolhas e decisões no processo de desenvolvimento destes estudos. No entanto, certos

fatores mostram ganhos ou perdas consistentes em desempenho (tal como o tipo de arquitetura). A busca por um maior grau de normalização metodológica é um dos maiores desafios no campo, e o desenvolvimento de novas pesquisas é essencial para encontrar esses fatores consistentes e facilitar a adoção destes modelos.

Embora o grau de utilização destes algoritmos tenha crescido recentemente, sua adoção para fins públicos ou governamentais, tal como o monitoramento por imagens de satélite, ainda é baixa. Fatores como a falta de normalização metodológica e a natureza complexa destes algoritmos contribuem para reduzir a confiança neles e dificultar a sua adoção. Portanto, o desenvolvimento de pesquisas nesta área é essencial para sanar estes problemas, elucidando a natureza destes algoritmos para facilitar seu uso público. Adicionalmente, o campo de DL ainda se encontra em um constante processo de desenvolvimento e, dada a possibilidade da criação de algoritmos cada vez mais complexos e sofisticados, a tendência é que futuras metodologias obtenham ainda melhores resultados.