

UNIVERSITY OF SPLIT
FACULTY OF ELECTRICAL ENGINEERING, MECHANICAL ENGINEERING
AND NAVAL ARCHITECTURE

Antonia Ivanda

**THE IMPACT OF THE CONVOLUTION DEGREE
ON THE RESULTS OF CONVOLUTIONAL NEURAL
NETWORKS IN REMOTE SENSING PROBLEMS**

DOCTORAL THESIS

Split, 2024.

UNIVERSITY OF SPLIT
FACULTY OF ELECTRICAL ENGINEERING, MECHANICAL ENGINEERING
AND NAVAL ARCHITECTURE

Antonia Ivanda

*The impact of the convolution degree on the results of
convolutional neural networks in remote sensing problems*

DOCTORAL THESIS

Split, 2024.

The research reported in this thesis was carried out at Department of Electronics and Computer Science, University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture.

Supervisor: prof. dr. sc. Ljiljana Šerić, FESB, University of Split, Croatia

Dissertation number: xxx-xxx

BIBLIOGRAPHIC INFORMATION

Keywords: remote sensing, deep learning, convolutional neural network (1D, 2D, 3D and 4D-CNN), multispectral imagery, classification, regression, segmentation, change detection, PRISMA methodology

Scientific area: Technical Sciences

Scientific field: Computer Science

Scientific branch: Artificial Intelligence

Institution of PhD completion: University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture

Supervisor of the thesis: prof. dr. sc. Ljiljana Šerić

Number of pages: 206

Number of figures: 73

Number of tables: 19

Number of references: 337

Committee for assessment of doctoral dissertation:

1. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
2. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
3. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
4. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
5. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town

Committee for defence of doctoral dissertation:

1. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
2. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
3. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
4. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town
5. XXXX. dr. sc. XXXX Yyyyy, Institution name and City/Town

Dissertation defended on: xx. month. 20xx.

The impact of the convolution degree on the results of convolutional neural networks in remote sensing problems

Abstract:

In recent years, space technologies have become increasingly integral to various industrial and scientific domains, emphasizing the critical role of remote sensing and Earth observation in guiding decision-making processes. This doctoral thesis aims to enhance the understanding of remote sensing applications by leveraging Convolutional Neural Networks (CNNs) across different convolution degrees (1D, 2D, 3D, and 4D). The primary research question explores how the selection of convolution dimensionality influences CNN performance across diverse tasks. 1D, 2D, 3D, and 4D convolutions are selected for recognizing spectral, spatial, and temporal features of data. A comprehensive systematic literature review was conducted, analyzing publications related to 1D, 2D, 3D, and 4D-CNNs, addressing their domains of application, machine learning techniques used, and corresponding datasets. Findings are synthesized in the form of ontology for CNN architecture selection providing a structured framework, which is further evaluated in four case studies. The 1D-CNN proved to be an appropriate method for extracting spectral information sufficient for predicting point parameters such as Secchi disk depth. The 2D-CNN demonstrated success in extracting spatial features from Sentinel-2 images for road detection. The 3D-CNN with its ability to extract spectral-spatial features was successfully applied to the problem of fire propagation potential detection. The 4D-CNN combined temporal with spectral and spatial features, which proved extremely useful in detecting burned areas. Finally, a comparison of all four convolutional neural networks for the task of semantic segmentation of land cover was performed, where the 3D-CNN stood out as the best. Results of all case studies are aligned with the central hypothesis that selecting appropriate convolution degree significantly enhances deep learning performance in environmental monitoring applications. These findings contribute to understanding of convolution strategies in remote sensing, offering insights for researchers and practitioners seeking to optimize neural network architectures for Earth observation technologies.

Keywords:

remote sensing, deep learning, convolutional neural network (1D, 2D, 3D and 4D-CNN), multispectral imagery, classification, regression, segmentation, change detection, PRISMA methodology

Utjecaj stupnja konvolucije na rezultate konvolucijskih neuronskih mreža u problemima daljinskih istraživanja

Sažetak:

Posljednjih godina svemirske tehnologije postaju sve više sastavni dio različitih industrijskih i znanstvenih domena, naglašavajući ključnu ulogu daljinskog istraživanja i promatranja Zemlje u procesu donošenja odluka. Ova doktorska disertacija ima za cilj unaprijediti razumijevanje primjene daljinskog istraživanja korištenjem konvolucijskih neuronskih mreža (CNN) kroz različite stupnjeve konvolucije (1D, 2D, 3D i 4D). Primarno istraživačko pitanje istražuje kako odabir dimenzionalnosti konvolucije utječe na performanse CNN-a u različitim zadacima. 1D, 2D, 3D i 4D konvolucije odabrane su za prepoznavanje spektralnih, prostornih i vremenskih značajki podataka. Provedena je sveobuhvatna sistematska analiza literature, analizirajući publikacije vezane uz 1D, 2D, 3D i 4D-CNN, njihove domene primjene, korištene tehnike strojnog učenja i odgovarajuće skupove podataka. Nalazi su sintetizirani u obliku ontologije za odabir CNN arhitekture pružajući strukturirani okvir i dodatno su evaluirani u četiri studije slučaja. 1D-CNN pokazao se kao prikladna metoda za izdvajanje spektralnih informacija dovoljnih za predviđanje točkastih parametara poput Secchi dubine diska. 2D-CNN pokazao je uspjeh u izdvajanju prostornih značajki iz Sentinel-2 snimaka za detekciju cesta. 3D-CNN sa svojom sposobnošću izdvajanja spektralno-prostornih značajki uspješno je primijenjen na problem detekcije potencijala širenja požara. 4D-CNN kombinirao je vremenske sa spektralnim i prostornim značajkama, što se pokazalo izuzetno korisnim u otkrivanju izgorjelih područja. Konačno, provedena je usporedba sve četiri konvolucijske neuronske mreže za zadatak semantičke segmentacije zemljišnog pokrova, gdje se 3D-CNN istaknuo kao najbolji. Rezultati svih studija slučaja u skladu su s centralnom hipotezom da odabir odgovarajućeg stupnja konvolucije značajno poboljšava performanse dubokog učenja u aplikacijama za praćenje okoliša. Ovi nalazi doprinose razumijevanju konvolucijskih strategija u daljinskom istraživanju, pružajući uvide istraživačima i stručnjacima koji žele optimizirati arhitekture neuronskih mreža za tehnologije promatranja Zemlje.

Ključne riječi:

daljinsko istraživanje, duboko učenje, konvolucijska neuronska mreža (1D, 2D, 3D i 4D-CNN), multispektralne snimke, klasifikacija, regresija, segmentacija, detekcija promjena, PRISMA metodologija

Acknowledgments

This PhD thesis

Contents

Abstract	v
Sažetak	vi
Acknowledgments	viii
List of Tables	xiv
List of Figures	xviii
List of Acronyms and Symbols	xix
1 INTRODUCTION	1
1.1 Research Hypothesis	2
1.2 Dissertation Outline	3
2 REMOTE SENSING	5
2.1 The Electromagnetic Spectrum	5
2.2 Sensor Resolution	7
2.2.1 Spatial Resolution	7
2.2.2 Spectral Resolution	8
2.2.3 Radiometric Resolution	9
2.2.4 Temporal Resolution	9
2.3 Sensor Types	10
2.3.1 Active Sensors	10
2.3.2 Passive Sensors	12
2.4 Data Preprocessing Techniques	12
2.4.1 Radiometric Preprocessing	13
2.4.2 Atmospheric Correction	13
2.4.3 Geometric Preprocessing	15
2.4.4 Image Enhancement	15
2.4.5 Data Fusion	16
2.4.6 Processing Levels	16
2.5 Overview of Satellite Missions	17
3 DEEP LEARNING	21
3.1 Image Analysis Techniques in Remote Sensing	21

3.1.1	Image Preprocessing	22
3.1.2	Pixel-based Classification	23
3.1.3	Semantic Segmentation	23
3.1.4	Object-based Classification	23
3.1.5	Target Recognition	24
3.1.6	Regression Analysis	25
3.1.7	Scene Understanding	25
3.2	CNN Key Concepts	25
3.3	Architecture of CNN	28
3.3.1	Convolutional Layer	29
3.3.2	Pooling Layer	29
3.3.3	Fully Connected Layer	30
3.4	CNN training process	30
3.4.1	Data Preprocessing	30
3.4.2	Data Augmentation	31
3.4.3	Parameter Initialization	31
3.4.4	Regularization	31
3.4.5	Optimization	32
3.5	CNN Applications in Remote Sensing	32
3.5.1	1D-CNN (Spectral/Temporal)	33
3.5.2	2D-CNN (Spatial)	34
3.5.3	3D-CNN (Spectral-Spatial)	34
3.5.4	4D-CNN (Spectral-Spatial-Temporal)	35

4 SYSTEMATIC LITERATURE REVIEW OF CNN ARCHITECTURES IN MULTISPECTRAL IMAGERY 37

4.1	Introduction	37
4.1.1	Problem formulation	38
4.2	Materials and Methods	38
4.2.1	Eligibility Criteria	39
4.2.2	Information Sources	39
4.2.3	Search Strategy	40
4.2.4	Data Extraction	41
4.2.5	Effect measures	41
4.3	Results and Discussion	41
4.3.1	Study Selection	41
4.3.2	Overview of Remote Sensing Publications	42
4.3.3	Meta-Analysis of Publications	47

4.3.4	Data Ontology for CNN Architecture Applications in Remote Sensing Publications	62
4.3.5	Limitations and Future Work	64
4.4	Conclusion	64
5	EVALUATION OF 1D-CNN MODEL	65
5.1	Introduction	65
5.2	Materials and Methods	68
5.2.1	Study Area	68
5.2.2	Secchi Disk Depth	68
5.2.3	Sentinel-3 OLCI Data	70
5.2.4	Dataset Construction and Preprocessing	72
5.2.5	1D-CNN Architecture	73
5.2.6	Hyperparameter Tuning	75
5.2.7	Accuracy Assessment	75
5.2.8	Overview of Regression Algorithms for Comparison with 1D-CNN	78
5.3	Results	79
5.3.1	Quantitative Algorithm Performance	79
5.3.2	Performance Analysis of 1D-CNN and Commonly Used Regression Algorithms	81
5.3.3	The Spatial Distribution of Secchi Disk Depth	84
5.4	Discussion	86
5.4.1	Accuracy of the 1D-CNN Model	87
5.4.2	Limitations	87
5.4.3	Applicability to Other Regions	88
5.4.4	Future Implications for Long-term Monitoring	88
5.5	Conclusion	89
6	EVALUATION OF 2D-CNN MODEL	91
6.1	Introduction	91
6.2	Materials and Methods	93
6.2.1	Study Area	93
6.2.2	Dataset Construction and Preprocessing	94
6.2.3	Model Implementation	95
6.2.4	Evaluation Metrics	97
6.2.5	Overview of Common CNN Architectures for Comparison with 2D-CNN	98
6.3	Results	100
6.3.1	2D-CNN Model Performance	100
6.3.2	Comparison with DeepLabv3, ResNet-50 and U-Net	102

6.4	Discussion	104
6.4.1	Accuracy of the 2D-CNN Model	104
6.4.2	Limitations	104
6.4.3	Applicability to Other Regions	105
6.4.4	Future Implications for Long-term Monitoring	105
6.5	Conclusion	105
7	EVALUATION OF 3D-CNN MODEL	107
7.1	Introduction	107
7.2	Materials and Methods	109
7.2.1	Study Area	109
7.2.2	Dataset Construction and Preprocessing	110
7.2.3	Model Implementations	111
7.3	Results	115
7.3.1	3D-CNN Model Performance	115
7.3.2	Comparison with 1D-CNN and 2D-CNN Models	117
7.3.3	Qualitative Evaluation of FPP Prediction Maps	119
7.4	Discussion	122
7.4.1	Accuracy of the 3D-CNN Model	122
7.4.2	Limitations	122
7.4.3	Future Research Directions	123
7.5	Conclusion	123
8	EVALUATION OF 4D-CNN MODEL	125
8.1	Introduction	125
8.2	Materials and Methods	127
8.2.1	Study Area	127
8.2.2	Dataset Construction and Preprocessing	127
8.2.3	Model Implementation	129
8.3	Results	130
8.3.1	4D-CNN Model Performance	130
8.3.2	Qualitative Analysis	134
8.4	Discussion	136
8.4.1	Performance and Insights of the 4D-CNN Model	137
8.4.2	Limitations	139
8.4.3	Future Research Directions	139
8.5	Conclusion	140
9	A COMPARATIVE ANALYSIS OF CNN MODELS FOR SEMANTIC SEGMENTATION	141

9.1	Introduction	141
9.2	Materials and Methods	142
9.2.1	Study Area	142
9.2.2	Dataset Construction and Preprocessing	142
9.2.3	CNN Architecture	144
9.3	Results	145
9.3.1	Performance Analysis of CNN Architectures	146
9.3.2	Qualitative Analysis	148
9.4	Discussion	149
9.5	Conclusion	150
10	CONCLUSION	151
	BIBLIOGRAPHY	155

List of Tables

2.1	<i>Overview of satellites for remote sensing and their basic parameters</i>	20
4.1	<i>Eligibility criteria.</i>	39
4.2	<i>Search strategy for different databases and CNNs</i>	40
4.3	<i>1D-CNN publications that implement regression</i>	55
4.4	<i>2D-CNN publications that implement regression</i>	56
4.5	<i>3D-CNN publications that implement regression</i>	57
4.6	<i>Publications that implement segmentation</i>	59
5.1	<i>Comparison 1D-CNN performance for different datasets (Adapted from [141])</i>	81
5.2	<i>Comparison of regression algorithms metrics (Adapted from [141])</i>	83
5.3	<i>Comparison of in situ secchi depth measurements with kd_z90max predicted values from C2RCC processor and Secchi disk depth from 1D-CNN model (Adapted from [141])</i>	86
6.1	<i>Comparison of models on the train dataset</i>	103
6.2	<i>Comparison of models on the test dataset</i>	103
7.1	<i>CNN model comparison for FPP prediction on training dataset</i>	118
7.2	<i>CNN model comparison for FPP prediction on testing dataset</i>	118
8.1	<i>Comparison of Remote Sensing Methods for Burned Area Detection</i>	138
9.1	<i>IoU Scores Comparison of Different CNN Architectures</i>	146
9.2	<i>Dice Coefficient Comparison of Different CNN Architectures</i>	146
9.3	<i>Computational Performance of CNN Architectures</i>	147
9.4	<i>Comprehensive Performance Analysis of CNN Architectures</i>	149

List of Figures

2.1	<i>Electromagnetic spectrum divided into bands (Adapted from [14])</i>	6
2.2	<i>Interaction between solar energy and the Earth’s surface</i>	7
2.3	<i>Campus of the University of Split in four different spatial resolution</i>	8
2.4	<i>Examples of spectral resolution for the campus of the University of Split</i>	9
2.5	<i>Examples of radiometric resolution for the campus of the University of Split</i>	10
2.6	<i>Sensor Types</i>	11
2.7	<i>Example of atmospheric correction on Sentinel–2 imagery, credits: "Modified Copernicus Sentinel data 2024/Sentinel Hub" [29]</i>	14
3.1	<i>Convolution operation on a 5×5 input image using a 2×2 kernel</i>	26
3.2	<i>Kernel movement with a stride of one</i>	27
3.3	<i>Zero padding</i>	28
3.4	<i>Base architecture of convolutional neural network</i>	29
3.5	<i>Overfitting and Underfitting of Data (Adapted from [77])</i>	32
3.6	<i>1D–CNN model architecture for spectral feature extraction in MSI data</i>	33
3.7	<i>2D–CNN model architecture for spatial feature extraction in MSI data</i>	34
3.8	<i>3D–CNN model architecture for spatial-spectral feature extraction in MSI data</i>	35
3.9	<i>4D–CNN model architecture for spatial-spectral feature extraction in time-series MSI data</i>	36
4.1	<i>Flow diagram illustrating the publication identification and screening process following the PRISMA guidelines (template is reused from Page et al. [101] with CC BY 4.0)</i>	42
4.2	<i>Distribution of publications</i>	43
4.3	<i>Heatmap of the most frequent journals of publications focusing on CNN applications in multispectral imagery</i>	43

4.4	<i>Heatmap of the most frequent conferences of publications focusing on CNN applications in multispectral imagery</i>	44
4.5	<i>Publication Trends Over Time for Each 1D, 2D, 3D and 4D-CNN</i>	45
4.6	<i>The word cloud of the most frequently used words in abstracts</i>	45
4.7	<i>Publication Distribution for each CNN by Domain</i>	46
4.8	<i>Publication Distribution for each CNN by Satellite</i>	46
4.9	<i>Satellite usage across different domains for 1D-CNN</i>	47
4.10	<i>Satellite usage across different domains for 2D-CNN</i>	48
4.11	<i>Satellite usage across different domains for 3D-CNN</i>	49
4.12	<i>Satellite usage across different domains for 4D-CNN</i>	49
4.13	<i>Number of publications using different machine learning techniques by each CNN</i>	50
4.14	<i>Accuracy of CNNs using classification across different domains</i>	51
4.15	<i>F₁-scores of CNNs using classification across different domains</i>	53
4.16	<i>Ontology visualization generated with WebVOWL</i>	63
5.1	<i>Map of the Adriatic Sea showing the locations of Secchi disk depth measurements (shown in Plate Carrée projection; adapted from [141])</i>	69
5.2	<i>Probability distribution of the measured Z_{SD} values in the Croatian, Slovenian, and Secchi Disk Project datasets (Adapted from [141])</i>	70
5.3	<i>Statistical characteristics of in situ Z_{SD} (m) through years</i>	70
5.4	<i>Cloud and land mask based on band Oa17 values for the area of the Kaštela Bay and Brač Channel (Croatian coast) on August 2, 2020 (Adapted from [141])</i>	73
5.5	<i>One dimensional data example – Sentinel-3 OLCI TOA reflectances corresponding to in situ Secchi values (Adapted from [141])</i>	74
5.6	<i>Proposed 1D-CNN architecture for the Secchi disk depth prediction (Adapted from [141])</i>	74
5.7	<i>Hyperparameters used in the 1D-CNN model (Adapted from [141])</i>	76
5.8	<i>Flowchart of the 1D-CNN model development and evaluation process (Adapted from [141])</i>	79
5.9	<i>A prediction error plot shows the difference between the predicted values and the measured values of Secchi depth in meters (m) on the final test dataset (Adapted from [141])</i>	82

5.10	<i>Histogram of errors of 1D-CNN model on the final test dataset (Adapted from [141])</i>	82
5.11	<i>The distribution of Secchi disk depth in the study area on September 3, 2021 (Adapted from [141])</i>	85
5.12	<i>The distribution of kd_z90max in the study region on September 3, 2021, created using the C2RCC processor in SNAP (Adapted from [141])</i>	85
5.13	<i>Comparison of Secchi disk depth predictions using in situ measurements, 1D-CNN model, and C2RCC processor based on data from Table 5.3 (Adapted from [141])</i>	87
6.1	<i>Study area in Split-Dalmatia County, Croatia (shown in Plate Carrée projection)</i>	94
6.2	<i>Monthly distribution of "road" and "no road" patches in Sentinel-2 dataset</i>	96
6.3	<i>Proposed 2D-CNN architecture for road detection</i>	96
6.4	<i>Loss and accuracy curves for the training and validation datasets over epochs for the 2D-CNN model used for road detection</i>	101
6.5	<i>2D-CNN model evaluation for road detection on train dataset</i>	101
6.6	<i>2D-CNN model evaluation for road detection on test dataset</i>	102
7.1	<i>Study area in Split-Dalmatia County, Croatia. Points represent small fire locations, while polygons represent larger fires from EFFIS database (shown in Plate Carrée projection)</i>	110
7.2	<i>Time differences between fire incidents and corresponding Sentinel-2 images (2017-2021). Left: EFFIS data; Right: Intervention data. Negative values indicate that images were taken before fire events.</i>	112
7.3	<i>Proposed 1D-CNN architecture for Fire Propagation Potential</i>	113
7.4	<i>Proposed 2D-CNN architecture for Fire Propagation Potential</i>	114
7.5	<i>Proposed 3D-CNN architecture for Fire Propagation Potential</i>	115
7.6	<i>Loss and accuracy curves for the training and validation datasets over epochs for the 3D-CNN model used for fire propagation potential</i>	116
7.7	<i>3D-CNN model evaluation for fire propagation potential prediction on train dataset</i>	117
7.8	<i>3D-CNN model evaluation for fire propagation potential prediction on test dataset</i>	118
7.9	<i>Fire propagation potential prediction map using 1D-CNN model</i>	120
7.10	<i>Fire propagation potential prediction map using 2D-CNN model</i>	121

7.11	<i>Fire propagation potential prediction map using 3D-CNN model</i>	121
8.1	<i>Study area representing fires in Split-Dalmatia County, Croatia, March 2022 (shown in Plate Carrée projection)</i>	128
8.2	<i>4D-CNN Architecture</i>	129
8.3	<i>Confusion matrices for the 4D-CNN model illustrating classification performance for various patch sizes in detecting burned and unburned areas</i>	131
8.4	<i>Performance metrics of the 4D-CNN model for various patch sizes on train and test datasets</i>	132
8.5	<i>ROC curves with AUC values for the 4D-CNN model on different patch sizes for both train and test datasets</i>	133
8.6	<i>Predicted burned area map using a 10×10 patch size</i>	135
8.7	<i>Predicted burned area map using a 32×32 patch size</i>	135
8.8	<i>Predicted burned area map using a 64×64 patch size</i>	136
9.1	<i>Study area in Split-Dalmatia County, Croatia (shown in Plate Carrée projection)</i>	143
9.2	<i>Unified CNN Architecture showing all model variants (1D/2D/3D/4D)</i>	145
9.3	<i>Comparison of True Land Cover Map with CNN Predictions</i>	148

List of Acronyms and Symbols

- 1D-CNN** - *One-Dimensional Convolutional Neural Network*
- 2D-CNN** - *Two-Dimensional Convolutional Neural Network*
- 3D-CNN** - *Three-Dimensional Convolutional Neural Network*
- 4D-CNN** - *Four-Dimensional Convolutional Neural Network*
- AI** - *Artificial Intelligence*
- ANN** - *Artificial Neural Networks*
- API** - *Application Programming Interface*
- AUC** - *Area Under the Curve*
- BAI** - *Burned Area Index*
- BOA** - *Bottom Of Atmosphere*
- CLC** - *CORINE Land Cover data*
- CNN** - *Convolutional Neural Network*
- CPU** - *Central Processing Unit*
- DBN** - *Deep Belief Network*
- DEM** - *Digital relief Height Model*
- DN** - *Digital Number*
- EFFIS** - *European Forest Fire Information System*
- ESA** - *European Space Agency*
- EU** - *European Union*
- FC** - *Fully Connected layer*
- FCN** - *Fully Convolutional Networks*
- FCNN** - *Feature Fusion-based Convolutional Neural Network*
- FPP** - *Fire Propagation Potential*
- GAN** - *Generative Adversarial Network*
- GAP** - *Global Average Pooling*
- GIS** - *Graphical Processing Unit*
- GPU** - *Geographic Information System*
- HSI** - *Hyperspectral Imaging*
- IDW** - *Inverse Distance Weighted interpolation*
- IoU** - *Intersection over Union*
- IR** - *Infrared Spectrum*
- LSTM** - *Long Short-Term Memory network*
- MAE** - *Mean Absolute Error*
- mIoU** - *mean Intersection over Union*
- MIR** - *Middle Infrared*
- ML** - *Machine Learning*

MLP - *Multi-Layer Perceptron*
MSE - *Mean Squared Error*
MSI - *Multi-Spectral Instrument*
NASA - *National Aeronautics and Space Administration*
NBR - *Normalized Burn Ratio*
NDVI - *Normalized Difference Vegetation Index*
NIR - *Near Infrared Spectrum*
NLP - *Natural Language Processing*
OLI - *Operational Land Imager*
OSM - *OpenStreetMap*
PRISMA - *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*
RAM - *Random-access memory*
R-CNN - *Region-based Convolutional Neural Network*
RF - *Random Forest algorithm*
RGB - *Red, Green and Blue*
RMSE - *Root Mean Squared Error*
RNN - *Recurrent Neural Network*
ROC - *Receiver Operating Characteristic curve*
RQ - *Research Question*
RRMSE - *Relative Root Mean Squared Error*
SAE - *Stacked Autoencoders*
SAR - *Synthetic Aperture Radar*
SNR - *Signal-to-Noise Ratio*
SST - *Sea Surface Temperature*
SVM - *Support Vector Machine algorithm*
SWIR - *Short-Wave Infrared*
TIR - *Thermal Infrared*
TIRS - *Thermal Infrared Sensor*
TOA - *Top of Atmosphere*
UAV - *Unmanned Aerial Vehicle*
USGS - *United States Geological Survey*
UV - *Ultra Violet Spectrum*
VGI - *Volunteered Geographic Information*
VIS - *Visible Spectrum*
Z_{SD} - *Secchi Depth*

1. INTRODUCTION

The expansion of space technologies and their influence on various industries has been evident since the beginning of the space age. Historical launch data from SpaceTrack's public catalog [1] reveals a large increase in the number of satellite launches since 1957, when the space age began. During the Cold War, the United States and Russia (the former Soviet Union) dominated space launches, with Russia peaking at nearly 100 annual launches in the 1980s, while the United States showed early dominance in the 1960s. Nowadays, we witness a large number of launches, particularly from the United States, reaching over 120 launches per year (2024), with multiple satellites being deployed in a single launch. China has also emerged as an important factor in the space industry, with more than 60 launches per year.

In the last two decades, the number of space companies has grown exponentially [2]. In addition to large independent agencies, such as the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA), private companies such as SpaceX [3] and PlanetLabs [4] are taking on a significant role in the space scene. SpaceTrack reports that the current number of active satellites has reached 10 700. According to astronomer Jonathan McDowell's data from November 20, 2024 [5], SpaceX alone operates 6 426 Starlink satellites in orbit, of which 6 371 are operational. Starlink satellites are designed to provide low-cost Internet and mobile phone services to remote locations.

The growing number of satellites produces huge amounts of data about the Earth, creating new opportunities for the development of various applications. Among them are the digital twins of the Earth that serve to monitor, predict, and assess changes in the environment [6]. Artificial intelligence (AI), which is currently increasingly popular and widely used, has emerged as a promising solution for handling large datasets. The volume, variety, and velocity of satellite data, also known as the "3V's of big data", exceed traditional processing capabilities and require a faster and more efficient way of analysis [7].

Development of parallel computer architecture, advanced AI algorithms, and increased data storage capabilities improved analysis of big datasets. GPU technology, which was originally made for computer graphics and the gaming industry, has become widely used in the processing of satellite images. For example, NVIDIA's CUDA architecture can use thousands of processing cores (GPUs) to simultaneously process different parts of satellite images, which significantly reduces the time required for complex operations [8].

Compared to traditional methods, AI algorithms have proven more suitable in remote

sensing applications due to their ability to identify complex patterns and handle non-linear relationships in data. However, in addition to efficiently processing huge amounts of data, AI algorithms must provide reliable and interpretable results that can influence important decisions (e.g., political) and scientific research. Consequently, there is an increasing emphasis on explainable AI, especially convolutional neural networks (CNNs) that form the backbone of many remote sensing applications. For instance, CNNs have shown promise in addressing diverse remote sensing challenges, such as feature extraction from satellite imagery and environmental monitoring. For now, CNNs are not fully explainable because it is still difficult to explain how a model manages to learn from features and achieve reliable predictions, which is why the CNN mechanism is often called a "black box". Therefore, it is necessary to work on the explainability of CNN models to ensure trust and reliability in their operations [9, 10].

This dissertation uses a methodological approach to address the challenges of artificial intelligence in remote sensing, with a particular focus on CNN models. The main focus is on the analysis of different degrees of CNN and how they affect the processing of remote sensing data. The research specifically studies the effects of different CNN architectures (1D, 2D, 3D and 4D) and determines for which tasks certain architectures are most suitable. The goal is to help scientists and experts to better understand the capabilities and limitations of each CNN degree in the field of remote sensing, which will allow a better selection of CNN architectures for specific task.

1.1. Research Hypothesis

This dissertation investigates the relationship between CNN architectures and their applications in remote sensing. The main hypothesis is the following:

- **(H0) selecting the appropriate degree of a convolutional neural network (CNN) based on the application domain and the characteristics of remote sensing data can result in a reliable model for monitoring Earth changes.**

To gain a deeper understanding, additional hypotheses have been proposed. Hypothesis **H1** focuses on a systematic review of publications related to using different CNN degrees in remote sensing and is defined as:

- **(H1) It is possible to gain insights into the suitability of 1D, 2D, 3D and 4D CNNs for addressing specific remote sensing problems.**

Based on a systematic review of existing works, it is possible to get insights into different CNN architectures applied to remote sensing data. Considering the large amount of remote sensing data, the review will be limited only to those publications that use multispectral data. Those results can be represented with ontology, which will help researchers to select an appropriate CNN architecture based on the application domain.

Hypotheses (H2)-(H5) focus on case studies for different CNN architectures and are defined as:

- **(H2) It is possible to use 1D-CNN on Sentinel-3 OLCI satellite data to accurately predict Secchi disk depth.**

The Secchi disk is a tool to measure water clarity and its depth is assumed to be a point-based parameter. The *spectral* features from satellite data are sufficient to achieve satisfactory results and can be successfully identified by 1D-CNN.

- **(H3) It is possible to detect roads by using a 2D-CNN model on Sentinel-2 satellite data.**

The roads are assumed to be objects in the spatial context of the observed area. Therefore, the *spatial* features that can be identified in satellite imagery should be sufficient to detect roads without depending on the spectral wavelength of each band.

- **(H4) It is possible to predict fire propagation potential (FPP) by using a 3D-CNN model on Sentinel-2 satellite data.**

By using a 3D-CNN model, it is possible to learn and process the complex *spatial-spectral* relationships within the satellite data. This enables the prediction of fire propagation potential, which is often influenced by various factors such as vegetation type, moisture content, and topography. These environmental factors are represented in the satellite data through their distinct spatial and spectral characteristics.

- **(H5) It is possible to perform land cover change detection by using a 4D-CNN model on multi-temporal Sentinel-2 satellite data.**

In particular, the assumption is that the model can detect the burned area after a fire event. Based on the *spatial-spectral-temporal* features of the satellite imagery, the model should be able to find spatial changes characterized by different spectral wavelengths (bands) through time.

Finally, hypothesis H6 is defined as:

- **(H6) By comparing 1D, 2D, 3D and 4D CNN architectures for land cover segmentation based on Sentinel-2 satellite data, a 3D-CNN architecture should prove to be the best model for this problem.**

Therefore, we hypothesize that a 3D-CNN architecture can better discriminate between land cover classes because of its ability to capture spatial patterns and spectral features simultaneously.

1.2. Dissertation Outline

The doctoral thesis has a total of ten chapters. The introductory chapter presents the motivation for this work, which lies in the development of the space industry and the increase in

the amount of satellite data that creates the need for more efficient processing using artificial intelligence. Also, this chapter defines the hypotheses of the doctoral thesis. The following one is Chapter 2, which provides brief theoretical foundations of remote sensing, with an emphasis on the electromagnetic spectrum, resolutions and types of sensors, basic data preprocessing techniques, and an overview of satellite missions important for this research. Building upon remote sensing fundamentals, Chapter 3 presents theoretical aspects of deep learning in the context of remote sensing. This chapter describes image analysis techniques used in remote sensing, CNN key concepts, and most commonly used layers in CNN architectures. Also, it includes important steps in the CNN training process and provides the main concepts of 1D, 2D, 3D and 4D-CNNs. This is followed by chapters that present the main research, contributions and results of this doctoral thesis.

Chapter 4 provides a systematic review of the available literature on the use of 1D, 2D, 3D and 4D-CNN architectures on multispectral images. For this purpose, the PRISMA methodology was used, which provides a framework for objective literature analysis. The chapter ends with an ontology and conclusions drawn from an extensive analysis of studies on the applicability of individual CNN architectures for a specific task domain. The following chapters present specific contributions of 1D, 2D, 3D, and 4D-CNN, which are quantitatively and qualitatively evaluated. Chapter 5 explores the application of 1D-CNN for estimating Secchi disk depth from Sentinel-3 OLCI data. The 1D-CNN architecture's contribution lies in effectively processing *spectral* information for a *regression* task. Chapter 6 demonstrates the application of 2D-CNN for road detection using Sentinel-2 imagery and OpenStreetMap data. The 2D-CNN architecture's contribution is in effectively processing *spatial* information for a *classification* task. Chapter 7 presents the application of 3D-CNN for predicting fire propagation potential using Sentinel-2 imagery. The 3D-CNN architecture's contribution lies in effectively processing both *spatial* and *spectral* information for a *classification* task. The chapter also provides comparative analysis with 1D-CNN and 2D-CNN models to demonstrate the advantages of 3D architecture in this application. As the final CNN case study, Chapter 8 demonstrates the application of 4D-CNN for burned area detection using multi-temporal Sentinel-2 imagery. The 4D-CNN architecture's contribution lies in effectively processing *spatial-spectral-temporal* information for a *scene understanding* problem, with particular emphasis on analyzing the impact of different spatial patch sizes on model performance. As a final contribution of this thesis, Chapter 9 presents a comparative analysis of all four CNN architectures on a common land cover *semantic segmentation* task using CORINE Land Cover data, providing both quantitative and qualitative assessment of their performance.

The thesis concludes with Chapter 10, which summarizes all scientific findings and contributions, and outlines potential directions for future research.

2. REMOTE SENSING

This chapter presents the theoretical aspects of remote sensing as a basis for further discussion. The first subchapter describes the electromagnetic spectrum, its ranges and the interactions of radiant energy with the Earth's surface in the context of remote sensing imagery. The introduction is followed by an overview of spatial, spectral, radiometric and temporal resolutions and how these affect the images acquired by remote sensing. The types of sensors are then described and their categorization into active and passive sensors based on the principles they use to detect electromagnetic energy. The next subchapter gives a brief overview of the basic satellite missions and their characteristics. The chapter concludes with a description of data preprocessing techniques related to the digital analysis of remote sensing data.

2.1. The Electromagnetic Spectrum

Remote sensing is the acquisition of information about the Earth's land and water surfaces by sensors installed on aircraft or satellites. These sensors receive electromagnetic energy that is emitted or reflected by the observed surface [11]. Electromagnetic energy captured by sensor can be characterized by its frequency or wavelength. In the field of remote sensing, it is common to define regions of the spectrum based on wavelength, often using micrometers ($1 \times 10^{-6}\text{m}$, μm) and forming the electromagnetic spectrum. Figure 2.1 shows the electromagnetic spectrum divided into discrete regions of continuous wavelengths called *wavebands* or *bands*. The visible region consists of a very small portion of the spectrum, only 0.4 to 0.7 μm and can be sensed by human eyes. It can be used for land cover mapping. This region can be further divided into three primary colors: blue (0.4 to 0.5 μm), green (0.5 to 0.6 μm), and red (0.6 to 0.7 μm) band. Besides the visible region, there are several regions of the electromagnetic spectrum of interest in remote sensing such as the ultraviolet (UV) spectrum (0.30 to 0.38 μm), the infrared (IR) spectrum (0.72 to 15 μm) and microwaves ($> 1 \text{ mm}$) [12]. The UV region covers the shortest wavelengths that are of practical use for Earth observation because it is largely scattered by the Earth's atmosphere. Visible UV radiation is emitted, for example, only by rocks and minerals. The infrared spectrum can be divided into the near-infrared (NIR) region, which is relevant for discriminating green vegetation and its health; the mid-infrared (MIR) region, which is useful for estimating soil and vegetation moisture contents and detecting high-temperature sources; and the thermal infrared (TIR)

region, which is commonly used to map surface temperatures. The microwave region has very long wavelengths that are not affected by atmospheric conditions (clouds, smoke, dust). Additionally, microwaves can penetrate to various depths of forest canopies and can be used in soil moisture and surface roughness analyses. Usually, a remote sensing instrument is designed to operate in one or more bands taking into account the characteristics of the study area [13].

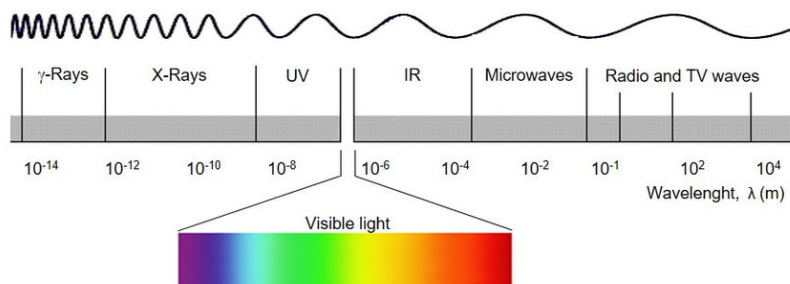


Figure 2.1. Electromagnetic spectrum divided into bands (Adapted from [14])

Electromagnetic waves that interact with the Earth's atmosphere and surface can be reflected from its surface, absorbed and/or transmitted through its surface. Reflection, absorption and transmission will depend on the nature of the surface (material type and condition), the energy's wavelength and the angle of illumination. The relationship among these interactions can be defined using the principle of energy conservation [15]:

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \quad (2.1)$$

where λ represents the wavelength, E_I is the incident energy, E_R is the reflected energy, E_A is the absorbed energy and E_T is the transmitted energy. The reflectance characteristics of the Earth's surface can be quantified by measuring the proportion of incident energy and reflected energy often called Top-of-Atmosphere (TOA) radiance. It is defined as the relative brightness of a surface measured within a specific wavelength interval and is referred to as spectral reflectance $R_{rs}(\lambda) = \frac{E_R(\lambda)}{E_I(\lambda)}$ commonly expressed as a percentage [16]. Transmission refers to the radiation passing through a medium without significant attenuation. A measure of the medium's ability to transmit energy is called transmittance. Absorption energy loss is caused by atmosphere which prevents or absorbs energy at a given wavelength. The most efficient absorbers of solar radiation are water vapor, carbon dioxide and ozone [12].

Figure 2.2 shows the path of solar radiation traveling at the speed of light (3×10^8 m/s), as it reaches the satellite's sensors. It can be noted that the incident solar radiation must pass through the atmosphere twice. It is affected by the permeability of the atmosphere, which depends on the physical properties of gases and the number of suspended particles and several physical processes (scattering, absorption and refraction).

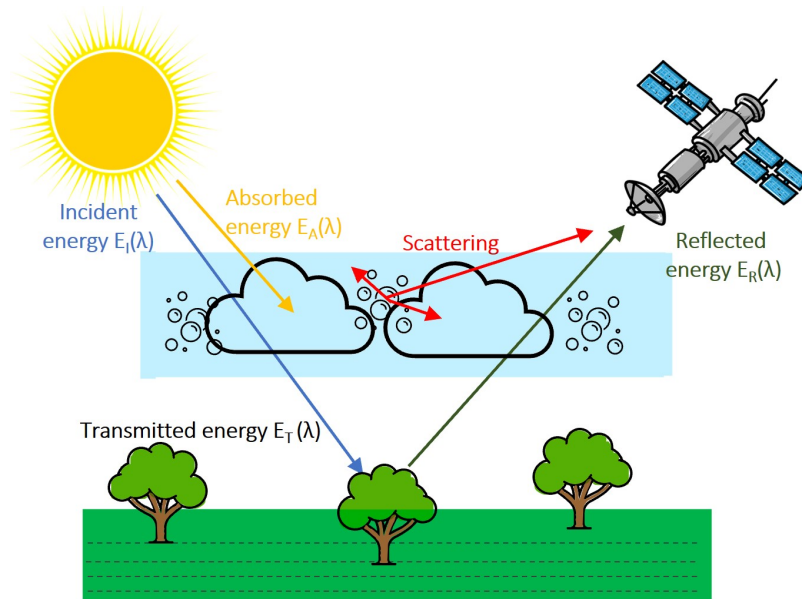


Figure 2.2. Interaction between solar energy and the Earth's surface

2.2. Sensor Resolution

Onboard sensors in remote sensing are capable of snapping larger area at a time with a predefined resolution. The resolution of a sensor refers to its ability to distinguish information, i.e. how well it records the fine details of the observed area [17]. It is important to consider the resolution of the sensor as an overall system rather than focusing on its individual components. Image resolution also depends on ground features, atmospheric conditions, lighting and the expertise of the image interpreter. In the scope of remote sensing image resolution can be categorized into spatial, spectral, radiometric and temporal resolution [12, 13, 18]. Each type of resolution is described in more detail in the following subsections.

2.2.1. Spatial Resolution

Spatial resolution refers to the smallest object that can be resolved by the sensor, or the size of the area used for the sensor's instantaneous field of view (IFOV). The IFOV is defined as the angular section observed by the sensor from which the sensor receives the energy at a given time. This means that spatial resolution depends on sensor precision and satellite distance from the Earth. The most common measure of the spatial resolution of remote sensing data is the pixel size, which is usually expressed in metres. Figure 2.3 shows the scene of the campus of the University of Split in Croatia for four different spatial resolutions (1 m, 10 m, 30 m and 100 m). It can be seen that the larger the pixel, the lower the spatial resolution.

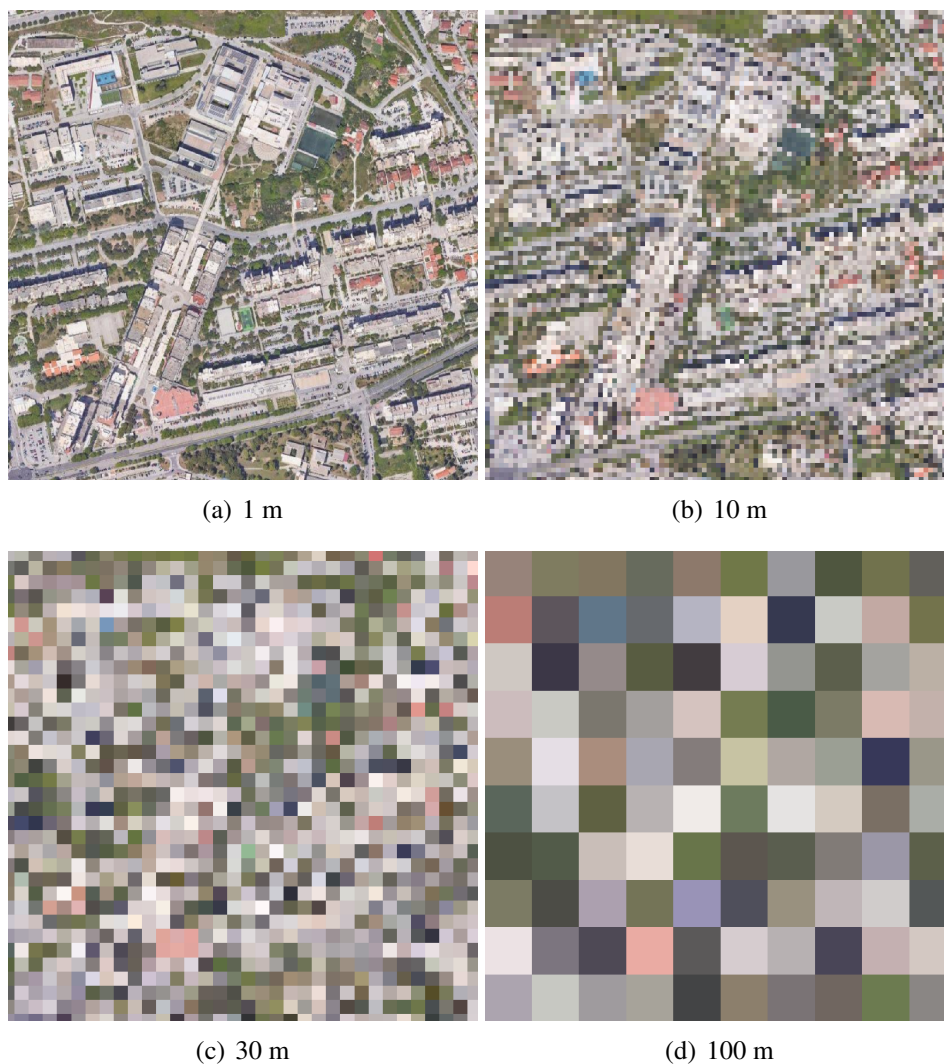


Figure 2.3. Campus of the University of Split in four different spatial resolution

2.2.2. Spectral Resolution

Spectral resolution refers to the ability of a sensor to distinguish finer wavelengths. It is determined by the number of bands and their spectral bandwidths. If a sensor has larger number and narrower range of spectral bands, it can better distinguish different features. Earth observation satellites equipped with multispectral sensors offer a wide range of bands, usually a panchromatic (PAN) band and bands in the visible–near–IR or thermal–IR spectrum. Sensors with high spectral resolution are commonly referred to as hyperspectral sensors. They usually have a better detection capability as they contain hundreds of spectral bands in continuous ranges. Figure 2.4 shows an example of spectral resolution for the scene on the campus of the University of Split. It can be seen that a grayscale image with one band does not provide much information, while an image with three bands, namely red, green and blue (RGB), can distinguish different objects in the scene (e.g. roofs, vegetation and playgrounds).

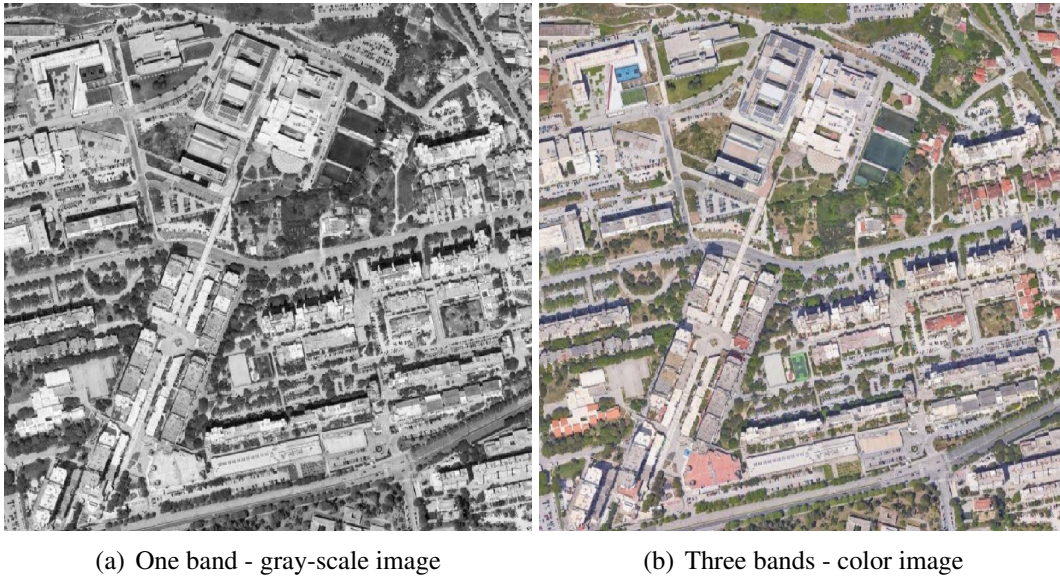


Figure 2.4. Examples of spectral resolution for the campus of the University of Split

2.2.3. Radiometric Resolution

Radiometric resolution refers to the sensitivity of the sensor, i.e. its ability to measure and distinguish electromagnetic energy represented by different radiation intensities within the same spectral band. It is defined as the range of values digitally encoded by the sensor. The number of bits per pixel is used to store the radiance as digital numbers (DN). Figure 2.5 shows the radiometric resolution for the University of Split campus in an 8-bit image, which has 2^8 or 256 possible values per pixel (i.e. 0–255), and in a 2-bit image, which has only 2^2 or 4 possible values per pixel (i.e. 0–3). It is noticeable that some details are missing in the 2-bit image. Thus, when the radiometric resolution of a sensor is higher, it is more sensitive to detecting small differences in reflected or emitted energy.

2.2.4. Temporal Resolution

Temporal resolution is a measure that refers to the frequency with which the sensor revisits the same part of the Earth's surface. Usually, days are used as the unit of temporal resolution for most satellites. Satellites with frequent revisits are considered to have a high temporal resolution. In addition, the temporal resolution can be influenced by the orbital characteristics of the satellite, the field of view of the sensor and the atmospheric conditions, as some sensors cannot detect the surface below the clouds. Furthermore, the temporal resolution depends on the targets that the sensor is tracking. For example, in the case of data collected by a meteorological satellite sensor, the data should be updated at short time intervals (e.g. 15 to 30 minutes) to take into account the dynamics of the observed phenomena.

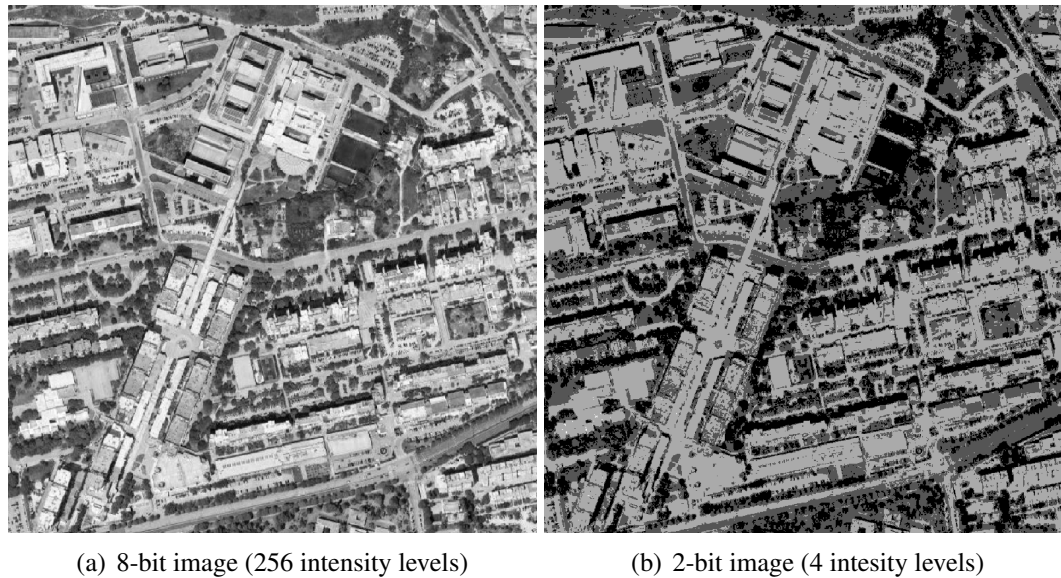


Figure 2.5. Examples of radiometric resolution for the campus of the University of Split

2.3. Sensor Types

The classification of sensors in remote sensing is generally based on their functional principle into active and passive sensors (Figure 2.6). Active sensors can provide an energy source, i.e. generate an energy pulse in the direction of the target of interest and detect the reflection of the observed target. Passive sensors can measure electromagnetic radiation originating from an external source, e.g. energy reflected or emitted from solar radiation, the Earth's surface or the atmosphere. There are different types of remote sensing images for the sensor types used, such as panchromatic, multispectral, hyperspectral and Synthetic Aperture Radar (SAR) images [19,20]. This subsection describes the characteristics of some commonly used instruments for active and passive sensors.

2.3.1. Active Sensors

Active sensors usually operate in the microwave part of the electromagnetic spectrum, which allows them to measure the reflection of the observed area without the influence of the atmosphere and solar light.

Radar

Radar or "Radio Detection and Ranging" is the most commonly used active sensor. It emits electromagnetic energy in the wavelength range from 0.1 cm to 1 m and detects the reflections scattered back from the observed area, which provide information about the target. The pixel coefficient in the radar image represents the backscatter coefficient of the target, and its value increases with the strength of the received signal. With a sufficiently high resolution,

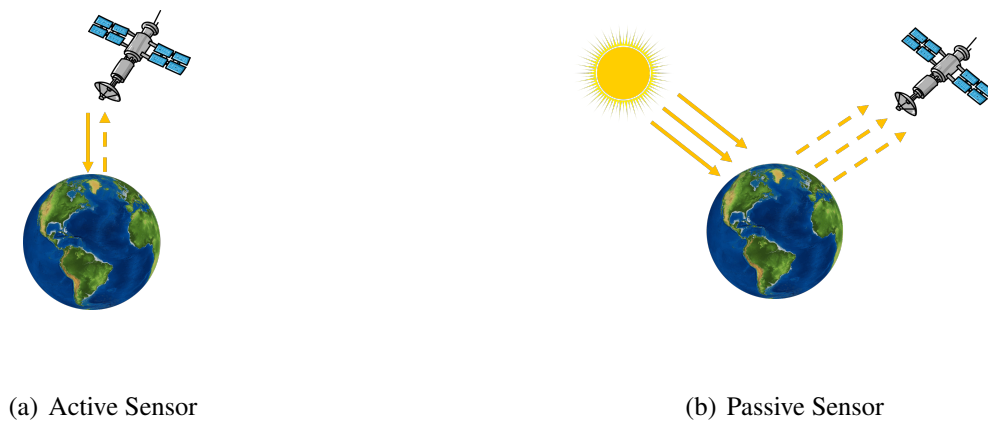


Figure 2.6. Sensor Types

the radar can recognize the size and shape of the target. However, to increase the observation altitude of the object and achieve sufficient spatial resolution, antennas of enormous size are required, which is a significant disadvantage of this approach. This problem can be solved by using virtual antennas that synthesize "longer" antennas, a technique called Synthetic Aperture Radar (SAR) [13,21].

SAR

SAR or "Synthetic Aperture Radar" is a radar system that has a synthetic aperture and is based on "augmenting" the length of the antenna, not in the physical sense, but by using the relative forward motion of a short antenna towards the target area and the Doppler effect. The Doppler effect is the change in wave frequency as a function of the relative velocities of the transmitter and the reflector. SAR uses the Doppler information to calculate the frequency shift to determine the position and scattering characteristics of the target. Signals that are backscattered over a certain time interval during the recording of an area are stored and subsequently processed so that the recorded elements can be distinguished and an image of the observed area can be reconstructed [11,22].

LiDAR

LiDAR or "Light Detection And Ranging", uses a laser to measure the distance between an object and the scanner using coherent light, and a mechanical optical mount to precisely scan the laser across the scene. The scanner also needs to know where it is in order to know the position of features in the world. In addition to estimating the brightness of the backscatter, the LiDAR sensor can also measure the angular position, frequency change and time of the reflected pulses. One of the advantages of LiDAR sensors is their speed, which can detect

tens of thousands to more than a million positions per second [12, 23].

2.3.2. Passive Sensors

Passive sensors used in remote sensing usually operate in the visible, infrared, thermal infrared and microwave regions of the electromagnetic spectrum.

Radiometer

A radiometer is a sensor that measures electromagnetic radiation with high radiometric resolution in the visible, infrared, or microwave regions of the electromagnetic spectrum [18]. It quantitatively measures the intensity of electromagnetic radiation in one broad spectral band (*single-band radiometer*) or in only a few bands (*multi-band radiometer*) [24].

Hyperspectral radiometer

A hyperspectral radiometer captures the reflected energy in hundreds to thousands of continuous narrow spectral bands in the visible, near-infrared and mid-infrared regions of the electromagnetic spectrum [25]. The high spectral resolution of the sensor makes it possible to distinguish details of the examined objects in a specific range. This is achieved on the basis of the spectral response of the observed objects in each of the narrow bands of the electromagnetic spectrum [26].

Spectrometer

A spectrometer is a sensor used for detecting, measuring, and analyzing the spectral content of incident electromagnetic radiation [26]. Four characteristics used to measure the spectral response of a spectrometer are: spectral range, spectral sampling interval, spectral bandwidth, and signal-to-noise ratio (SNR) [27]. The spectrometer measures the radiance in many narrow spectral bands, often from the visible (VIS) part to the short-wave infrared (SWIR) part of the electromagnetic spectrum. This sensor is characterized by high spectral resolution but a low radiometric resolution compared to a radiometer [24].

2.4. Data Preprocessing Techniques

Most of the data collected through remote sensing of the Earth must undergo basic processing from the moment the images are received on Earth to the moment the user loads them into their computer and they become usable. In the context of digital analysis of remote sensing data, two main image preprocessing operations are performed, namely radiometric and geometric image processing. In addition to these operations, this subsection describes the

common preprocessing steps in digital image processing, including atmospheric correction, image enhancement and data fusion.

2.4.1. Radiometric Preprocessing

Radiometric calibration removes radiometric distortions in remote sensing images to ensure the most accurate representation of ground conditions. It converts recorded voltages or digital numbers (DN) into absolute values of radiance or reflectance. The radiance values of the image are adjusted to eliminate distortions caused by sensor failures or atmospheric interference. When observing the Earth's surface with visible or near-visible radiation, each sensor records a mixture of both types of brightness. The brightness that is of interest for remote sensing is the brightness derived from the Earth's surface, and the brightness that "interferes" with the image brightness value is the brightness of the atmosphere itself. For example, a digital radiance value of "62" may result partly from a surface reflectance of "50" and partly from atmospheric scattering of "12". The above-mentioned brightnesses are difficult to distinguish from each other. Therefore, it is necessary to identify and separate them using atmospheric correction so that the main analysis can focus on examining the exact surface of the brightness. This means that the DNs need to be converted into actual surface brightness values so that consistent and accurate measurements can be made to detect changes in climate characteristics and environmental conditions [12, 28].

2.4.2. Atmospheric Correction

As mentioned in the Section 2.4.1, any sensor that observes the Earth's surface in the visible or near-visible part of the electromagnetic spectrum detects two types of radiation, one of which is radiation from the atmosphere itself. Therefore, the focus of atmospheric correction is on images taken in the visible and near-infrared parts of the spectrum, as radar images are not affected by clouds, precipitation and other atmospheric conditions. Clouds in the atmosphere block information about the earth's surface, so different methods are used for their detection and their corresponding shadows. Mainly the threshold method, atmospheric radiation models and different statistical methods are used to create a cloud mask of the observed area. In addition to clouds, the quality of optical images is also influenced by aerosols and water vapor, which scatter and absorb the radiation reflected from the surface. The aerosol distribution mainly affects short-wave signals, while water vapor affects near-infrared signals. According to the authors of [18], atmospheric correction mainly consists of two parts: 1) estimation of atmospheric parameters and 2) determination of surface reflectance. Figure 2.7 shows the atmospheric correction of a Sentinel-2 satellite image with the Sen2Cor processor.

Many atmospheric correction algorithms are developed for a specific satellite. For example, the Sen2Cor processor developed by European Space Agency (ESA) is used for

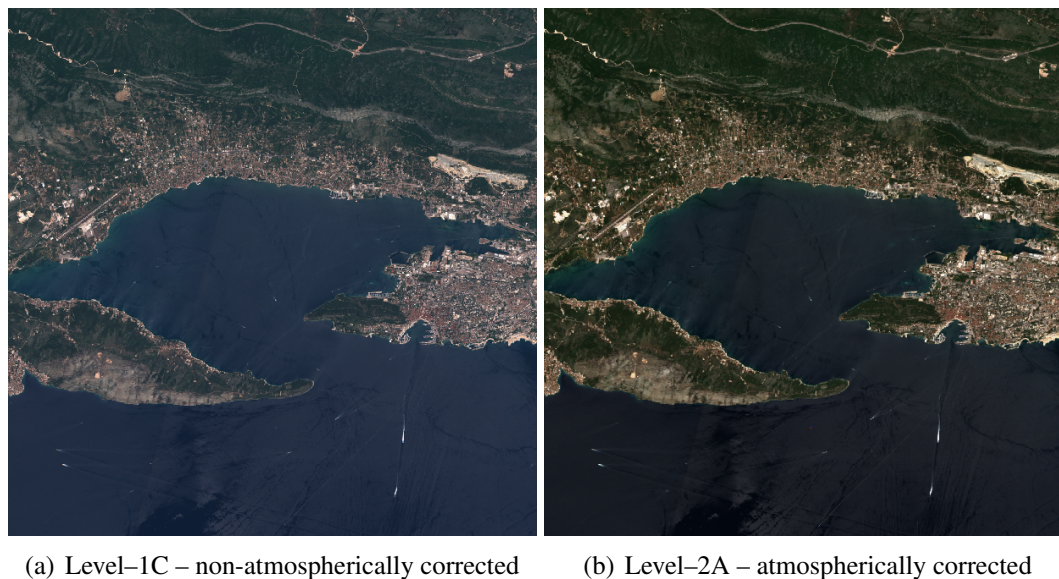


Figure 2.7. Example of atmospheric correction on Sentinel-2 imagery, credits: "Modified Copernicus Sentinel data 2024/Sentinel Hub" [29]

the atmospheric correction of Sentinel-2 satellite images. The aim of the Sen2Cor processor is to correct the Sentinel-2 Level-1C Top-of-Atmosphere (TOA) products from atmospheric effects to obtain a Level-2A Bottom-of-Atmosphere (BOA) product. To detect clouds, Sen2Cor uses the Scene Classification (SCL) module, which uses a set of spectral reflectance thresholds, ratios and indices such as the Normalised Difference Snow Index (NDSI) and the Normalised Difference Vegetation Index (NDVI) for each pixel and all spectral bands except B06, B07 and B09. Sen2Cor performs the atmospheric correction using a set of so-called look-up tables (LUT) generated by the library *libRadtran* for the calculation of solar and thermal radiation in the Earth's atmosphere [30]. There are atmospheric correction processors intended for the correction of images from different satellites. For example, the MACCS-ATCOR Joint Algorithm (MAJA) processor is used for the atmospheric correction of images from the Formosat-2, Landsat, VEN μ S and Sentinel-2 satellites. MAJA is a spectral-temporal method consisting of two components, namely the Multi-sensor Atmospheric Correction and Cloud Screening (MACCS) algorithm and modules of the Atmospheric and Topographic Correction (ATCOR) software. To detect low clouds, the MAJA processor uses a composite reference image containing pixels without clouds to check whether the blue and red spectral bands of the observed image have exceeded the defined threshold and whether there is a low correlation between the reflections of neighboring pixels [31]. One of the best known processors for the purpose of atmospheric correction of water images in land and coastal areas is the Case 2 Regional CoastColour (C2RCC) processor. The C2RCC algorithm is based on a series of neural networks trained on a simulated data set with TOA reflections. The mentioned algorithm generates Case-2 products containing inherent optical properties, absorption and scattering of different particles as well as

relevant optical concentrations (e.g. phytoplankton pigment, total suspended solids and yellows). It is most commonly used with images acquired by the Sentinel-3 OLCI, Sentinel-2 MSI, Landsat-8 OLI and MERIS sensors [32].

In addition to the algorithms developed for the target satellite, there are those developed specifically for atmospheric correction of satellites recording in a specific range of the electromagnetic spectrum, such as MODerate resolution atmospheric TRANsmission (MODTRAN), the 6S model [33] and Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [34]. MODTRAN is one of the better known and widely used atmospheric correction algorithms. It was developed for modeling atmospheric propagation in the range from the ultraviolet to the infrared part of the electromagnetic spectrum. MODTRAN solves the radiative transfer equation, i.e. the phenomenon of energy transfer in the form of electromagnetic radiation, which includes the effects of absorption, emission and scattering of molecules and particles, surface reflections and emissions, solar or lunar illumination and spherical refraction [35].

Atmospheric correction is certainly one of the most important steps in the preprocessing of remote sensing images. This chapter gives a brief overview of the commonly used atmospheric correction algorithms. It is important to emphasise that there is still no uniform algorithm for all remote sensing images. Therefore, this area is still an interesting topic for many researchers, considering that even the smallest failure in the algorithm can lead to incorrect results, which in turn leads to poor performance of machine and deep learning algorithms in classifying and recognising certain objects and phenomena in images.

2.4.3. Geometric Preprocessing

Geometric processing means the creation of planimetrically correct versions of images obtained by remote sensing. No image captured by a sensor can fully reflect the true characteristics of the landscape. In addition, many factors such as variations in platform height, position, velocity, rotation and curvature of the Earth can distort the geometric properties of the data. Some of the mentioned distortions are systematic and can be corrected by analyzing the properties of the sensor or the platform, while others are random and must be corrected by using ground control points (GCPs). GCPs represent locations in the input image that can be precisely located on the ground and on planimetrically accurate maps. The goal of geometric processing is to match remote sensing images with other images and maps to make them suitable for measuring distance and area [12, 18].

2.4.4. Image Enhancement

Image enhancement is the process of emphasizing or improving the visual appearance and characteristics of digital images, such as edges or contrast. Enhancement does not increase embedded information and data capacity and is often only successful for a specific purpose

and may be inappropriate for another image or purpose. In addition, image enhancement enables the identification of real and virtual objects that are visible on satellite images but sometimes do not appear, e.g. due to the shadows cast by trees, buildings, mountains and the like. Image enhancement techniques can be divided into three broad categories: spatial domain methods, frequency domain methods and color enhancement [12, 36].

2.4.5. Data Fusion

Data fusion refers to processes in which images of a particular scene with different resolutions, originating from two or more sensors, are combined to create an interpretation of the scene that cannot be obtained from a single sensor. Thus, the limitation of one sensor can be overcome by combining images taken at the same or at a different time, with different spectrum, radiometric calibration or spatial resolution, depending on the characteristics of the individual sensors with which the images were taken [37]. According to Pohl and Van Genderen (1998) [38], the fusion of remote sensing data can be divided into three different levels: pixel level, feature level and decision level. Pixel-level data fusion aims to improve the spatial information in multispectral images by using high resolution panchromatic images. In feature-level data fusion, different features (e.g. edges, lines) are extracted from different data sources and then fused into one or more feature shapes to replace the original data. In decision-level data fusion or interpretation-level fusion, the input images are processed individually for information extraction to generate feature vectors, which are then subjected to pattern recognition to complete the feature description of the targets.

2.4.6. Processing Levels

Most remote sensing data requires at least minimal processing before it can be used. Various data providers have adopted a common set of processing levels to describe the type of image processing and to enable users to more easily select the appropriate level for their project. The processing levels are hierarchical, meaning that higher-level data starts with the processing involved in the images of the previous level and adds other features to them.

Raw and corrected remote sensing images can be classified into data products at different levels, namely [39, 40]:

- **Level-0** – represents raw instrument and payload data, just as it was collected at the sensor at full resolution, with any or all communication artifacts removed (e.g. duplicate data or communications headers). Since some basic data processing needs to be applied before it can be used, this kind of data will generally not be distributed to users, unless the user is studying the sensor device and not necessarily the features of the Earth.

- **Level-1A** – represents data over which radiometric correction has been applied and detector variations within the sensor have been removed. The data are raw and displayed in full resolution, time-referenced and marked with auxiliary information containing radiometric and geometric calibration coefficients.
- **Level-1B** – represents images over which corrections have been applied for variations such as distortions in image geometry caused by scan line misalignment or non-uniform pixel sizes. The applied corrections improve the geometric quality of the image, which is highly desirable for their further use and processing.
- **Level-2A** – represents enhanced versions of Level-1B images that are systematically mapped to a standard cartographic map projection based on a prediction of where the satellite was when the image was taken. With some satellite data, such as Sentinel-2 satellite images, this level of processing also includes scene classification and atmospheric image correction.
- **Level-2B** – presents precisely georeferenced images, which undergo a process of geometric correction or image correction, the image analyst positions the image on the existing base map by selecting pairs of well-defined GCP points from the observed image and the base map. The positional accuracy of a Level-2B image generally corresponds to the spatial resolution of the source data (e.g. 30 m for Landsat images).
- **Level-3A** – useful in the case of images that represent high relief, so in addition to manually locating GCP points, it is necessary to provide a digital relief height model (DEM) in order to take into account relief displacement at different heights. This process is generally called orthorectification.
- **Level-3B** – represents images with the same attributes as Level-3A scenes, but covering a larger area, i.e. scenes that are combined into a mosaic.

2.5. Overview of Satellite Missions

Satellite missions consist of one or more satellites that are launched into space by rockets and can orbit the Earth for 5 to 12 years in a predefined orbit [24]. Common critical requirements that every satellite mission should fulfill are: coverage or response time, resolution, sensitivity and lifetime in orbit. The coverage of the satellite mission refers, for example, to the number of satellites used, the communication architecture, the field of view of the useful terrain and the altitude. The resolution depends on the size of the instruments, the altitude and the position control. Sensitivity relates to payload size, complexity, processing and thermal control. The lifetime of a satellite mission in orbit depends on weight, component selection, power budget and propulsion [41].

According to Space-Track.org's public catalog (as of 16 November 2024), there are currently more than 10,700 active satellites orbiting Earth among a total of 47,000 tracked objects in space [1]. Given the number of active satellites in orbit, this thesis only provides an overview of the satellites that are of interest for future research and whose data is publicly available. As part of the ESA Copernicus program, the European Space Agency (ESA) has developed a "family" of satellites called Sentinel, which consists of a total of seven satellite missions with radar and multispectral instruments to detect and monitor land, oceans and the atmosphere [42]. This section describes the theoretical aspects of the radar satellite mission Sentinel-1 and the two multispectral satellite missions Sentinel-2 and Sentinel-3.

The Sentinel-1 mission currently operates with Sentinel-1A (launched 2014) using synthetic aperture radar (SAR) in the C-band spectrum. Following Sentinel-1B's mission end in December 2021, Sentinel-1C is scheduled for launch on December 3, 2024, with Sentinel-1D planned for late 2025 to replace Sentinel-1A [43]. The two main products of Sentinel-1 are Single-Look Complex (SLC) images and Ground Range Detection (GRD) images. SLC images consist of focused SAR data containing the full phase and associated complex data. These images are georeferenced using the satellite orbit and position data at the time of acquisition and corrected for terrain obliquity. GRD images contain generalized focused SAR data projected into the WGS84/UTM coordinate system using an ellipsoidal model. These images have approximately square pixels and less noise, but lower spatial resolution than SLC images. The advantage of radar is the ability to perform remote sensing and observation regardless of atmospheric conditions, as the microwaves penetrate through fog, rain and clouds [44].

The Sentinel-2 mission uses the MSI (Multispectral Imager) instrument mounted on two identical satellites, Sentinel-2A and Sentinel-2B, to monitor the Earth's surface. The satellite captures images in 13 broad bands, ranging from visible (VIS), near-infrared (NIR), vegetation red edge to short-wave infrared (SWIR) bands. The spatial resolution is 10 m, 20 m and 60 m, depending on the spectral band. The two satellites are in the same sun-synchronous orbit, but on diametrically opposite sides, ensuring a revisit time of five days in the equatorial region. The satellite data from the Sentinel-2 mission can be used, for example, for monitoring forests and vegetation, water management, climate change, risk management (e.g. floods and forest fires) or urban mapping [45, 46].

The Sentinel-3 mission uses several instruments for global environmental monitoring, including two important optical instruments: the Ocean and Land Color Instrument (OLCI) and the Sea and Land Surface Temperature Radiometer (SLSTR). The OLCI is a sensitive optical instrument consisting of 21 spectral bands designed to capture the color of ocean and land surfaces. The SLSTR instrument measures the surface temperature of oceans, land and ice. It collects data in nine spectral bands, three in the visible and near-infrared regions, three in the short-wave infrared region, one in the mid-infrared region, and two in the thermal infrared region. The spatial resolution of Sentinel-3 satellite is 300 m, 500 m or 1 km,

depending on the spectral band. In addition, two special channels enable the detection of fire. Two identical satellites, Sentinel-3A and Sentinel-3B, are currently active, which are in the same orbit with a phase delay of 180 degrees. In this way, the mission enables a high degree of availability of data products with a revisit time of less than two days at the equator [47].

Similar to the ESA, the US government's National Aeronautics and Space Administration (NASA) also operates numerous satellite missions for Earth observation. In the following text, the Landsat-8 and MODIS satellite missions are described.

The Landsat-8 satellite was developed in cooperation between NASA and the US Geological Survey (USGS). It has two instruments: Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). The OLI instrument records in nine spectral bands in the visible (VIS), near-infrared (NIR) and short-wave infrared (SWIR) spectral regions as well as in a panchromatic band. Panchromatic and multispectral images are recorded with a resolution of 15 meters and 30 meters respectively. This instrument can be used for a variety of applications such as cirrus cloud detection, coastal aerosol monitoring, active fires and water purity measurement. The TIRS instrument uses two thermal bands in the thermal infrared region and the collected data has a spatial resolution of 100 meters. This sensor makes it possible to measure the temperature of the Earth's surface [48].

MODIS (Moderate Resolution Imaging Spectroradiometer) serves as the primary instrument on NASA's Terra and Aqua satellites, which continue to operate in extended mission phases. Terra maintains observations despite transitioning to reduced operations in 2023, while Aqua operates in 'free-drift' mode since December 2021 with expected operations until September 2026. MODIS is a radiometer with 36 bands measuring in visible and infrared region of electromagnetic spectrum in the wavelength range 0.4 to 14.5 μm . The spatial resolution of Aqua and Terra MODIS is 250 m, 500 m and 1 km, depending on the spectral band. Its data is used to create a wide range of products for the ocean (e.g., chlorophyll fluorescence, suspended solids, and organic matter concentrations), land (e.g., land cover type, various vegetation indices, fires), and atmosphere (e.g., cloud mask, cloud optical thickness). The missions' complementary orbits (Terra in morning, Aqua in afternoon) enable twice-daily global coverage [49–51].

Table 2.1 provides an overview of the satellites described above, including Envisat (no longer operational) [52], PROBA-V (ended in 2020) [53], SPOT 6/7 and WorldView-2/3. It should be noted that SPOT and WorldView satellites are commercial satellites. SPOT is owned by Airbus Defence and Space, but has an agreement with ESA for the distribution of the data products obtained from these missions, while the WorldView satellites are owned by DigitalGlobe. The SPOT satellites are identical and each of them is equipped with two identical New Astrosat Optical Modular Instruments (NAOMI) [54]. Both WorldView satellites have panchromatic (PAN) and multispectral (MS) sensors, while WorldView-3 has two additional SWIR and CAVIS (Clouds, Aerosols, Vapours, Ice, and Snow) sensors [55, 56].

Table 2.1. Overview of satellites for remote sensing and their basic parameters

Agency	Satellite	Sensor/ Instrument	Spatial resolution	Temporal resolution (day)	Spectral resolution (bands)
Airbus Defence and Space	SPOT 6 and SPOT 7	NAOMI	1.5m and 8m	twice a day	5 (PAN, RGB, NIR)
ESA	Sentinel-1	SAR	10-40m	12	C-Band
	Sentinel-2	MSI	10m, 20m and 60m	≤ 5	13 (VIS, NIR, SWIR)
	Sentinel-3	OLCI	300m, 500m	< 2	21 (VIS, NIR)
		SLSTR	and 1km	< 4	11 (VIS, SWIR, MWIR, TIR)
	Envisat	MERIS	300m	≤ 3	15 (VIS, NIR)
	PROBA-V	VGT-P	100m, 333m, and 1km	≤ 2	4 (SWIR, Blue, Red, NIR)
NASA	Landsat-8	OLI, TIRS	15m, 30m, 100m	16	11 (VIS, NIR, SWIR, PAN, TIR)
	Aqua	MODIS	250m, 500m, and 1km	daily; reduced operations	36 (VIS, NIR, SWIR, MWIR, LWIR)
	Terra				
DigitalGlobe	WorldView-2	NAOMI	PAN: 0.46 m MS: 1.8m	up to 1.1 days	9 (VIS, NIR, PAN)
	WorldView- 3		PAN, MS, SWIR: < 4 m CAVIS: 30m	daily	PAN: 1, VIS: 8, SWIR: 8, CAVIS: 12

3. DEEP LEARNING

Deep learning is a subfield of machine learning within artificial intelligence that focuses on creating deep neural network models capable of making accurate data-driven decisions. These decisions are made by identifying patterns and extracting features from large data sets, often referred to as *Big Data* [57]. This chapter describes convolutional neural networks (CNNs), which are specialized deep neural networks for the task of image recognition and processing.

The first section explores image analysis techniques in remote sensing and covers the areas of image preprocessing, pixel-based classification, object-based classification, target recognition, scene understanding and regression analysis. Following this, the subsequent section describes the key concepts and principles that define how CNNs operate. Afterward, the architecture of the convolutional neural network is described, including the details of the layers. Following the architecture description, a section on the CNN training process is provided, which consists of several important steps to reduce training time and improve model accuracy. The last section discusses the architectures of 1D, 2D, 3D and 4D CNNs in the context of remote sensing. In particular, their applications in analyzing spectral/temporal data (1D), spatial data (2D), spectral-spatial data (3D) and spectral-spatial-temporal data (4D) are discussed.

3.1. Image Analysis Techniques in Remote Sensing

Remote sensing data presents new challenges for deep learning. By studying the Earth as an ever-changing system, it is possible to use modern technologies to create lasting records that could help current and future scientific research. The application of deep learning algorithms is now widespread in the world of geosciences, and experimental results confirm the excellent performance of their application in the analysis of remote sensing big data. Despite its great potential, deep learning cannot be directly applied to all remote sensing tasks. For example, hyperspectral images contain a large number of spectral bands, which means that a large number of neurons must be made available in a pre-trained network to analyze such images. Besides the large amount of data, the problem that scientists face when applying deep learning in remote sensing is the small number of labeled samples and the application of a pre-trained network to data captured by different sensors.

There are various approaches to analyzing remote sensing images. This section describes image preprocessing, pixel-based classification, semantic segmentation, object-based classification, target recognition, regression analysis and scene understanding. Image preprocessing improves image quality by using methods based on deep learning for subsequent classification and recognition tasks. Pixel-based classification assigns labels to individual pixels, while object-based classification divides the image into meaningful objects before labeling. Target recognition and scene understanding use features extracted from objects and raw digital pixel numbers (DN – digital numbers) from a database of high and low resolution images. Continuous values, such as chlorophyll, can be predicted using regression analysis, which provides quantitative assessments [58].

3.1.1. Image Preprocessing

Remote sensing images are not always immediately suitable for further analysis due to many factors such as sensor limitation and atmospheric influence. In order to minimize these limitations there is a need to preprocess remote sensing images to improve image quality before further tasks such as classification and image recognition. Section 2.4 describes key preprocessing techniques including radiometric and geometric processing, as well as common digital image processing steps like atmospheric correction, image enhancement, and data fusion. Furthermore, image preprocessing often involves techniques such as remote sensing image restoration and pan-sharpening.

Remote sensing data restoration aims to recover images from their corrupted versions and improve subsequent analysis. This process addresses the degradation caused by atmosphere, imaging system artifacts such as speckles and striping, and instrument noise such as thermal and quantization noise. Techniques such as denoising, despeckling, destriping and deblurring are used to mitigate these effects and improve the interpretability of remote sensing imagery [59].

Pan-sharpening technique combines high-resolution panchromatic images with lower-resolution multispectral images to produce high-resolution multispectral images while keeping important spectral features. Most algorithms work well when the images are acquired in the same time with the same sensor and both the spatial and spectral information is preserved. However, when fusing multiple data, these techniques can improve spatial resolution but often affect spectral consistency. Huang et al. [60] introduced a novel pan-sharpening technique for preprocessing remote sensing images by using a stacked modified sparse denoising autoencoder (S-MSDA) to train the relationship between high-resolution and low-resolution image patches. This deep learning-based approach has demonstrated superior performance over traditional and state-of-the-art methods.

3.1.2. Pixel-based Classification

Pixel-based classification categorizes a pixel based on its spectral information and the similarities between the individual classes. Typically, multispectral data is used where feature vectors of each pixel are extracted, consisting of grayscale values that are used as the numerical basis for categorization. The specified feature vectors are then compared with the prototype vectors of each class using the spectral information of the pixels.

Pixel-based classification reaches its limits when interpreting high-resolution scenes and is therefore often used for low-resolution images, especially when objects are involved. The reason for this is that a pixel cannot represent an entire object, making pixel-based classification less effective for high-resolution images. The most common methods for pixel-based classification are Minimum Distance/Nearest Neighbor Classifier, Parallelepiped Classifier, Iterative Self-Organizing Data Analysis Technique (ISODATA) and Maximum Likelihood Classifiers (MLC) [61].

3.1.3. Semantic Segmentation

Semantic segmentation in remote sensing enables pixel-level image classification, which is critical for applications such as urban planning, disaster assessment, and crop monitoring. Deep learning methods, particularly convolutional neural networks (CNN), attention mechanisms, generative adversarial networks, and models based on data fusion, have advanced its application in remote sensing tasks. Various architectural frameworks, such as fully convolutional networks (FCN), U-Net and DeepLab, have made substantial contributions to this field. In these approaches, CNNs learn to automatically extract hierarchical features and perform end-to-end pixel-wise classification while preserving spatial context. Despite these advances, challenges remain, including variability in image resolution and illumination, limited availability of pixel-wise labeled data, computational demands due to high-resolution images, and difficulty in detecting small objects and fine edges [62].

3.1.4. Object-based Classification

Object-based image analysis is based on objects consisting of a large number of homogeneous pixels grouped in a meaningful way. Image objects reveal shape features, i.e. their spatial distribution, which is necessary for classification. This method consists of two parts, namely image segmentation and classification based on object features in the spectral and spatial domain.

Unlike semantic segmentation which directly assigns class labels to pixels, traditional segmentation algorithms merely group similar pixels into meaningful objects without knowledge of their classes. It is a challenging task due to the mixed pixels and spectral similarity of many land cover types. There are numerous segmentation algorithms in the literature, such

as Fuzzy C-Means, Level Set and Watershed Transformation. Conventional segmentation algorithms usually group pixels iteratively into regions based on predefined similarity criteria, which are often difficult to determine and practically subjective. The result of image segmentation directly affects the performance of subsequent classification tasks. Conventional segmentation methods often do not provide a perfect partitioning of the image, resulting in over or under-segmentation, with over-segmentation being more common.

Image objects are characterized and classified using a variety of features that include color, texture, shape, and contextual properties in various forms. Spectral features such as brightness, mean, standard deviation, soil-adjusted vegetation index (SAVI), normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) are crucial components in feature extraction. Shape features such as area, length, main direction and shape index as well as texture features such as entropy calculated over multiple bands (red, green, blue, near infrared) greatly enhance the classification process. Classification methods often use both a nearest neighbor (NN) classifier, which uses user-defined pattern objects for class assignment, and fuzzy membership functions, which describe feature characteristics within predefined intervals [61, 63].

3.1.5. Target Recognition

Target recognition in remote sensing satellite imagery is a process that involves the identification and classification of individual targets within a scene (e.g. objects, vehicles, people). Image size, different lighting conditions and complex backgrounds can significantly affect target recognition. Methods that rely on manual labeling and classification of features in large scenes encounter efficiency and accuracy problems. In contrast, algorithms based on deep learning have significantly improved the accuracy and speed of target detection and recognition. They can be divided into two categories: regional and end-to-end target detection.

Algorithms based on regional detection (e.g. Region-based CNN or R-CNN and Fast R-CNN) generate candidate images and classify them using CNNs, which improves detection accuracy. However, they require more processing time as numerous suggestions of regions are needed.

Algorithms based on end-to-end target recognition simplify the recognition process and improve processing efficiency by converting target recognition into a regression problem and integrating feature extraction, classification and localization into a single deep CNN. The target recognition method is widely used in military applications, smart city construction, land management and disaster management [64].

3.1.6. Regression Analysis

Regression analysis in remote sensing uses the same basic concepts as regression in other applications. It is used with numerical dependent variables, so the strength and characteristics of the relationship between the dependent variable and one or more independent variables must be determined to better understand and predict the dependent variable. For example, in a remote sensing task, the dependent variable might be chlorophyll concentration or the probability of a natural disaster, while the independent variable might be a band or an index (e.g. Normalized Difference Vegetation Index – NDVI).

When applying regression to remote sensing data, several steps are typically followed, such as: collecting data on known instances of the dependent variable (such as plant or animal species), selecting and defining independent variables, creating equations that describe the relationship between the dependent and independent variables, and finally, by using equations and independent variables, it is possible to create a georeferenced map of the dependent variable on the entire observed area [65].

3.1.7. Scene Understanding

Scene understanding refers to the perception of images obtained through remote sensing to achieve effective analysis, recognition and representation of the geographic scene. In addition to perception, methods such as visual analysis, image processing and pattern recognition are used to recognize patterns and features at different levels of abstraction. The goal is to accurately recognize and understand different classes of objects in the scene and their complex spatial relationships and occurrence of event scenario in series of images.

Therefore, the scene understanding method uses a hierarchical approach ranging from visual level details to the conceptual level, linking pixel-level information to abstract scene-level concepts through semantic processing and reasoning. Identifying spatial structures and relationships within scenes is a key aspect for discovering unique patterns that help to classify and distinguish different types of scenes. Although remote sensing images provide a large amount of information about a particular scene, they are often subject to change under different acquisition conditions and need to be interpreted for accurate labeling and scene analysis [66].

3.2. CNN Key Concepts

To better understand convolutional neural networks and their application, it is necessary to define the fundamental concepts underlying their architecture. These key concepts include filters (or kernels), activation functions, stride and padding, which define the operations within the CNN layers and the way the network processes and transforms the data.

The parameters of the CNN layers consist of a set of learnable filters. Filters or channels consist of kernels that usually have a small spatial dimension, but extend over the entire depth of the input data and have three dimensions: Length (L), Width (W) and Depth (D). A kernel is a two-dimensional tensor with width and length ($W \times L$), where the usual sizes are 3×3 , 5×5 and 7×7 . These grids contain discrete numbers or values that are artificially generated and serve as weights of the kernel at the beginning of the CNN training process, which are adjusted in each training epoch. Increasing the number of filters can improve the CNN's ability to extract significant features, but also increases computational complexity [67].

Figure 3.1 illustrates a 5×5 input image with a 2×2 kernel sliding over it. At each position, the dot product is computed between the kernel and the corresponding section of the input matrix. Here is an example calculation of the first value of the output feature map:

$$(2 * 1) + (0 * 0) + (1 * 0) + (3 * 1) = 2 + 3 = 5$$

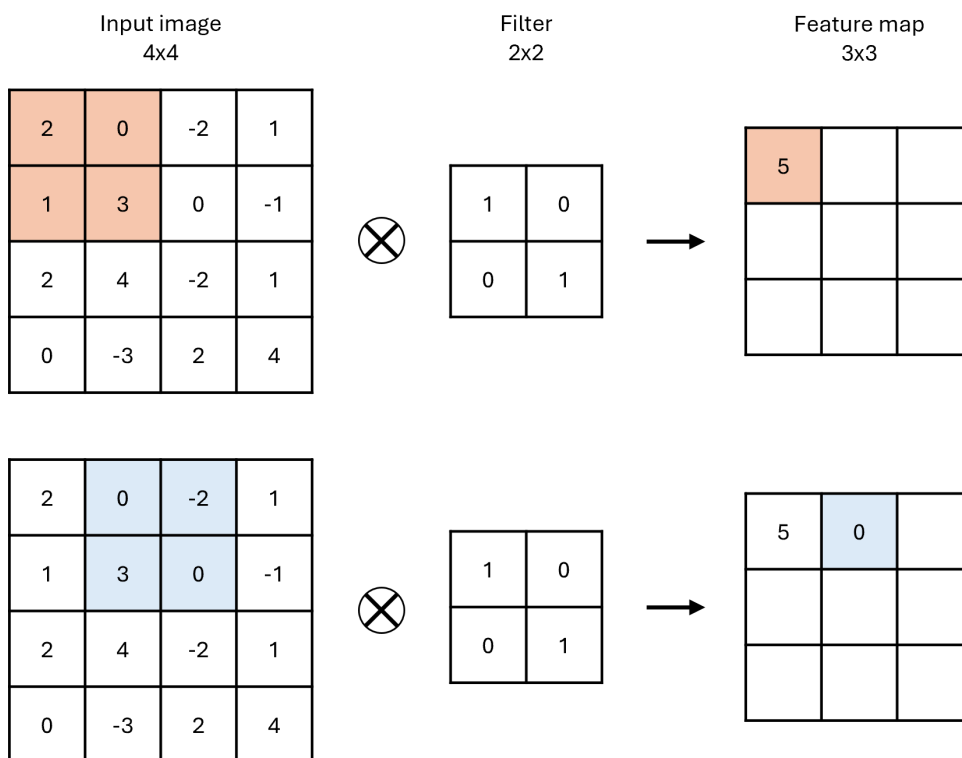


Figure 3.1. Convolution operation on a 5×5 input image using a 2×2 kernel

In deep learning, the input is passed from one neuron to the next through an activation function and this process is repeated until the output layer is reached. During this process and its repetition, the linear relationship is transformed into a non-linear relationship using the activation function. The activation function improves the representational capability through non-linear operations in CNNs and gives them the ability to perform complex calculations. Activation functions are mainly divided into linear and non-linear activation functions. Linear activation functions are not commonly used in CNNs due to the discontinuous properties

of their derivatives. Examples of non-linear activation functions are the Sigmoid, Tangent, Softmax and Rectified Linear Unit (ReLU) functions. The ReLU function is the most commonly used today because it has a simpler definition of the function and gradient, avoids the occurrence of gradient vanishing when applying the backpropagation algorithm, which is the case for the Sigmoid and Tangent activation functions since their gradient is very close to zero, except in the middle of the function, and it creates a so-called sparse representation, which means that the ReLU function can produce complete zero values for negative inputs, while the Sigmoid and Tangent activation functions learn how to approximate a zero output [68].

The stride is a parameter that defines the movement of the kernel or filter over the input matrix that represents an image. It indicates how many rows and columns the kernel or filter moves at each step, starting from the upper left corner of the input matrix. For example, a stride of one moves the kernel by one pixel, as shown in Figure 3.2. A larger stride results in a smaller output dimension of the matrix, which can effectively reduce the size of the image and the amount of computation. However, this operation can result in losing some important information.

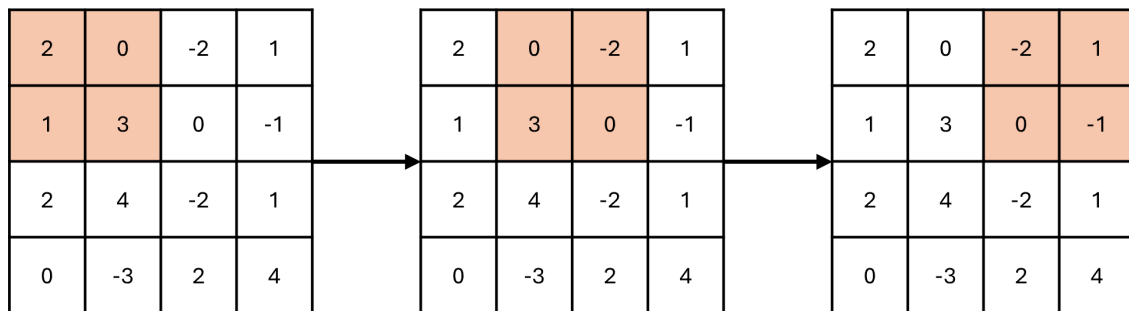


Figure 3.2. Kernel movement with a stride of one

Padding is a common technique used in CNN as it prevents the feature map from shrinking at each layer, by adding a certain number of pixels to the edges of the input data. This ensures that the size of the output data matches the size of the input data, usually using a value of 0 or copying edge pixels to fill the matrix size (Figure 3.3). For example, if a $3 \times 3 \times 1$ image is padded, it expands to a $5 \times 5 \times 1$ matrix. When a $2 \times 2 \times 1$ kernel is applied over this padded image, the resulting feature map will have dimensions of $4 \times 4 \times 1$. This means that the output image has a larger dimension than the input image. If the same procedure is performed without padding, the output could have a smaller dimension than the input image and become a $2 \times 2 \times 1$ image, which could result in loss of important information about the edges [69].

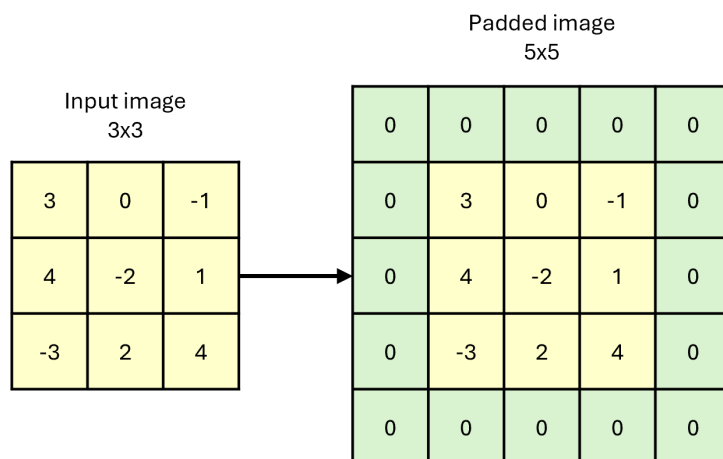


Figure 3.3. Zero padding

3.3. Architecture of CNN

A Convolutional Neural Network is a deep, feed-forward network developed by studying the visual cortex of the brain with the aim of processing and recognizing images. The CNN technique was developed in the 1980s and 1990s, but has become more widespread in recent years due to increased computing power and the availability of large amounts of training data [70]. CNN networks have a grid-like topology and are specialized for image recognition, which can be visualized as a two-dimensional grid of pixels. In addition to visual perception, they are also successfully used in other tasks such as speech recognition or natural language processing (NLP) and time-related data, which can be visualized as a one-dimensional grid with regular time intervals [71].

A CNN achieves better results in image recognition for complex tasks when its feature extraction neural network has more layers, meaning when the neural network is "deeper," which significantly affects the training process. This type of network uses a mathematical operation called 'convolution', which is a specialized type of linear operation used instead of general matrix multiplication in at least one layer of the CNN network. The feature extraction neural network consists of a large number of pairs of convolutional layers and pooling layers. Besides the feature extraction network for the input image, every CNN has a neural network that classifies the image based on its features and generates the output [72].

Figure 3.4 shows the base architecture of a convolutional neural network. This architecture consists of five main components: an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer. CNN for specific task can have more than one of each listed layers to achieve more deeper architecture. The process of transforming input data into output through these layers is known as forward propagation. The following sections briefly describe the most commonly used layers in the CNN architecture.

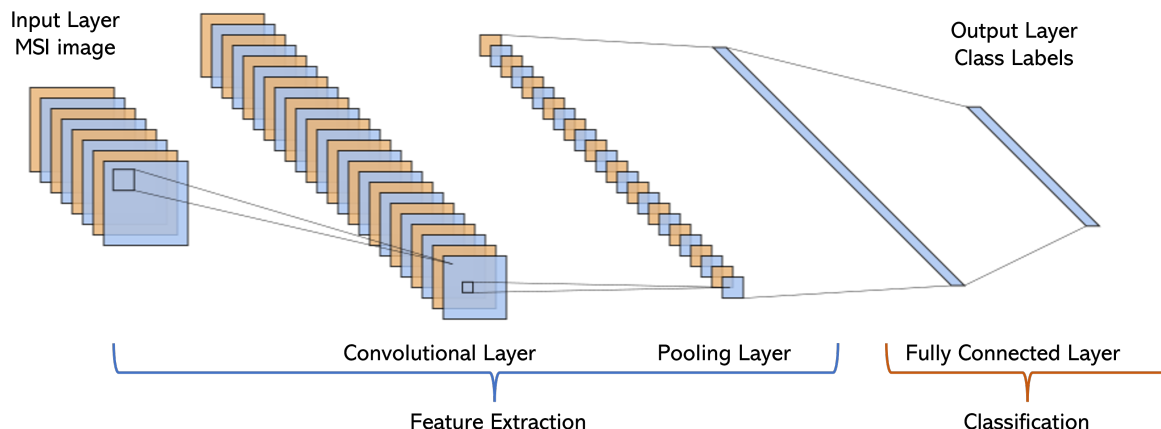


Figure 3.4. Base architecture of convolutional neural network

3.3.1. Convolutional Layer

A convolutional layer is a core component of the CNN architecture that uses the mathematical convolution operation, usually denoted by an asterisk (*), where the equation takes the following form [71]:

$$s(t) = (x * w)(t) \quad (3.1)$$

where the first argument (x) is called the input and the second argument (w) is called the kernel, while the output after applying the convolution is called the feature map.

Typically, a convolutional layer consists of a collection of convolutional filters or kernels, where the input layer, which is usually an image represented as a multidimensional dataset, is convolved using filters to produce an output feature map. The number of extracted features increases with the number of convolution kernels in the layer and they vary depending on which convolution kernel or filter is used.

3.3.2. Pooling Layer

The pooling layer selects only the most relevant results of convolutional layer. It is an important step to continuously reduce the spatial dimension of the feature map, resulting in fewer parameters and computations in the network. At the same time, it retains only the important features or dominant information, which helps to control overfitting and reduce the time required to train the network. This layer allows CNNs to extract the most important features of a particular object, even if its shape is distorted or it is presented at a different angle.

There are different types of pooling layers, e.g. max pooling, average pooling, stochastic pooling and spatial pyramid pooling. The max pooling operation is the most popular and works by selecting the maximum value from each block of the input matrix. Similar to the convolutional layer, it is necessary to define the size of the pooling filter and its step size, which "slides" over the input feature map [73].

3.3.3. Fully Connected Layer

At the end of each CNN architecture there is usually a fully connected (FC) layer in which each neuron is connected to all neurons of the previous layer. A one-dimensional array of numbers (or vectors) from the output feature map of the last convolutional layer or pooling layer is used as input to a fully connected layer. The features generated by applying convolutional layers and pooling layers are mapped from a subset of the fully connected layers to the final network outputs, such as the probability of each class in a classification task. When dealing with classification task, the final fully connected layer usually has the same number of output nodes as the number of classes [74].

3.4. CNN training process

The CNN training process is a critical phase that involves preparing the data and configuring the model to achieve optimal performance. This section includes subsections on data preprocessing and augmentation, parameter initialization, regularization and optimization. Each of these steps is important to improve the training efficiency and accuracy of the CNN model.

3.4.1. Data Preprocessing

Data preprocessing refers to the transformation of a raw data set, which includes a train, validation and test data set, in order to make the data set cleaner, more detailed, easier to learn and in a standardized format. The transformation takes place before the data is entered into the CNN model, whose performance depends on good preprocessing of the data, which can increase the accuracy of the model.

The most commonly used techniques in preprocessing data are: mean-subtraction and normalization. The method of subtracting the mean from each individual data point or feature is used to move the data to the zero center, i.e. zero centering of the data is performed. Mathematically, this process can be described as follows:

$$X' = X - x^* \tag{3.2}$$

$$x^* = \frac{1}{N} \sum_{i=1}^N x_i \tag{3.3}$$

where N represents the size of the training dataset and x^* represents the mean of the data.

In the normalization method, each dimension of the data is divided by its standard deviation. The operation can be implemented mathematically as follows:

$$X'' = \frac{X'}{\sqrt{\frac{\sum_{i=1}^N (x_i - x^*)^2}{N-1}}} \quad (3.4)$$

where the parameters N , X' and x^* have the same meaning as in the equations 3.2 and 3.3 [75].

3.4.2. Data Augmentation

Data augmentation is a technique in which synthetic data is constructed through various operations and transformations of the available data set, thereby expanding the size of the training set for a CNN model. These synthetic data are new versions or new data samples of the available data set, usually containing small changes in the data to which the predictions of the model should be invariant. This technique is extremely useful when only a very limited amount of training data is available and when it is necessary to perform a regularization of the CNN model to avoid the problem of overfitting. Some of the operations used in data augmentation are cropping, rotating, flipping, translation, scaling and noise injection [75,76].

3.4.3. Parameter Initialization

Parameter initialization occurs before the training of a CNN begins and can influence on how fast the CNN will converge and with what accuracy. Therefore, a good initialization of the network parameters can make the training of the network more efficient. One of the simplest methods to initialize the parameters is to set the weights of all layers to zero. However, this method has been shown to be ineffective because the output and gradients are always the same during backpropagation through the network, resulting in all weight updates being identical. In this way, the network would not learn any useful features.

To avoid this problem, the weights are not initialized with the same value, but different techniques are used to initialize the weights randomly. Examples of common methods for initializing parameters are random initialization, Xavier initialization, Gaussian distribution initialization and unsupervised pre-training initialization [75].

3.4.4. Regularization

Overfitting is a major problem for CNN models to achieve good model generalization. Model generalization can be achieved by properly adapting the deep learning algorithm to new or previously unseen inputs that come from the same data distribution as the training data. A model is said to be overfitted if it performs exceptionally well on the training data but does not perform well on the test data, i.e. the unseen data. The opposite of this is an underfitted model, which is not well trained on the training data. A model that performs well on both training and test data can be considered a balanced model and is called a just-fitted model. These three types of models are shown in Figure 3.5.

Regularization helps to avoid overfitting by using intuitive concepts such as dropout, dropping weights or connections between neurons, data augmentation and batch normalization [77].

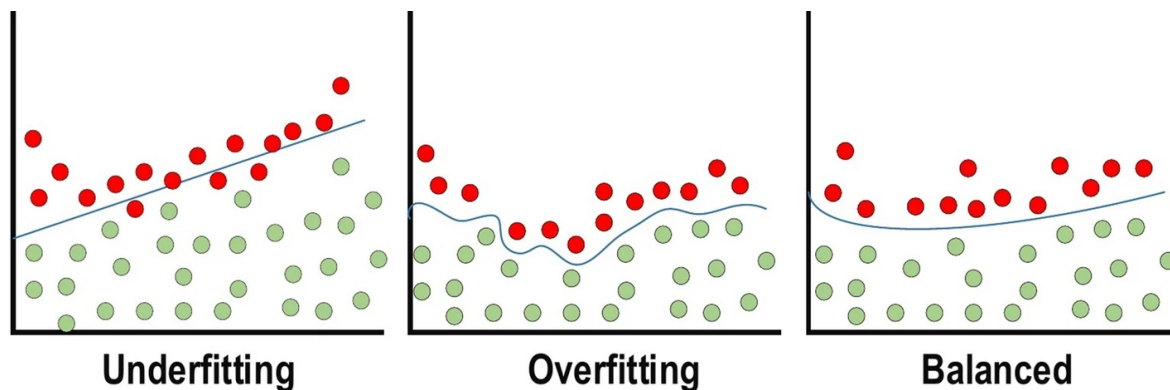


Figure 3.5. Overfitting and Underfitting of Data (Adapted from [77])

3.4.5. Optimization

The CNN model is trained by iteratively updating the parameters of all layers in the network. Therefore, it is very important to choose a good learning algorithm (optimizer) and its extensions (such as AdaDelta, AdaGrad and Momentum) to improve the model output.

The gradient descent algorithm is commonly used to optimize neural networks. In order to minimize the error in the network, the model parameters are updated in the opposite direction of the gradient of the loss function. In this process, the model iteratively searches for a local optimal solution at each training epoch. During each training iteration, the predicted output of the model is compared with the desired output to minimize the loss function, propagating the error backwards. The size of each update step is called the "learning rate" and the complete cycle of updating the parameters over the entire training dataset is called a "training epoch".

One of the most popular metrics to measure performance is the cross entropy, which reaches a value of zero when the desired and predicted outputs match exactly, which is the main goal of any optimization technique. Other algorithms used to optimize CNN models are Batch Gradient Descent, Stochastic Gradient Descent, Mini-batch Gradient Descent, Momentum and Adaptive Moment Estimation (Adam) [75, 78].

3.5. CNN Applications in Remote Sensing

The application of CNNs in the field of remote sensing offers a significant advance in the analysis of multispectral and hyperspectral images. By using deep learning CNNs can recognise complex patterns in spectral, spatial and temporal data across different image datasets.

In this way, it is possible to effectively capture the relationships between neighbouring pixels and temporal changes in captured scenes, which contributes to better classification and analysis accuracy of remote sensing images. This section outlines the main CNN concepts: those that focus on spectral or temporal data (1D), those that capture spatial relationships between pixels (2D), those that integrate spectral and spatial data (3D), and those that incorporate spectral, spatial, and temporal data (4D). Systematic overview of their applications across various domains will be further elaborated in Chapter 4.

3.5.1. 1D-CNN (Spectral/Temporal)

One-dimensional convolutional neural networks (1D-CNNs) have been shown to be effective in extracting spectral features from multispectral or hyperspectral remote sensing images. Spectral signatures, often represented as spectral response curves, provide unique reflectance values for each pixel that aid in the identification and classification of objects or materials on the Earth's surface. The main advantage of 1D-CNNs over conventional methods lies in their ability to automatically detect and extract spectral features across multiple wavelengths. Conventional techniques often rely on manual feature extraction, which is time-consuming and not very adaptable. Although spectral information improves classification performance, they cannot always distinguish objects with similar spectral signatures but different spatial structures, leading to potential misclassifications [79].

1D-CNNs are also increasingly used for analyzing temporal data and time series by learning hierarchical patterns directly from sequential data. This capability facilitates effective modeling of temporal dependencies and improves classification accuracy in applications such as land use/land cover mapping from multitemporal remote sensing imagery [80]. Furthermore, while 1D-CNNs can be applied to spatial data at the pixel level [81], their use in remote sensing research is more common for spectral and temporal data.

Figure 3.6 shows an example of a 1D-CNN architecture consisting of an input layer that represents the spectral values of a pixel as a one-dimensional array. This is followed by two one-dimensional convolution layers (1D Conv), a pooling layer and a fully connected (FC) layer. The result is a map representing the land cover classification.

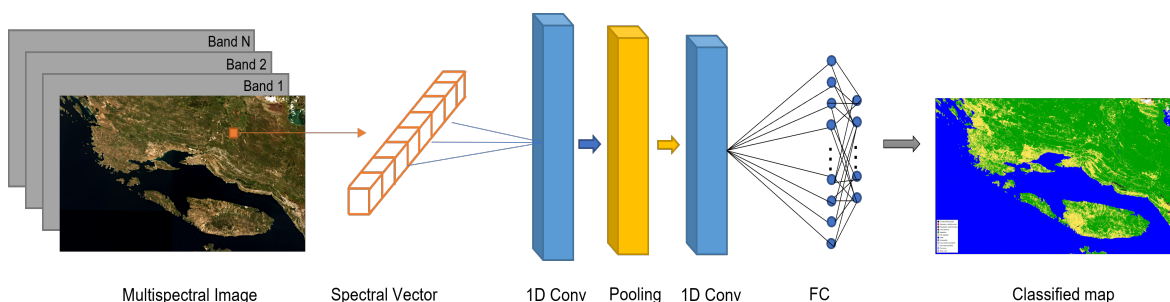


Figure 3.6. 1D-CNN model architecture for spectral feature extraction in MSI data

3.5.2. 2D-CNN (Spatial)

Two-dimensional convolutional neural networks (2D-CNNs) are used for feature extraction, object classification and spatial pattern analysis in remote sensing images. In contrast to 1D-CNNs, which only focus on the spectral characteristics of the observed objects, 2D-CNNs take into account the spatial arrangement and relationship between pixels to classify and identify features on the Earth's surface. Unlike traditional machine learning algorithms that use pixel-based inputs, 2D-CNNs utilize patch-like inputs to identify and extract the best classification features (e.g., shape, size, texture). Using patches to merge the spatial information of neighboring pixels has proven to be effective in classifying different species with similar spatial features, such as: agricultural fields, urban areas and water bodies [82].

Figure 3.7 illustrates a typical 2D-CNN architecture for spatial feature extraction, which begins with a convolutional layer that applies filters to the input image. This is followed by a pooling layer that reduces the spatial dimensions while preserving important features, allowing the network to focus on the most relevant information. The hierarchical feature extraction process continues through multiple layers, each learning increasingly abstract representations, and culminates in a fully connected layer that outputs class probabilities or segmentation maps [83]. This layered approach allows 2D-CNNs to effectively capture and interpret the complex spatial structures present in remote sensing imagery. On the other hand, the disadvantage of 2D-CNN is that it extracts spatial information from one spectral band or index at a time, failing to take advantage of additional information present in other bands.

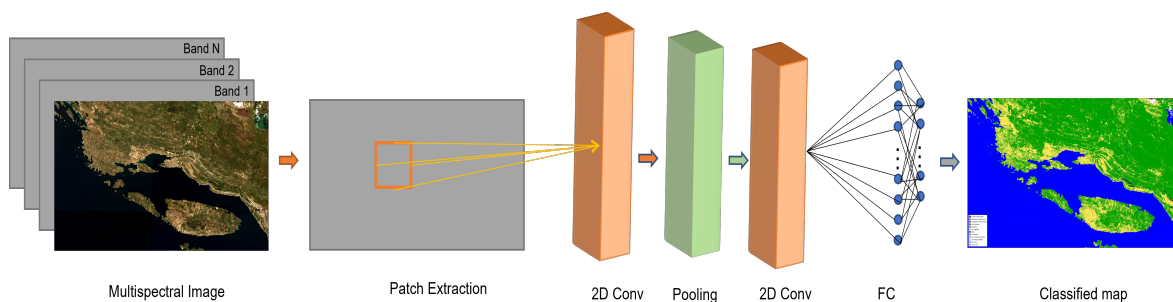


Figure 3.7. 2D-CNN model architecture for spatial feature extraction in MSI data

3.5.3. 3D-CNN (Spectral-Spatial)

A three-dimensional convolutional neural network (3D-CNN) in remote sensing image analysis enables the extraction of spectral and spatial information formed into a data cube. A data cube is usually formed from a large number of images stacked on top of each other. The number of stacked images depends on whether they were acquired with a multispectral or hyperspectral sensor, with each individual image representing a spectral band of the selected scene.

Therefore, 3D filters are formed to capture the spatial (width and height) image patch and the spectral dimension (depth). Figure 3.8 shows an example of a 3D-CNN architecture, which shows that data cubes are fed into the 3D convolutional layer during patch extraction. The efficiency of model learning can be affected by the size of the image patches used as input, where too much data can lead to noise and too little data can limit the receptive field [84]. Therefore, selecting an optimal patch size is important to balancing the trade-off between capturing sufficient contextual information and avoiding unnecessary noise.

By simultaneously analyzing spectral signatures and spatial relationships, 3D-CNNs can achieve a more comprehensive understanding of complex scenes compared to 1D-CNNs and 2D-CNNs. Furthermore, 3D-CNNs can be used to improve the accuracy of land cover classification, anomaly detection, and environmental monitoring, making them a powerful tool for the advancement of remote sensing applications. In addition, 3D-CNNs can also take advantage of the temporal dimension and enable the classification of images acquired at different times [85].

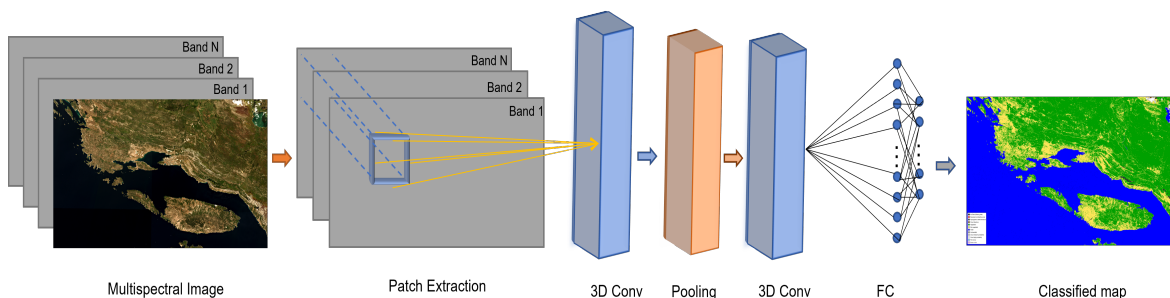


Figure 3.8. 3D-CNN model architecture for spatial-spectral feature extraction in MSI data

3.5.4. 4D-CNN (Spectral-Spatial-Temporal)

A four-dimensional convolutional neural network (4D-CNN) enables the simultaneous monitoring of spectral, spatial, and temporal features when analyzing remote sensing images. Figure 3.9 shows a 4D-CNN architecture where the input data consists of 3D cubes containing the spectral-spatial information of the scene over different times (t_1, t_2, \dots, t_N). In this way, the model can learn complex patterns in different spectral-spatial contexts over time. These patterns are crucial for capturing the dynamic changes and trends observed in remote sensing data [86].

4D-CNNs require large datasets with consistent temporal coverage to effectively learn from changes over time. This can be a challenge, especially in cases where remote sensing is affected by clouds or there is no continuous data collection. Furthermore, the inclusion of an additional dimension increases the computational complexity of the model. A disadvantage of 4D-CNNs is that they are not natively available in common deep learning frameworks such as TensorFlow [87] or PyTorch [88]. Therefore, 4D-CNNs need to be implemented

with a series of 3D convolutional layers along the time dimension. This involves stacking multiple 3D convolutional layers that operate on remote sensing images at different points in time so that the network can learn the temporal dynamics of the scene [89]. The advantage of 4D-CNNs over the previously described CNNs lies in their ability to handle various remote sensing tasks that involve temporal analysis, such as land cover change detection, disaster monitoring, and agricultural monitoring. This makes 4D-CNNs a powerful tool for advancing remote sensing applications through a comprehensive understanding of temporal dynamics and trends.

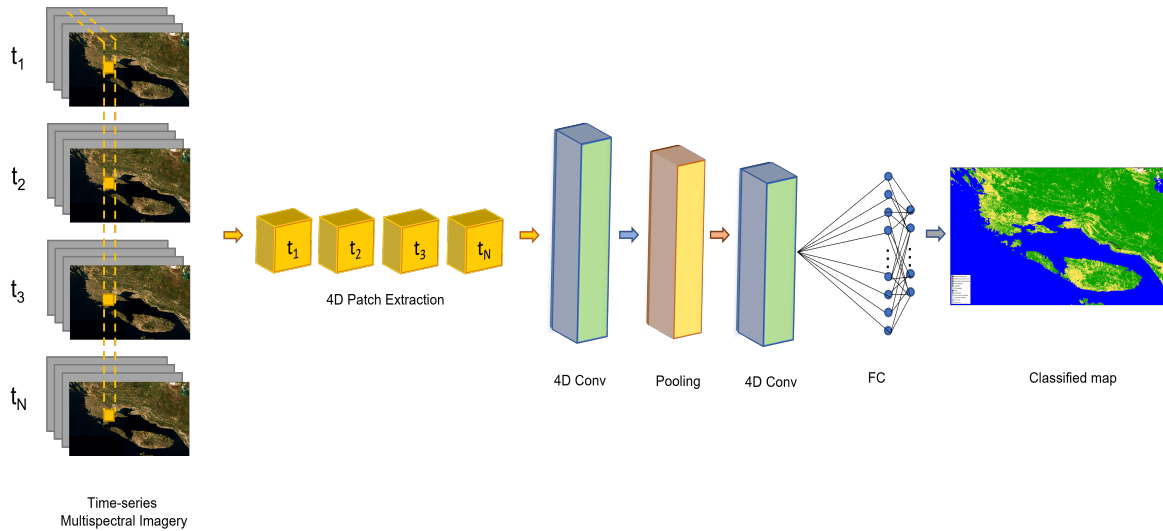


Figure 3.9. 4D-CNN model architecture for spatial-spectral feature extraction in time-series MSI data

4. SYSTEMATIC LITERATURE REVIEW OF CNN ARCHITECTURES IN MULTISPECTRAL IMAGERY

This chapter, based on the systematic review [90], presents an overview of 1D, 2D, 3D and 4D-CNN architectures applied to multispectral images. It summarizes the most important aspects of the types of multispectral image (MSI) data used and their application domains. In the introduction, the relevant literature is explained and the research questions are formulated. The Methods and Materials section describes the methods used, which are organized according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. In the Results and Discussion section, the most important results are described and analyzed. The chapter concludes with a summary of the key findings and recommendations for future work.

4.1. Introduction

In recent years, interest in the classification of hyperspectral (HSI) and multispectral (MSI) images using deep learning methods has increased significantly. Although many articles have been published on this topic, only a limited number of review articles provide a comprehensive overview of the current state of research in this area. Most review articles focus on the classification of hyperspectral images and explore different approaches, algorithms, techniques, and applications in this area [91–94]. For instance, in [95], the authors conducted a comprehensive comparison of several convolutional neural network architectures, including 1D-CNN, 2D-CNN, 3D-CNN, and feature fusion-based CNN (FCNN), for the purpose of hyperspectral image classification. In [96], the authors provided a brief overview of several deep learning models that can be utilized for hyperspectral image classification, such as CNNs, stacked autoencoders (SAE), and deep belief networks (DBN). They systematically analyzed the state-of-the-art deep learning approaches from two perspectives: pixel-wise image classification and scene-wise image classification.

However, as of this writing and to the best of the author’s knowledge, there is currently no review article that focuses solely on the application of deep learning techniques to multispectral images. There are several review articles that address the broader topic of deep learn-

ing and remote sensing images, which encompasses hyperspectral, multispectral, unmanned aerial vehicle (UAV), and synthetic aperture radar (SAR) imagery [58, 97, 98]. Additionally, authors in [99, 100] have provided a comprehensive review and resources for high-resolution multispectral imagery in the context of scene classification, object detection, segmentation, and image retrieval.

Although there are multiple review articles that cover the topic of MSI classification, a comprehensive analysis of the CNN-based approach has not yet been provided. This indicates a need for a more in-depth exploration and analysis of CNN-based MSI classification. To address this gap, this section focuses specifically on different types of CNNs based on the dimensionality of input data. Specifically, it examines 1D, 2D, 3D, and 4D CNNs, which respectively cover spectral, spatial, spectral-spatial, and spectral-spatial through time classification.

To identify the type of CNN (1D, 2D, 3D, or 4D), the term "degree of convolution" is introduced to specify the dimensionality of the input data in a convolutional neural network. In CNNs, the term "dimension" can refer to the number of dimensions in the input data or feature map, as well as the number of filters or channels in a layer that do not necessarily match the input data or feature map's dimension. To avoid ambiguity, this thesis uses the term "degree" to identify the type of CNN (1D, 2D, 3D, or 4D) based on the dimensionality of the input data.

4.1.1. Problem formulation

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were used for a systematic review of the literature on the application of different degrees of convolutional neural networks to multispectral images. PRISMA provides a framework for conducting and reporting systematic reviews and meta-analyses [101]. The aim of this systematic review is to answer the following research questions (RQ):

- **RQ1:** In which application domains have different CNN models been successfully applied to process multispectral image data?
- **RQ2:** What are the commonly utilized MSI datasets for training CNN models in the context of multispectral satellite imagery?
- **RQ3:** How does the degree of CNN impact the performance of classification, regression, or segmentation tasks for multispectral satellite imagery?

4.2. Materials and Methods

This section provides a comprehensive description of the study approach, including eligibility criteria, information sources, search strategies, data extraction and effect measures in line

with the rationale and objectives stated in the PRISMA guidelines.

4.2.1. Eligibility Criteria

The main eligibility criteria for inclusion in this study were scientific publications that applied 1D, 2D, 3D or 4D CNNs in the field of remote sensing. The broader term 'remote sensing' was employed to capture studies that may not explicitly mention multispectral imagery but rather refer to satellite names or datasets relevant to the field. Studies not focused on multispectral satellite imagery (e.g., hyperspectral or UAV imagery) were excluded.

Additionally, publications that focused only on well-known CNN models such as AlexNet, ResNet and VGG were not considered eligible, but studies that originally used different degrees of CNN or adapted them in their models were included. Only articles and conference papers published in English up to March 2024 were considered. The detailed eligibility criteria for this review is summarized in Table 4.1.

Table 4.1. Eligibility criteria.

Type of data	Multispectral satellite imagery
Algorithms or Techniques	Convolutional neural networks (1D, 2D, 3D, and 4D) for image classification, object detection, semantic segmentation or regression
Comparator	RQ1: e.g. Vegetation, Urban RQ2: e.g. Sentinel, Landsat, MODIS RQ3: Evaluation metrics for classification, regression and segmentation
Outcome	Classification and characterization of the diverse applications of CNNs within the domain of remote sensing
Timing	All
Environmental or Geographical Context	All
Study Design	Original data Relevant articles and conference papers
Publications	Studies published in English only

4.2.2. Information Sources

A search of the Web of Science, IEEE Xplore and Scopus databases was conducted on May 10, 2024 to find relevant studies. These databases were selected due to their comprehensive coverage of scientific literature in various fields. The search included publications available until March 31, 2024.

4.2.3. Search Strategy

The search strategy used a combination of keywords related to convolutional neural networks and remote sensing. Specifically, the search query included variations of 'Nd-cnn' OR 'Nd cnn' OR 'Nd convolutional' AND 'remote sensing', where 'N' stands for the degree of CNN, ranging from one to four. All publications were restricted to those that were published in English and had the source type 'Journal' or 'Conference Proceedings'. The same query was performed in all three databases: Web of Science, IEEE Xplore and Scopus. An exception was the Scopus database, where the search was limited to publication titles, keywords and abstracts due to the very large number of publications indexed in this database. To cope with the large volume of search results in the Scopus database, filters were also applied to keywords to exclude irrelevant content (e.g. electrocardiogram, hyperspectral images). Given the limitations of specifying the search date in IEEE Xplore, publications up to the search date were manually reviewed to exclude those published after March 31, 2024, to ensure the relevance and timeliness of the search results. Table 4.2 shows the basic search queries for all three databases, where Nd refers to 1D, 2D, 3D and 4D CNNs.

Table 4.2. Search strategy for different databases and CNNs

Database	Search Query
Web of Science	((((ALL=(Nd - cnn)) OR ALL=(Nd cnn)) OR ALL=(Nd convolutional))) AND ALL=(remote sensing)) AND (LA=="ENGLISH") AND DT=="ARTICLE" OR "PROCEEDINGS PAPER") AND DOP=1990-01-01/2024-03-31
IEEE Xplore	(((("All Metadata":"Nd-cnn" OR "All Metadata":"Nd cnn" OR "All Metadata":"Nd convolutional") AND ("All Metadata":"remote sensing")))) AND ("ContentType":"Journals" OR "ContentType":"Conferences")
Scopus	((TITLE-ABS-KEY (Nd-cnn) OR TITLE-ABS-KEY (Nd AND cnn) OR TITLE-ABS-KEY (Nd AND convolutional)) AND ALL (remote AND sensing)) AND ((PUBYEAR > 1992 AND PUBYEAR < 2024) OR PUBDATETXT (january 2024) OR PUBDATETXT (february 2024) OR PUBDATETXT (march 2024)) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j") OR LIMIT-TO (SRCTYPE, "p"))

4.2.4. Data Extraction

The extracted data from the included studies follows standardized forms: Publication Type (Journal or Conference), Authors, Article Title, Source Title, Publication Year, Author Keywords, Abstract, Research Areas, Open Access, Satellite, Domain, Application, Machine Learning Technique, Algorithm Complexity, Accuracy, F_1 -Score, Precision, Recall, Producer Accuracy, User Accuracy, Pixel Accuracy, Kappa, IoU, MIoU, Dice, R-Square, RMSE, MAE and Parameter Description.

4.2.5. Effect measures

The performance of CNNs include evaluation metrics across different tasks such as classification, segmentation and regression. For classification and segmentation tasks, the collected data was analyzed using the following metrics: accuracy, F_1 -score, precision, recall, producer accuracy, user accuracy, pixel accuracy, kappa, Intersection over Union (IoU), Mean Intersection over Union (MIoU) and Dice coefficient. These metrics are expressed as percentages (%). For regression tasks, we extracted data from the collected publications for metrics R-squared (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) all of which are expressed according to the predicted unit (e.g. chlorophyll-a concentration as mg/m^3). It is important to note that in some publications, it was not possible to access every measure for the performed tasks.

4.3. Results and Discussion

4.3.1. Study Selection

A systematic review of the literature from the Web of Science, IEEE Xplore, and Scopus databases retrieved 678 articles for 1D-CNN related studies, 1213 articles for 2D-CNN related studies, 2013 articles for 3D-CNN related studies, and 48 articles for 4D-CNN related studies. Studies included in this review were published until March 31st, 2024. Due to the limitation of the IEEE Xplore database, which does not allow filtering by date or month but only by year, the publications were manually reviewed and those published after March 31st were removed. After that, the title of each publication was screened, resulting in 405 studies related to 1D-CNN, 765 studies related to 2D-CNN, 1219 studies related to 3D-CNN, and 37 studies related to 4D-CNN. In the next stage of abstract screening and full-text assessment (if applicable), studies that were not relevant to this systematic review were excluded. Thus, in the final review process, 49 studies were included for 1D-CNN, 58 for 2D-CNN, 66 for 3D-CNN, and 3 for 4D-CNN (Figure 4.1).

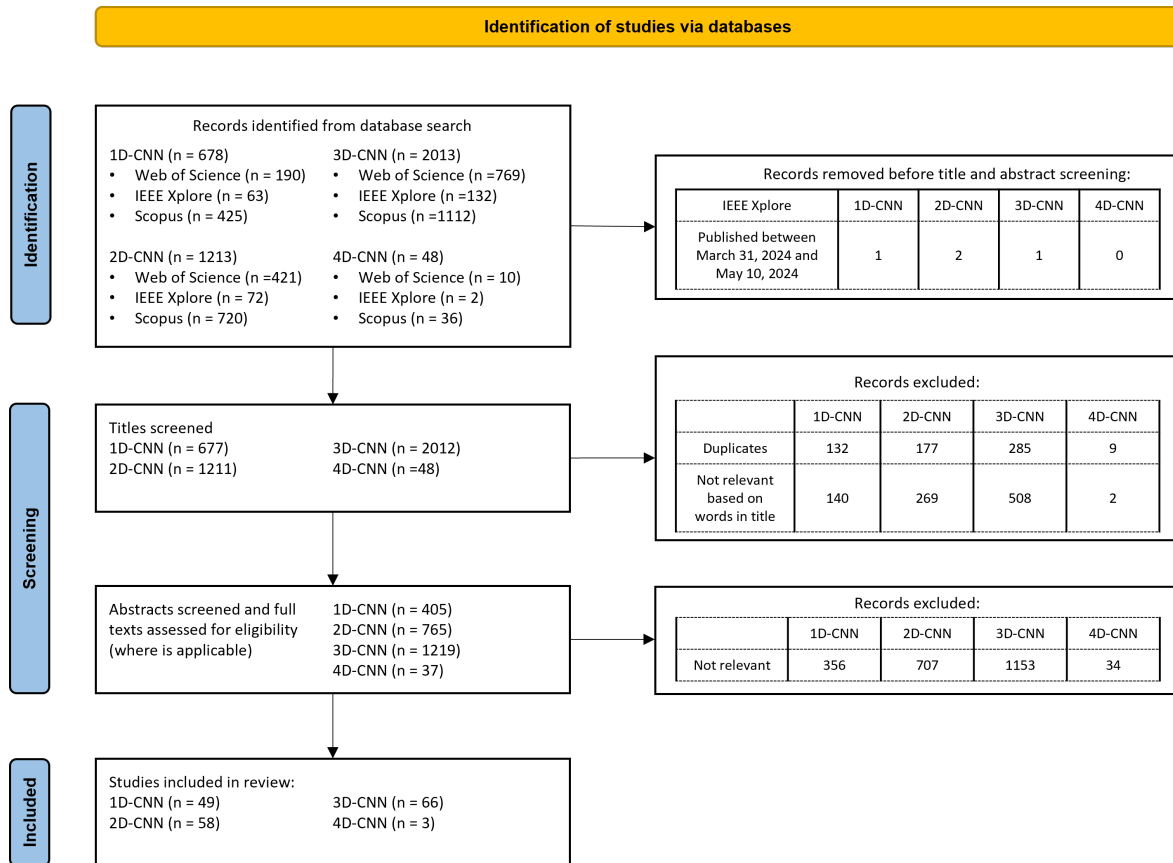


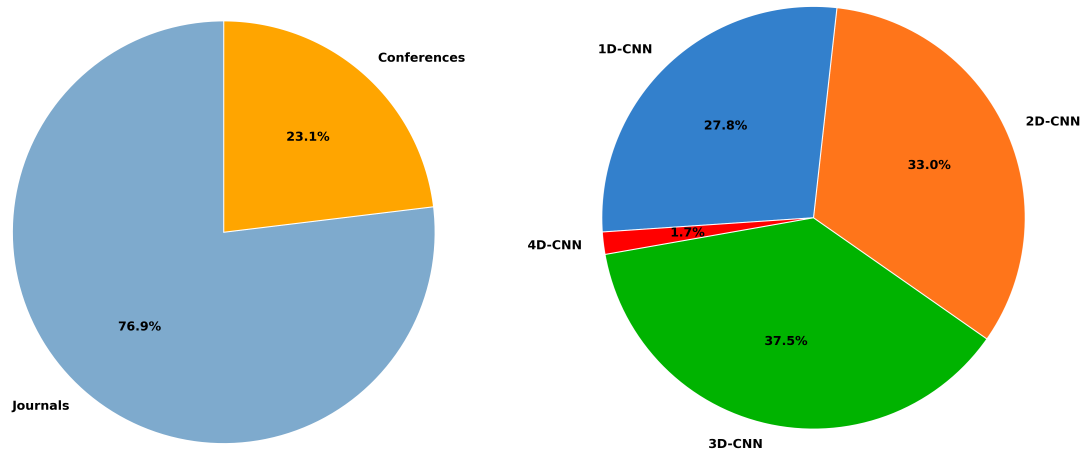
Figure 4.1. Flow diagram illustrating the publication identification and screening process following the PRISMA guidelines (template is reused from Page et al. [101] with CC BY 4.0)

4.3.2. Overview of Remote Sensing Publications

The systematic review included a total of 143 unique publications, comprising approximately three quarters (110) of journal articles and one quarter (33) of conference articles (Figure 4.2).

In examining the distribution of articles using convolutional neural networks, it is evident that publications using 3D-CNNs dominate. It is interesting that 3D-CNN allows the extraction of spatial features over multiple bands simultaneously, potentially providing better classification accuracy compared to 1D-CNNs and 2D-CNNs [102]. In contrast, 4D-CNNs shows the lowest representation with less than 2% of the total publications included in our systematic review. 4D-CNNs are under explored in terms of their application across diverse domains. While we found applications in Vegetation and Water domains, there's a lack of studies in Agriculture, Urban, and Geohazards domains. Additionally, the temporal aspect of 4D-CNNs is not fully utilized in most studies. The reason for this could be the complexity of 4D-CNNs and their unavailability in some popular deep learning frameworks such as Tensorflow or PyTorch. As a result, 4D-CNNs must be implemented using a sequence of 3D convolutions along the temporal dimension, involving stacking multiple 3D convolutional

layers that operate on the multispectral images at different points in time. This allows the network to learn the temporal dynamics of the scene [89]. While this approach may be computationally complex and not be as efficient as using native 4D convolutional layers, it still allows for the joint exploitation of spatial, spectral, and temporal information in multispectral imagery.



(a) Proportion of publications by publication type (b) Proportion of publications by CNN architecture

Figure 4.2. Distribution of publications

Figure 4.3 presents the heatmap illustrating the distribution of the most common journals used over the years. It is interesting to note that the majority of publications are published in journals associated with the term 'remote sensing' in their names. The most prevalent journal sources for the publications were Remote Sensing and ISPRS Journal of Photogrammetry and Remote Sensing.

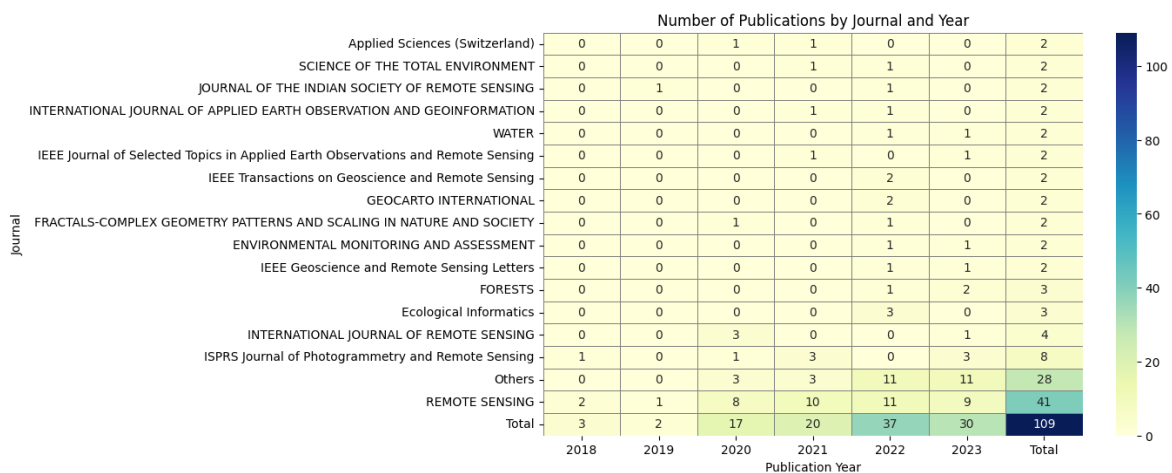


Figure 4.3. Heatmap of the most frequent journals of publications focusing on CNN applications in multispectral imagery

Similarly, Figure 4.4 presents a heatmap for the most selected conferences over years for publishing remote sensing related articles. As for the conference sources, the Interna-

tional Geoscience and Remote Sensing Symposium (IGARSS) had the highest number of published articles. The label "Others" is assigned to journals or conferences that have been identified as sources with only one publication.

Over the past decade, CNNs have become increasingly popular in remote sensing applications, particularly in processing MSI data which suggest these numbers of published articles. A notable increase in research output can be observed from 2020 to 2023, with 5 or more conference papers published each year and at least 17 journal articles released annually, indicating a consistent trend in research output. This is further illustrated by the overall number of publications for each degree of CNN over the years in Figure 4.5. Moreover, while the number of publications peaked in 2022, there was a decrease in publications for 3D-CNN thereafter (24 → 9), but an increase was observed for 1D-CNN (16 → 24).

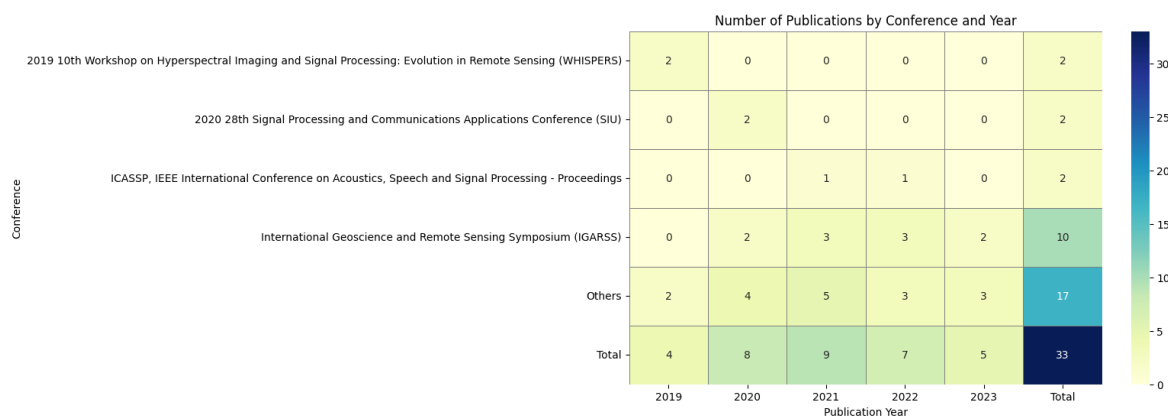


Figure 4.4. Heatmap of the most frequent conferences of publications focusing on CNN applications in multispectral imagery

Figure 4.6 provides a visual representation of the highest-frequency terms found in the titles and abstracts of peer-reviewed literature with larger font sizes indicating higher frequency. It can be noticed that the terms with the highest frequency are 'data' and 'model' which suggest the importance of selecting quality data and models in remote sensing research. Furthermore, terms that had a slightly lower frequency are 'method,' 'classification,' 'deep learning,' and 'neural network' which suggest common approaches used in publications. This observation underlines the importance of methodological choices in analyzing and interpreting remote sensing data.

Figure 4.7 shows a publication distribution for each CNN over different applications that can be classified in one of six different domains: Agriculture (e.g., crop classification), Geohazards (e.g. wildfire prediction), Urban (e.g. urban land cover maps), Vegetation (e.g. forest classification), Water (e.g. estuary water quality classification), and Others (e.g. spatiotemporal image fusion). Most publications used first three degree of CNNs to solve problems in the domain of Agriculture, specifically for crop classification and crop yield prediction. The second most used domain is Vegetation, where the 4D-CNN is applied for segmentation. Vegetation usually includes studies involving the analysis of plants such as trees, shrubs,

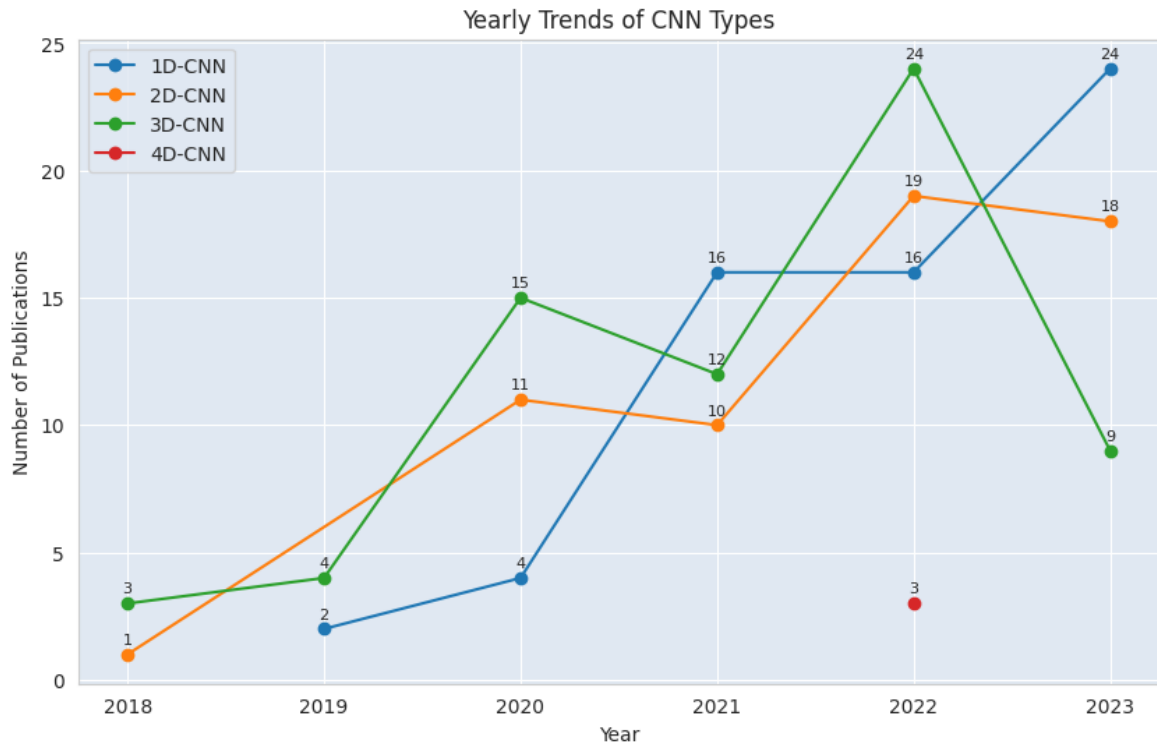


Figure 4.5. Publication Trends Over Time for Each 1D, 2D, 3D and 4D-CNN

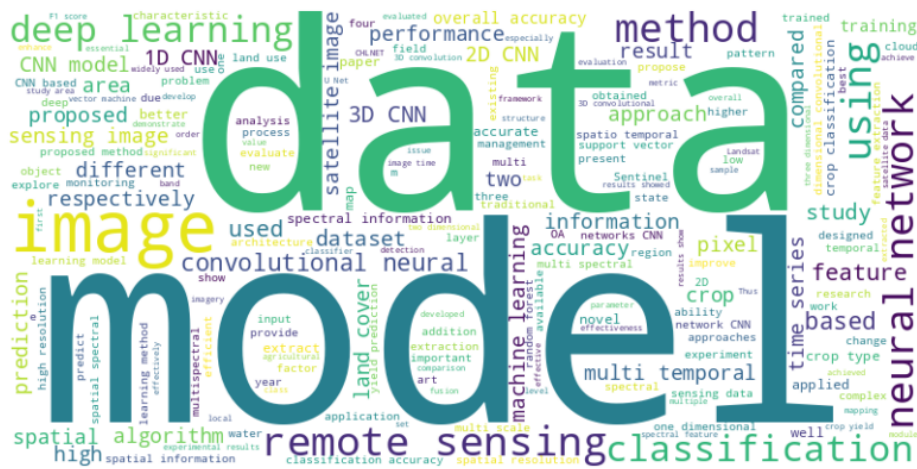


Figure 4.6. The word cloud of the most frequently used words in abstracts

grasslands and natural vegetation [103]. Agriculture, on the other hand, includes studies focused on crop breeding, agricultural land use monitoring and crop yield forecasting, often relying on vegetation data to to assess crop health, productivity, and land use [104]. Thus, both domains involve monitoring and analysis of plants life.

The distribution of publications of each CNN categorized by satellite data sources is presented in Figure 4.8. The plot shows that the majority of publications utilized data from the Sentinel-2 satellite when implementing 1D, 2D, and 3D CNNs. Sentinel-2 satellites, operated by ESA, have 13 bands with spatial resolutions ranging from 10 m to 60 m and

a temporal resolution of 5 days [46]. Notably, satellites Landsat-8 and MODIS have been preferred across all degrees of CNNs, even though there are almost half as many publications compared to Sentinel-2. Landsat-8 carries a two sensor payload, the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS), where OLI captures nine spectral bands including one panchromatic band while TIRS has two thermal bands. It has a temporal resolution of 16 days, with a 15 meter spatial resolution for the panchromatic band and a 30 meter spatial resolution for the multispectral bands [105]. On the other hand, MODIS is a NASA instrument launched on two satellites – MODIS Terra and MODIS Aqua. It has 36 spectral bands and a 1–2 day temporal resolution. It collects data at three spatial resolutions: 250, 500, and 1000 meters [106].

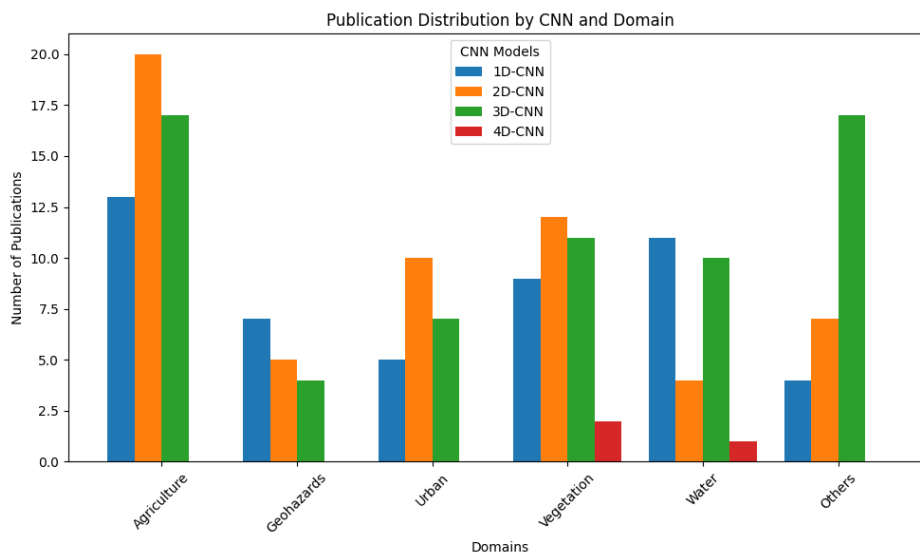


Figure 4.7. Publication Distribution for each CNN by Domain

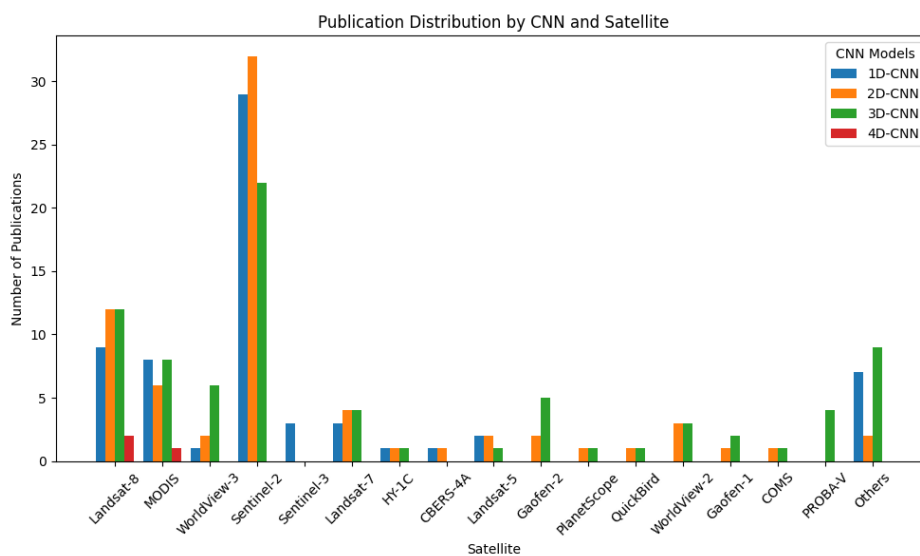


Figure 4.8. Publication Distribution for each CNN by Satellite

4.3.3. Meta-Analysis of Publications

As can be seen in Figure 4.8, the highest representation is of the Sentinel-2, Landsat-8, and MODIS satellites across different degrees of CNNs. Similar situation can be observed at the individual usage of satellites for each degree of CNN in particular application domains. Specifically, for 1D-CNN, in Figure 4.9 it can be seen that Sentinel-2 is the most represented in all domains, with the highest application in the Agriculture domain. Landsat-8 is most represented in the 'Water' domain, followed by 'Vegetation', 'Others', and 'Urban'. As for MODIS, it is equally used in the 'Agriculture', 'Geohazards', and 'Water' domains.

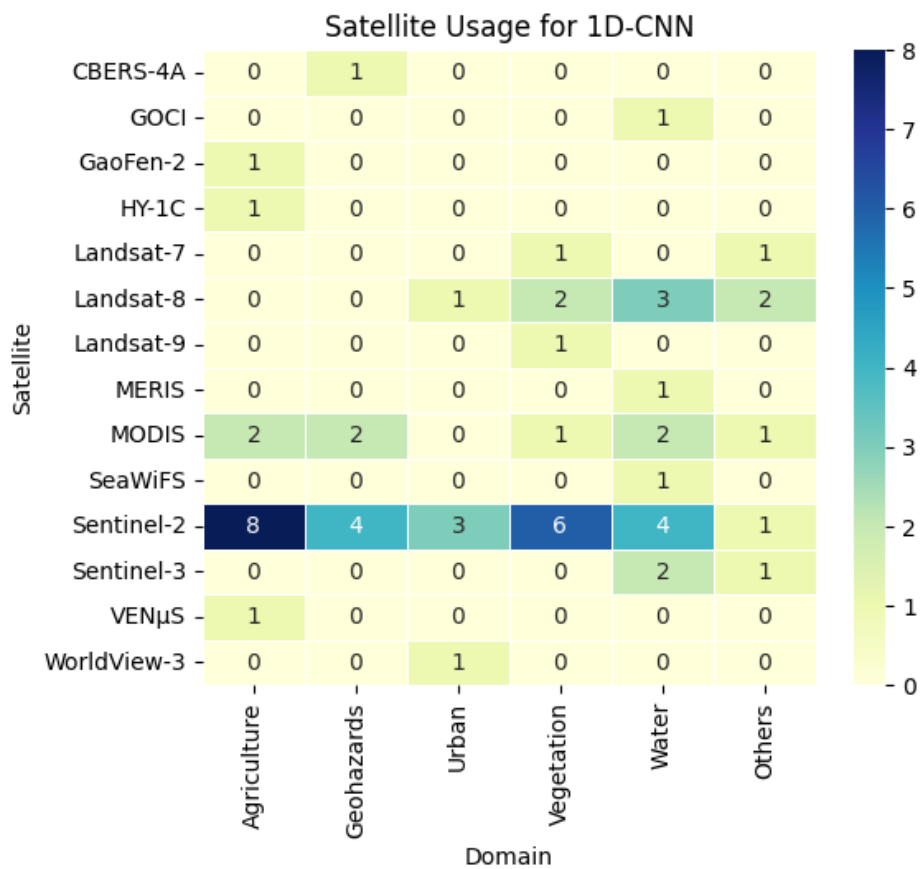


Figure 4.9. Satellite usage across different domains for 1D-CNN

Regarding the use of 2D-CNN, when the individual application domains are examined in Figure 4.10, it can be seen that in the 'Agriculture' domain, Sentinel-2 is the most represented satellite, followed by Landsat-8 and MODIS. In the 'Geohazards' domain, for example, Landsat-7 is identified as the most represented satellite, while in the 'Urban' domain, in addition to the three most common satellites (Sentinel-2, Landsat-8, and MODIS), Landsat-5 is also represented.

Figure 4.11 shows that more different satellites are used in the study for 3D-CNN compared to 1D and 2D CNN. For example, in the 'Agricultural' domain, in addition to the three most commonly used satellites, the Gaofen-2 satellite is also used. The 'Others' domain is

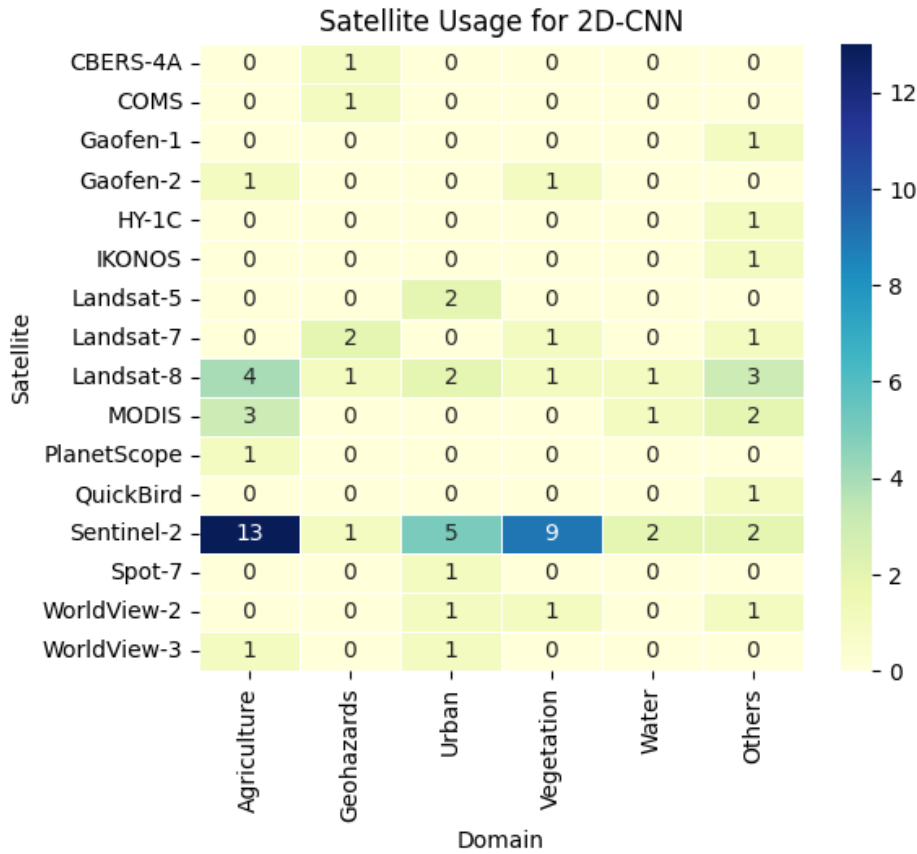


Figure 4.10. Satellite usage across different domains for 2D-CNN

also interesting, where many publications have used satellites such as Gaofen-1, Landsat-8, PROBA-V, and WorldView-3. For instance, studies that used Gaofen-1 data were applied in cloud masking [107, 108]. As for Landsat-8 data, they were used for Image Preprocessing and Enhancement purposes [109–112]. PROBA-V data were mainly utilized for improving satellite image quality [113–116], while data from the WorldView-3 satellite were used for Image Analysis and Processing [117–119] and Image Compression [110].

The number of publications that utilized 4D-CNN is too small, so there is not enough data for general conclusions (Figure 4.12). There are a total of three publications where 4D-CNN was applied, of which two used Landsat-8 data for problem-solving in the 'Vegetation' domain and one used MODIS data for problem-solving in the 'Water' domain. Both publications related to using 4D-CNN on Landsat-8 data in the 'Vegetation' domain were focused on land cover classification [86, 89], while the publication related to MODIS data in the 'Water' domain used 4D-CNN for cyanobacteria bloom prediction [120].

As shown in Figure 4.9, Sentinel-2 is the most frequently used satellite data source across different CNN architectures. Further analysis reveals that a significant portion of these Sentinel-2 based studies focus on agricultural applications. For instance, in the 1D-CNN category (Figure 4.10), Sentinel-2 data is predominantly used in the Agriculture domain. This trend is consistent across 2D-CNN (Figure 4.11) and 3D-CNN (Figure 4.12) applica-

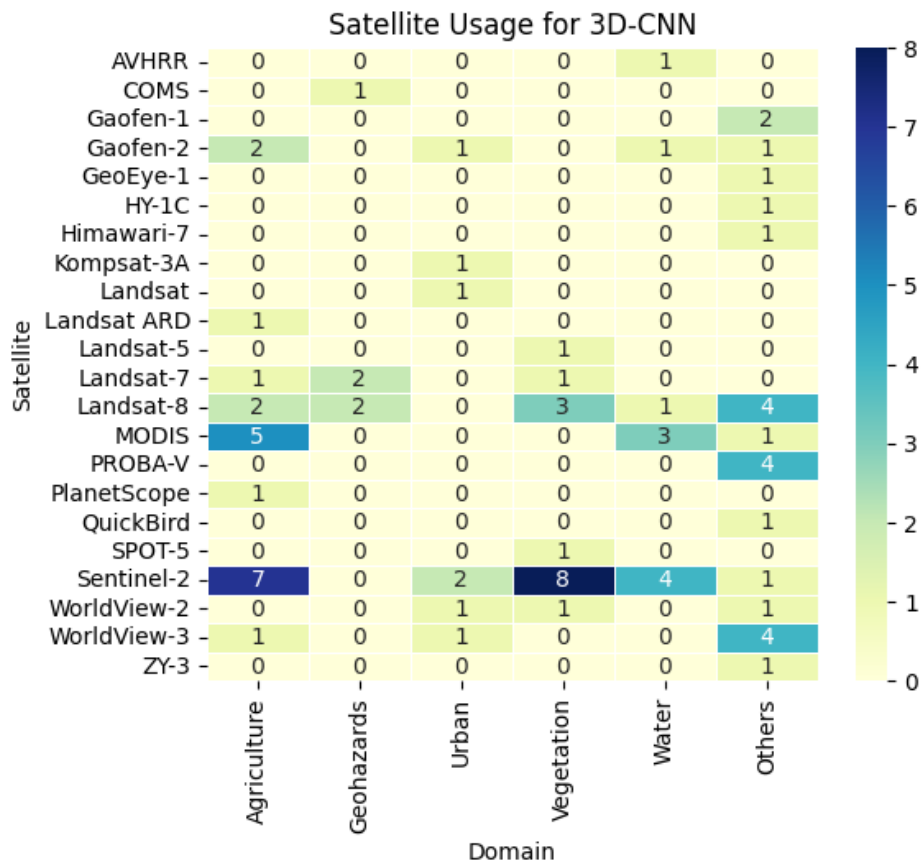


Figure 4.11. Satellite usage across different domains for 3D-CNN

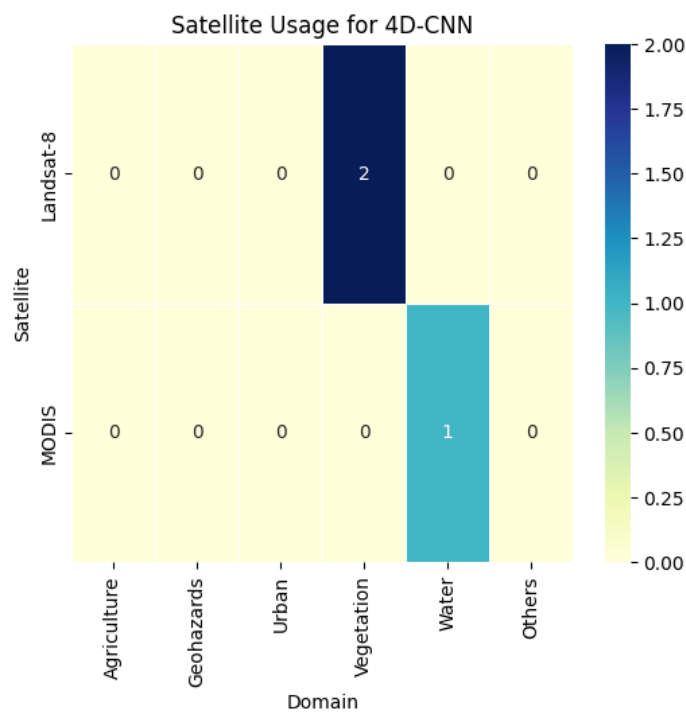


Figure 4.12. Satellite usage across different domains for 4D-CNN

tions as well, highlighting the particular suitability of Sentinel-2 data for agricultural studies using various CNN architectures.

In this systematic review, from each publication is extracted the type of the machine learning techniques used and metrics related to that technique. The distribution of publications using machine learning techniques classification, regression and segmentation over different CNNs is shown in Figure 4.13. There are also included compression and image enhancement, which are not traditionally categorized as machine learning techniques. These are techniques related to image processing that often use CNNs to achieve tasks and in this systematic review these techniques are used in publications related to 3D-CNN. The findings indicate that classification is the most frequently used machine learning technique, with more than 30 publications for 1D, 2D, and 3D CNNs. Regression is identified as the second most used machine learning technique, while segmentation is the least used. Since classification is the focus of most publications, further analysis of metrics for classification-based publications is conducted. Insufficient publications related to regression prevent a constructive analysis, and the lack of available data for metrics makes the sample too small for any concrete conclusions. Consequently, a full dataset for regression and segmentation metrics, where available, is provided and analyzed in the sections below.

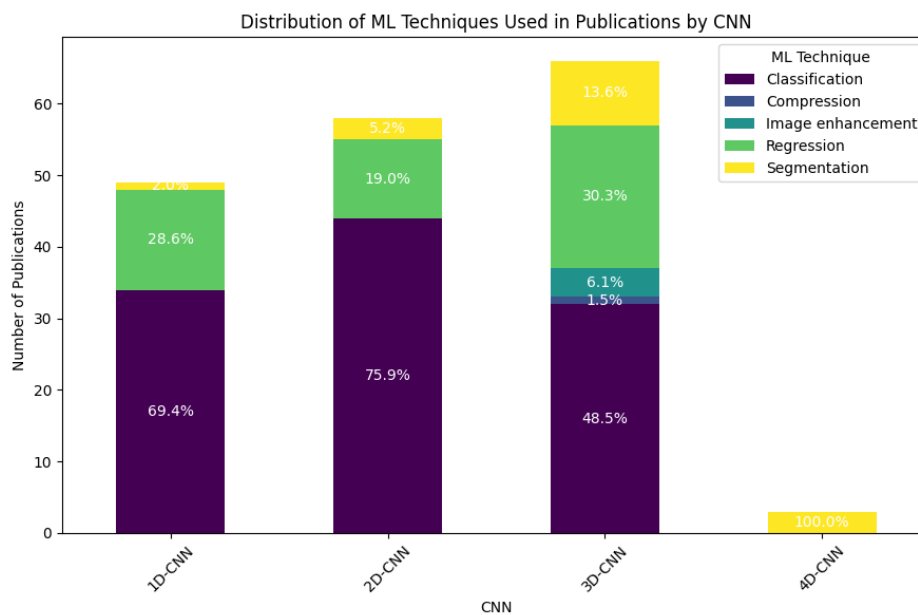


Figure 4.13. Number of publications using different machine learning techniques by each CNN

Analysis of Publications Implementing Classification

Classification involves examining the connections among a group of 'objects' to determine whether the data can be accurately summarized by a limited number of categories representing similar objects [121]. From the perspective of remote sensing imagery, it is a process

of categorizing pixels in an image of raw satellite data to obtain a given set of labels. There are different types of classification, of which the most popular are supervised and unsupervised [122]. In this systematic review, the types of classifications used in publications were not analyzed, only performance metrics accuracy and F_1 -score.

Accuracy is a measure that refers to the total data accurately predicted by the trained classifier when tested on unseen data. It ranges from 0 to 1, or in percentage terms from 0% to 100%, where accuracy closest to 1 or 100% suggests that the classifier is more accurate [123]. Figure 4.14 shows a box and whisker plot pointing the central tendency, dispersion, and potential outliers of accuracy for different CNNs degrees over different domains.

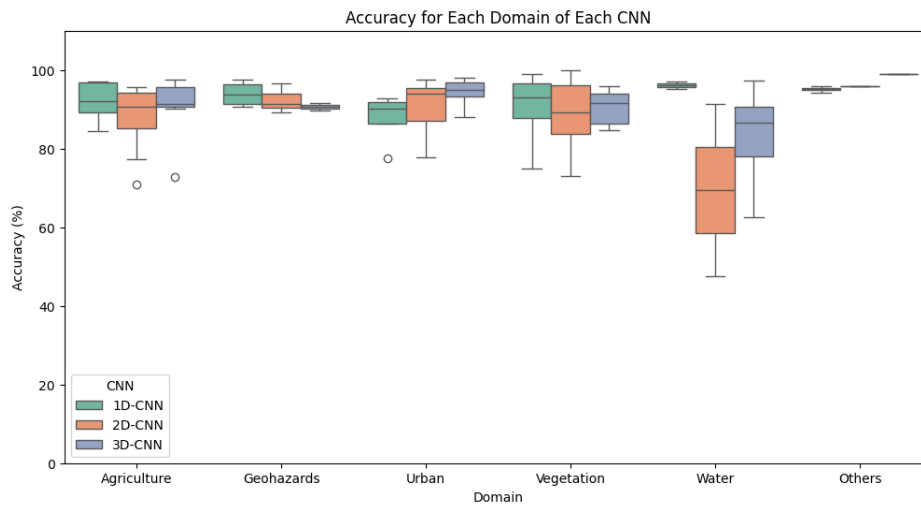


Figure 4.14. Accuracy of CNNs using classification across different domains

It can be seen that the 1D-CNN tends to achieve an accuracy of around 90% in the examined publications over different domains. The median accuracy values for the 1D-CNN in each domain are: 'Agriculture' – 92.00%, 'Geohazards' – 93.78%, 'Urban' – 90.28%, 'Vegetation' – 93.00%, 'Water' – 96.17%, and 'Others' – 95.18%. For the 'Urban' domain, the plot shows slightly negative skewness, suggesting that there are some publications with less than 90% accuracy, which is further confirmed by one publication falling outside the whisker. For the 'Agriculture' domain, the plot shows slightly positive skewness, indicating that there are several publications with higher accuracy compared to other publications. Examining the accuracy of the 2D-CNN for each domain, it can be noticed that the boxes for 2D-CNN across domains are taller compared to those for the other two CNNs. This suggests that there is greater variability in the accuracies reported in publications for the 2D-CNN compared to the other CNNs. The median accuracy values for the 2D-CNN in each domain are: 'Agriculture' – 90.59%, 'Geohazards' – 91.38%, 'Urban' – 94.09%, 'Vegetation' – 89.18%, 'Water' – 69.44%, and 'Others' – 95.94%. For the 'Agriculture' and 'Urban' domains, the plot shows negative skewness, and in the 'Agriculture' domain, one dot is outside the whisker, indicating that one publication falls outside the range. Positive skewness is observed in the 'Geohazards' domain, indicating that most accuracies are concentrated

on the lower end of the value range, but there are a few high-value accuracies pulling the distribution tail to the right. Accuracies of 3D-CNN do not have median accuracy values around 90% for all domains as was the case for 1D-CNN. Instead, the median accuracies are 91.30% for 'Agriculture', 90.59% for 'Geohazards', 95.00% for 'Urban', 91.57% for 'Vegetation', 86.53% for 'Water', and 98.87% for 'Others'. Moreover, it shows slightly negative skewness in the 'Vegetation' and 'Water' domains, while it exhibits positive skewness in the 'Agriculture' domain, with one accuracy as an outlier.

Based on the available median accuracy values, 1D-CNN proved to be the best in the 'Water' domain with a median accuracy of 96.17%. In contrast, in the 'Water' domain, 2D-CNN achieves a median accuracy of only 69.44%. In the 'Geohazards', and 'Urban' domains, 2D-CNN achieves higher median accuracy compared to the other CNNs. For the 'Vegetation' domain, all three CNNs show high median accuracy values. Based on the results, 1D-CNN and 2D-CNN can be used for problems that require consistent performance across all domains, while 3D-CNN is recommended for solving problems that require high accuracy in the 'Agriculture' and 'Others' domains.

The F_1 -score is the harmonic mean of the precision and recall metrics. Precision is the ratio of correctly predicted positive observations to the total number of predicted positive observations. Recall is the ratio of correctly predicted positive observations to all observations in the actual class. The best value achieved by the F_1 -score is 1 (perfect precision and recall) and the worst value is 0 [124]. Figure 4.15 displays a box and whisker plot of the F_1 -score metric for each CNN and application domain. Due to the unavailability of F_1 -score data for every individual publication, generalizing conclusions for all application domains to avoid bias is not feasible. However, domain-level analysis is conducted where data are available for all three CNNs. In the Agriculture domain, the median F_1 -scores are 85.33% for 1D-CNN, 91% for 2D-CNN, and 90.08% for 3D-CNN. In the Geohazards domain, the median F_1 -scores are 93.96% for 1D-CNN, 83.66% for 2D-CNN, and 74.81% for 3D-CNN. In the Vegetation domain, the median F_1 -scores are 94% for 1D-CNN, 83.68% for 2D-CNN, and 83.21% for 3D-CNN. Based on these findings, we can single out 2D-CNN and 3D-CNN with F_1 -scores around 90% for 'Agriculture', suggesting that 2D-CNN and 3D-CNN are likely to capture spatial patterns in crop classifications. On the other hand, 1D-CNN showed high F_1 -scores (>90%) compared to 2D-CNN and 3D-CNN in the 'Vegetation' and 'Geohazards' domains. These findings need to be interpreted with caution because the F_1 -score data is imbalanced, preventing us from making a general conclusion.

Analysis of Publications Implementing Regression

Regression is a statistical method used to investigate the relationship between species and the environment, based on observations of species and environmental variables at a series of locations. In regression analysis, data on a particular species is analyzed separately and how it is related to environmental variables. The goal of regression analysis is to describe

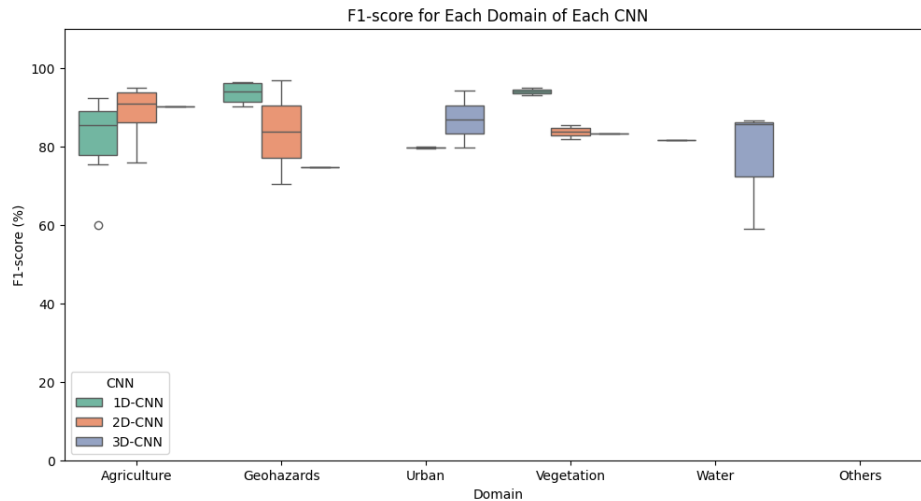


Figure 4.15. F_1 -scores of CNNs using classification across different domains

the response variable (abundance of species) as a function of one or more environmental explanatory variables. The response function cannot predict responses without errors, but efforts should be made to minimize those errors [125].

In this systematic review, certain metrics were extracted from publications used to evaluate models, such as R-squared, RMSE, and MAE. R-squared, also known as the coefficient of determination, determines the proportion of variance in the dependent variable that can be explained by the independent variables. It provides information on how well the observed values match the predicted values and can be expressed as a value or percentage, ranging from 0 to 1 or 0% to 100%. A value closer to 1 or 100% is desirable, indicating a better fit between the observed and predicted values. RMSE or the Root Mean Squared Error is the square root of the mean of the squares of all the errors. It indicates how close the line of best fit is to the set of points. MAE or Mean Absolute Error provide the average value of the absolute difference between the observed value and the predicted values. MAE or Mean Absolute Error provides the average value of the absolute difference between the observed value and the predicted values [126].

Since regression predicts the value of the response variable for specific species, a general analysis was not possible. For example, publications were grouped by application domains, but there was heterogeneity within the same domain (e.g., within the Water domain, parameters such as secchi depth and temperature were analyzed). Therefore, averaging metric values is not recommended as it could mask important application-specific differences and lead to misinterpretation, resulting in a lack of validity in the comparison. Tables 4.3, 4.4, and 4.5 highlight publications that have used regression with 1D, 2D, and 3D-CNNs, respectively. For each study, key information (e.g. publication type, year, satellite, and metrics such as R-squared, RMSE, and MAE) is presented and grouped by domain. Many values are missing. For those shown in Table 4.3, it can be seen that for the 1D-CNN in the 'Water' domain, the prediction of chlorophyll value expressed in milligrams per cubic meter is dominant in

the regression analysis. For publications related to 2D-CNN (Table 4.4), most studies are in the 'Others' domain, focusing on parameters related to image quality such as reflectance. As for the publications related to 3D-CNN (Table 4.5), in addition to the 'Others' domain, most of them were related to the 'Agriculture' domain. The studies mainly focused on predicting soybean and crop yield parameter values.

Analysis of Publications Implementing Segmentation

Image segmentation is a technique defined as the process of dividing or partitioning an image into homogeneous parts, called segments. This is particularly useful for applications such as image compression or object detection, as processing the entire image for these types of applications is inefficient. Therefore, image segmentation is used to segment parts of the image for further processing [127]. According to [128], in remote sensing, it is often viewed as a tool for detecting landscape changes and land use or land cover classification. In Table 4.6 all publications were comprised related to segmentation for different types of CNN grouped by domains. It can be seen that even on this really small representative sample of sixteen publications, the most publications were in domain agriculture and vegetation, which is in accordance to [128]. It is interesting that the most of these publications were using Landsat-8 data for segmentation.

Moreover, following metrics were extracted: Producer Accuracy, User Accuracy, Accuracy, F_1 -score, Precision, Recall, Kappa, IoU, MIoU and Dice. Accuracy, F_1 -Score, Precision and Recall have same definition as for classification. Producer's and User's Accuracies are commonly used in segmentation tasks. Producer's accuracy indicates how well a training set pixel is classified for a given coverage type. User's accuracy indicates the probability that a pixel classified as belonging to a certain class actually represents that class on the ground. The Kappa coefficient indicates how much the classification is better compared to a classification where each pixel is randomly assigned a class value [129]. The Dice coefficient is a metric used to compare the similarity of two samples. The Dice coefficient is twice the overlapping area of the two segmentations divided by the total number of pixels in the two images. Intersection over Union (IoU) measures the overlapping area between the predicted segmentation and the true segmentation, representing the overlapping area divided by the union area of the predicted segmentation and the true segmentation. Mean Intersection over Union (MIoU) is the average of the IoU values calculated for each class [130].

Table 4.3. 1D-CNN publications that implement regression

Domain	Study	Publication Type	Year	Satellite	R-squared	RMSE	MAE	Parameter
Agriculture	Jeong et al. [131]	Journal	2022	MODIS	0.86	0.61	N/A	Rice yield (Mg/ha)
	Sabo et al. [132]	Journal	2023	MODIS	N/A	N/A	N/A	N/A
Geohazards	Xu et al. [133]	Journal	2023	MODIS	0.907	0.31	N/A	Drought
Urban	Vulova et al. [134]	Journal	2021	Landsat-8	0.824	25	N/A	Reference evapotranspiration (mm/h)
Vegetation	Zhou et al. [135]	Journal	2021	MODIS	N/A	N/A	N/A	Rice yield (Mg/ha)
	Fathi et al. [136]	Journal	2023	Sentinel-2	0.745	6.085	4.895	Soybean yield (Bu/Ac)
Water	Maier et al. [137]	Journal	2021	MODIS	0.624	19.3	14.6	Chl-a (mg/m^3)
	Fan et al. [138]	Journal	2022	SeaWIFS	0.874	8.178	1.487	Chl-a (mg/m^3)
				MERIS	0.915	12.523	1.465	
				MODIS	0.894	6.727	1.448	
	Mukonza et al. [139]	Journal	2022	Landsat-8	0.93	0.15	N/A	Temperature (degree)
				Sentinel-3	0.91	0.2	N/A	
	Salah et al. [140]	Journal	2023	Sentinel-2	N/A	4.12	1.06	Chl-a (mg/m^3)
	Ivanda et al. [141]	Journal	2023	Sentinel-3	0.89	2.3	1.4	Secchi (m)
Zeng et al. [142]	Journal	2023	MODIS	0.874	18.968	1.494	Chl-a (mg/m^3)	
Salah et al. [143]	Conference paper	2023	Sentinel-2	N/A	11.87	6.71	Chl-a (mg/m^3)	
Others	Ojaghi et al. [81]	Journal	2023	Sentinel-3	0.97	3	N/A	Reflectance

* Mg/ha: Metric Tons per Hectare; Bu/Ac: Bushels per Acre; mg/m^3 : milligrams per cubic meter

Table 4.4. 2D-CNN publications that implement regression

Domain	Study	Publication Type	Year	Satellite	R-squared	RMSE	MAE	Parameter
Agriculture	Sagan et al. [144]	Journal	2021	WorldView-3 PlanetScope	N/A	N/A	N/A	N/A
	Nejad et al. [102]	Journal	2023	MODIS	0.73	6.25	5.09	Crop yield (Bu/Ac)
	Sabo et al. [132]	Journal	2023	MODIS	N/A	N/A	N/A	N/A
Geohazards	Lee et al. [145]	Journal	2020	COMS	0.94	8.32	6.09	Maximum sustained wind speed (kts)
Vegetation	Fathi et al. [136]	Journal	2023	Sentinel-2	0.758	N/A	N/A	Soybean yield (Bu/Ac)
Water	Zhong et al. [146]	Journal	2022	Sentinel-2	0.9	1.03	N/A	Bathymetry (m)
Others	Papadomanolaki et al. [147]	Journal	2021	WorldView-2 IKONOS	N/A	N/A	N/A	N/A
	Zhu et al. [111]	Journal	2022	Landsat-8 MODIS	0.9906	0.0258	N/A	Reflectance
	Zhang et al. [148]	Journal	2022	QuickBird	0.78	1.059	N/A	Similarity of spectral features
	Zhao et al. [112]	Journal	2023	Landsat-8 HY-1C	0.7921	N/A	N/A	Atmospheric correction
	Wang et al. [149]	Journal	2023	MODIS Sentinel-2 Landsat-8	N/A	N/A	N/A	N/A

* Bu/Ac: Bushels per Acre; kts: a knot is a unit of speed equal to one nautical mile per hour

Table 4.5. 3D-CNN publications that implement regression

Domain	Study	Publication Type	Year	Satellite	R-squared	RMSE	MAE	Parameter
Agriculture	Terliksiz et al. [150]	Conference paper	2019	MODIS	N/A	0.81	N/A	Soybean yield (Bu/Ac)
	Qiao et al. [151]	Journal	2021	MODIS	0.755	0.755	N/A	Crop yield (MT/ha)
	Sagan et al. [144]	Journal	2021	WorldView-3, PlanetScope	N/A	N/A	N/A	N/A
	Nejad et al. [102]	Journal	2023	MODIS	0.78	5.93	4.39	Crop yield (Bu/Ac)
	Wang et al. [152]	Conference paper	2023	MODIS	N/A	5.468	N/A	Crop yield (kg/ha)
	Wang et al. [153]	Journal	2023	MODIS	N/A	5.33	N/A	Crop yield (kg/ha)
Geohazards	Lee et al. [145]	Journal	2020	COMS	0.8856	11.34	8.65	Maximum sustained wind speed (kts)
Vegetation	Fernandez-Beltran et al. [154]	Journal	2021	Sentinel-2	0.9526	107.26	N/A	Crop yield (kg/ha)
Water	Fei et al. [155]	Journal	2022	AVHRR	N/A	0.352	0.2641	SST (degrees)
	Wang et al. [156]	Journal	2023	MODIS	N/A	N/A	N/A	N/A
Others	Molini et al. [113]	Conference paper	2019	PROBA-V	N/A	N/A	N/A	N/A

Table 4.5. (Continued)

Domain	Study	Publication Type	Year	Satellite	R-squared	RMSE	MAE	Parameter
Others	Chen et al. [157]	Journal	2020	GeoEye-1, WorldView-2	N/A	N/A	N/A	N/A
	Zhang et al. [108]	Journal	2021	Gaofen-1, Sentinel-2	N/A	N/A	N/A	N/A
	He et al. [118]	Journal	2021	WorldView-3	N/A	N/A	N/A	N/A
	Zhang et al. [158]	Journal	2021	Himawari-7	0.865	0.314	N/A	Aerosol optical depth
	Ibrahim et al. [116]	Conference paper	2022	PROBA-V	N/A	N/A	N/A	N/A
	Igeta et al. [119]	Conference paper	2022	WorldView-3	N/A	N/A	N/A	N/A
	Zhu et al. [111]	Journal	2022	Landsat-8, MODIS	0.9906	0.0258	N/A	Reflectance
	Zhang et al. [148]	Journal	2022	QuickBird	0.7868	0.0975	N/A	Similarity of spectral features
	Zhao et al. [112]	Journal	2023	HY-1C, Landsat-8	N/A	N/A	N/A	N/A

* Bu/Ac: Bushels per Acre; MT/ha: metric tons per hectare; kg/ha: kilograms per hectare; kts: a knot is a unit of speed equal to one nautical mile per hour; SST: Sea Surface Temperature

Table 4.6. Publications that implement segmentation

CNN	Domain	Study Type Year	Satellite	Producer Accuracy	User Accuracy	Accuracy	F ₁ -score	Precision	Recall	Kappa	IoU	MIoU	Dice
1D	Others	Bahl et al. [159] C/2022	Landsat-8	N/A	N/A	93.65	88.98	93.79	84.65	N/A	N/A	N/A	N/A
2D	Urban	Ghandorh et al. [160] J/2022	Spot-7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	79.51	88.65
	Vegetation	Li et al. [161] J/2022	Gaofen-2	N/A	N/A	94.13	82.78	N/A	N/A	88.86	N/A	73	N/A
		Saralioglu et al. [162] J/2022	WorldView-2	N/A	N/A	95.6	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3D	Agriculture	Mohammadi et al. [163] C/2021	Landsat ARD	93.7	93.6	N/A	N/A	N/A	N/A	91.8	N/A	N/A	N/A
		Gallo et al. [164] J/2023	Sentinel-2	N/A	N/A	70.33	70.3	70.35	70.33	64	N/A	53.02	N/A

Table 4.6. (Continued)

CNN	Domain	Study Type Year	Satellite	Producer Accuracy	User Accuracy	Accuracy	F ₁ -score	Precision	Recall	Kappa	IoU	MIoU	Dice
		Mohammadi et al. [165] J/2023	Landsat-7, Landsat-8	89.4	90.8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Vegetation	Kalinicheva et al. [166] J/2020	Sentinel-2, SPOT-5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Meshkini et al. [167] C/2021	Landsat-5, Landsat-7, Landsat-8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Li et al. [161] J/2022	Gaofen-2	N/A	N/A	97.45	89.49	N/A	N/A	94.68	N/A	82.45	N/A
		Meshkini et al. [168] J/2022	Landsat-8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Saralioglu et al. [162] J/2022	WorldView-2	N/A	N/A	95.6	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 4.6. (Continued)

CNN	Domain	Study Type Year	Satellite	Producer Accuracy	User Accuracy	Accuracy	F₁-score	Precision	Recall	Kappa	IoU	MIoU	Dice
	Water	Wang et al. [120] J/2022	MODIS	N/A	N/A	64.29	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4D	Vegetation	Giannopoulos et al. [86] J/2022	Landsat-8	N/A	N/A	61.56	61.56	61.56	61.56	N/A	44.47	N/A	N/A
		Giannopoulos et al. [89] J/2022	Landsat-8	N/A	N/A	89.16	77.96	N/A	N/A	N/A	N/A	N/A	N/A
	Water	Wang et al. [120] J/2022	MODIS	N/A	N/A	71.43	N/A	N/A	N/A	N/A	N/A	N/A	N/A

* C – Conference paper; J – Journal

4.3.4. Data Ontology for CNN Architecture Applications in Remote Sensing Publications

According to [169], there are a number of different definitions of ontology, but it can be described as a formal and explicit representation of concepts within a certain domain, including classes (or concepts), their properties (attributes) and restrictions on these properties. Ontologies have proven to be a powerful tool for representing knowledge for a chosen field, integrating data from different sources, and supporting various semantic applications [170].

In this study, the ontology enables systematic organization of publication information extracted via the PRISMA methodology, standardizing data to facilitate consistent comparison and categorization of publications. The Protégé program (version 5.6.4) was used for creating the ontology. This tool allows easy modeling and visualization of ontology structures, which facilitates the organization and analysis of data, especially in complex areas such as publication analysis [171]. In this study, Protégé was used for creating the ontology, and WebVOWL [172] was used for the ontology visualization.

Figure 4.16 shows the ontology created for CNN architecture applications in remote sensing publications. The ontology consists of seven main classes: Publication, CNN, Dataset, ML_Technique, Domain, Metric, and Application. The Publication class stores bibliographic information (title, abstract, authors, year, etc.). The CNN class includes dimensional variations (1D to 4D-CNN) and connects to ML_Techniques through usesTechnique property. The Dataset class, identified by satellite name, links to CNNs through providesDataFor and to Domain class which enumerates specific areas (Agriculture, Geohazards, Urban, Vegetation, Water, Others). The Application class represents practical CNN implementations, connecting to CNN, Dataset, and Domain classes. The ML_Technique class includes Classification, Regression, Segmentation, Compression, and ImageEnhancement, each associated with specific metrics. The Metric class, with its subclasses (ClassificationMetric, RegressionMetric, SegmentationMetric), captures various performance measures. ClassificationMetric and SegmentationMetric share common evaluation metrics including accuracy, F1-score, precision, recall, kappa, pixel accuracy, user accuracy, and producer accuracy (all ranging from 0 to 1). SegmentationMetric additionally includes specific measures such as IoU, MIoU and Dice coefficient. RegressionMetric uses different evaluation measures: RMSE, MAE and R2Score.

The developed ontology facilitates systematic analysis of CNN architectures in remote sensing applications by establishing clear relationships between publications, their implemented architectures, and achieved results. This structured representation enables researchers to effectively compare different CNN approaches across various domains and applications, while maintaining consistent measurement of their performance through standardized metrics.

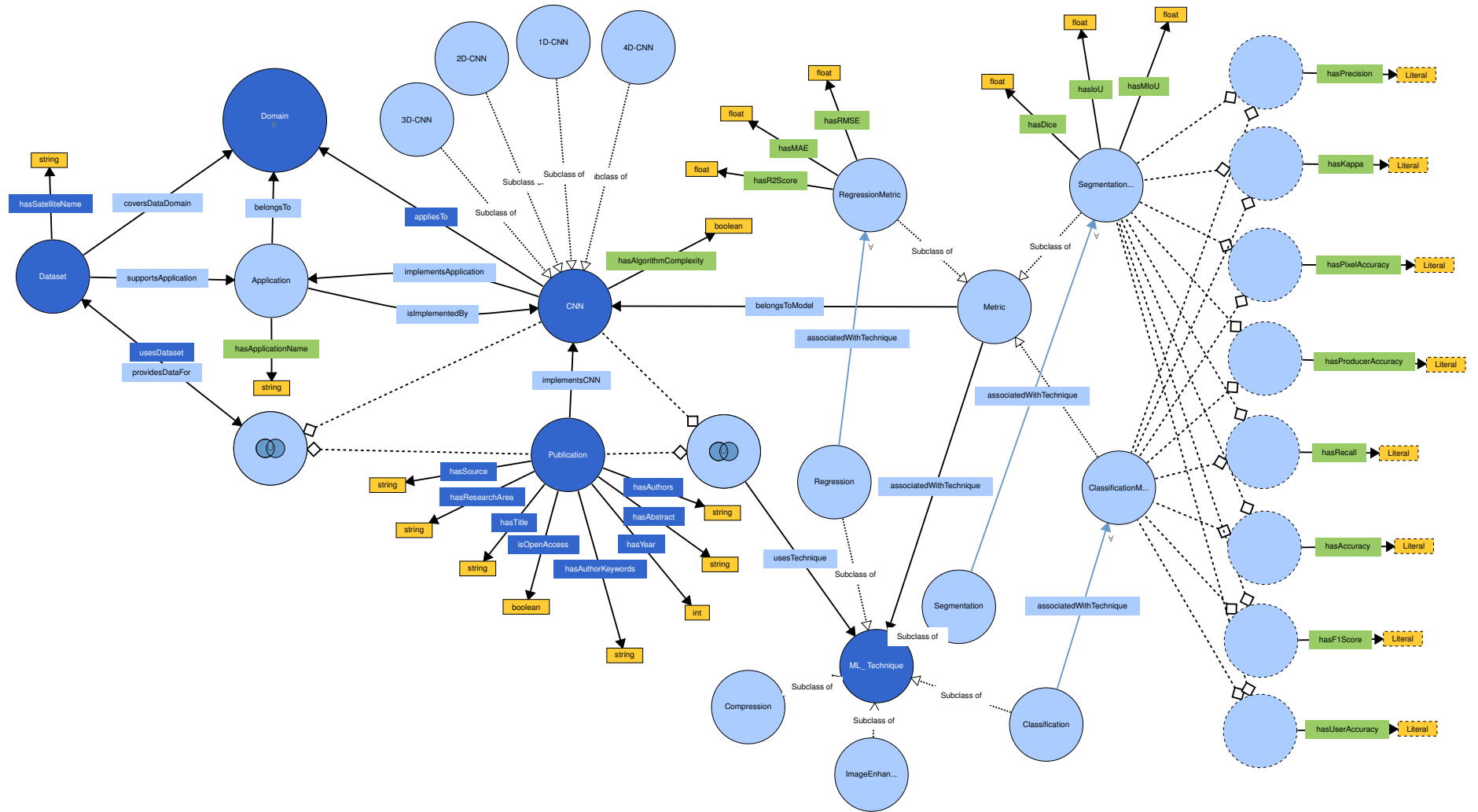


Figure 4.16. Ontology visualization generated with WebVOWL

4.3.5. Limitations and Future Work

This systematic review covers a wide range of keywords related to the topic in three different databases, so it may have missed some relevant literature. For example, studies that do not explicitly mention the term 'remote sensing' but only the name of the satellite could have been overlooked. Moreover, non-English studies were excluded, potentially affecting the geographic diversity of included publications. Additionally, publications that were not open access or inaccessible could not be further examined to extract relevant information. Due to the large number of studies, specific knowledge gaps within each satellite, domain, CNN, and machine learning technique are not extensively explored. Future work could narrow its focus to allow for a more detailed assessment of knowledge gaps, such as by concentrating on a single CNN model within a specific domain. Additionally, an evaluation of the risk of bias or quality of the reviewed studies should be included.

4.4. Conclusion

This research investigated the trends in the application domain of CNNs, the type of satellite data and employed machine learning techniques, with an emphasis on understanding how the degree of convolution affects model outcomes in relation to the characteristics of specific remote sensing tasks. Most of the papers were published in peer-reviewed journals and used a 3D CNN implementation. In addition, the majority of publications on 1D, 2D and 3D CNNs focused on solving problems in the field of agriculture and using Sentinel-2 satellite data. However, due to the limited sample size (three publications), no definitive conclusions could be drawn for 4D CNNs.

The predominant machine learning technique used by CNN is classification. In the 'Urban' domain, 3D-CNN achieved the highest accuracy (95.00%), followed closely by 2D-CNN (94.09%). The 'Water' domain showed interesting variation, with 1D-CNN demonstrating the best performance (96.17%), while 2D-CNN had notably lower accuracy (69.44%). Due to the small number of publications and the diversity of problems within the domains addressed by CNNs for the machine learning regression technique, no general conclusions could be drawn. The same applies to the publications that used the machine learning segmentation technique, as there were only 16 studies.

This systematic review, supported by ontological knowledge organization, identified a significant research gap: no previous studies have compared the performance of 1D, 2D, 3D, and 4D convolutional neural networks (CNNs) on the same or similar tasks. This emphasises the need for comprehensive benchmarking studies to evaluate these CNN variants on representative tasks using the same dataset. Such effort is crucial for understanding the strengths and limitations of each CNN type and for guiding future applications in this field.

5. EVALUATION OF 1D-CNN MODEL

This chapter describes the development and evaluation of a 1D convolutional neural network (1D-CNN) specifically designed for the estimation of Secchi disk depth (Z_{SD}) using multispectral Sentinel-3 OLCI images as a representative *regression analysis*. This work contributes to developing new methods for predicting spectral phenomena in remote sensing by utilizing the capabilities of CNNs. Although Z_{SD} is typically a point measurement, applying the model to satellite imagery demonstrates its potential to monitor changes over larger areas and overcome the limitations of traditional methods that are manual, costly, and dependent on human effort and capabilities.

This chapter is written based on paper [141] and is organized into several sections to evaluate the performance and applicability of the 1D-CNN model for predicting Secchi disk depth. It begins with an overview of recent studies on Secchi disk depth and the application of remote sensing in water quality analysis. This is followed by a detailed explanation of the materials and methods, including the study area, the creation of the dataset – a combination of official measurements and citizen contributions – and the model architecture. The results section presents the quantitative performance of the 1D-CNN model and comparisons with other regression algorithms, as well as the parameter kd_{z90max} , which is often used as a proxy for Secchi disk depth, obtained using the Case-2 Regional CoastColour (C2RCC) algorithm. The chapter concludes with a discussion of the accuracy of the model, its limitations, its potential application in other regions and its implications for long-term water quality monitoring.

5.1. Introduction

The Secchi disk depth (Z_{SD}) is a quantitative measure that indicates the 'transparency' or 'clarity' of the water in a body of water such as a lake or sea. Secchi disk depth is the depth at which a circular disk, known as a Secchi disk, is no longer visible when lowered into the water [173]. It serves as an important indicator of water quality as it reflects the amount of suspended particulate matter (SPM) present, which includes materials such as algae, clay, silt and others. In general, the clearer the water, the deeper the Secchi disk can be seen [174]. The relationship between Z_{SD} and SPM is complex because SPM is influenced by several factors, including Z_{SD} , chlorophyll-a (Chl-a) and colored dissolved organic matter (CDOM).

While Z_{SD} can indicate potential water quality issues, it cannot serve as a direct indicator for measuring SPM. For a comprehensive assessment of water quality, it is necessary to analyze additional parameters such as Chl-a, CDOM and SPM [175]. However, measuring these parameters requires specialized instruments and additional laboratory analysis, making the process time-consuming and costly. Nevertheless, these water quality parameters in combination with Secchi depth provide a more detailed understanding of water quality and can help identify specific sources of pollution or other water quality issues. In addition, Secchi depth is often used in research studies dealing with aquatic ecology and the functioning of aquatic ecosystems. It enables the study of the impact of human activities on water bodies, such as pollution, nutrient loading and land use changes. In agricultural areas, for example, excessive nutrient input from fertilizers can lead to strong algae growth and a decrease in water clarity. In addition, Secchi depth is valuable for assessing the effectiveness of water management strategies and identifying areas where water quality improvements are needed [176]. Overall, the combination of Secchi depth and other water quality parameters contributes to a more comprehensive and informed understanding of water quality conditions, enabling better water quality management and protection measures.

Water quality assessment traditionally involves analyzing the physical, chemical and biological properties of field samples in a laboratory. Waters are often categorized as Case-1 or Case-2 based on their optical properties. Case-1 waters are mainly found in open oceans and stratified shelf seas, while Case-2 waters include coastal areas and estuaries influenced by suspended sediments and colored dissolved organic matter [177–179]. It is worth noting that the classification boundaries between Case-1 and Case-2 are not always clear [180].

Earth observation has proven to be a valuable tool [181] to meet the challenges of water management. Improvements in computer technology and applications have enabled remote sensing techniques to monitor and detect changes on a large scale, which is not possible with traditional *in situ* measurements. Remote sensing enables the monitoring and assessment of water quality indicators such as Chl-a [182, 183], turbidity [183, 184], total suspended matter (TSM) [185], Secchi disk depth (Z_{SD}) [186–188], CDOM [189], etc. In addition, it is possible to estimate certain parameters by measuring another related parameter. For example, a correlation can be established between Secchi depth and turbidity [190]. *In situ* Measurements of turbidity can be measured with instruments such as nephelometers or spectrophotometers [191].

Three approaches to interpreting 'water color' are described by [192]: empirical, semi-empirical, and analytical. Empirical methods use statistical relationships and spectral properties to estimate parameters from remotely sensed data [193]. For instance, nonlinear models based on $K_d(\text{PAR})$ and $1/Z_{SD}$ have been developed [194–196]. Semi-empirical methods are often used to study the optical properties of Case-2 waters, combining physical and spectral information to create algorithms that relate to measured water parameters. However, the complexity of Case-2 waters, especially in inland turbid lakes, can sometimes cause these

algorithms to fail [197]. Analytical methods, on the other hand, estimate the concentration of water constituents based on absorption and backscatter coefficients, using reflected light from remotely sensed data [198]. Efforts have been made to link Secchi depth to PAR or visible solar radiation (VSR) [199–203], but measuring these parameters is often expensive and time-consuming [202].

Machine learning methods have shown promise in estimating the optical properties of water bodies, as discussed by [204]. Thus, [205] evaluated the clarity of Brazilian inland waters by applying various machine learning algorithms, including Random Forest, Extreme Gradient Boosting and Support Vector Machines, together with semi-analytical algorithms for querying Z_{SD} using Sentinel-2 imagery. Similarly, [206] and [207] used random forest regression on Sentinel-3 OLCI and Landsat-4, -5, -7, and -8 imagery to estimate Z_{SD} .

While machine learning algorithms have proven successful on various environmental problems, they often require large amounts of high-quality data to perform well. Official measurements may not be sufficient for achieving desired performance. To overcome this challenge, citizen science is becoming increasingly popular. In this approach, the public is encouraged to collect data using simple, low-cost devices [208]. Programs such as LAKE-WATCH [209] and the Secchi Disk Project [210] have been launched to harness citizen science. These initiatives can generate significant amounts of data for more frequent and focused observations [211]. Recent research suggests that data collected by volunteers can be as reliable as official data [209, 212]. However, satellite remote sensing makes it possible to observe larger areas. Although existing satellite products generally align with measured data [213], there is still a need for calibration or the development of new products using different data sets. The validation of Landsat-8 Z_{SD} products using official and citizen science data emphasises the need for large data sets to achieve a more accurate Z_{SD} estimation [214]. Expanding data collection through the involvement of citizen scientists can therefore improve the calibration of remote sensing products, leading to better environmental monitoring and management.

Recent studies have highlighted significant progress in using deep learning to predict Secchi depth, with promising results in estimating water quality parameters. These studies show that deep learning, particularly CNNs and recurrent models, can effectively capture complex relationships between remote sensing reflectance (R_{rs}) data, optically active constituents (OACs), and non-OACs to retrieve accurate water quality information [215–219]. For point parameters like water quality indicators (e.g., Chl-a, Z_{SD}), it is appropriate to use the spectral vector of an individual spatial pixel to identify a specific parameter value. A 1D-CNN is an effective method for this, as it processes raw spectral vectors and outputs deep spectral features for classification [96]. Additionally, 1D-CNNs have been successfully applied to detect water quality parameters such as Chl-a concentration in inland water bodies [137, 220], classify points (sea-land) using bispectral bathymetric echo [221], and classify land use and land cover in hyperspectral imagery using spectral signatures [79].

5.2. Materials and Methods

5.2.1. Study Area

Figure 5.1 shows the study area of the northern Adriatic Sea, which is located near the coasts of Croatia and Slovenia. The Adriatic Sea is a part of the Mediterranean Sea between the Balkan and the Apennine Peninsula extending southeast to the Otranto Sill, where a water exchange with the Mediterranean Sea takes place [222]. The maximum depth of the Adriatic Sea is 1233 m, while the average is 259.5 m. The depth of the sea increases from the north-western basin, which is only 15 m deep, to the south-eastern basin, where the depth reaches 780 m [223].

The Adriatic Sea is famous for its natural beauty and rich biodiversity, which is why it is often the destination of many tourists during the summer months. It is also home to many plant and animal species and thus provides food for the population living along the coast and is suitable for the development of numerous industries. Traditionally, it is considered a clean and unpolluted sea. Nevertheless, it is important to continuously monitor the status and trends of marine quality in these waters. Timely detection of deterioration in sea quality can prevent negative impacts on the health of the people and species living in the sea and protect biodiversity.

This study area was selected for several key reasons. Regular monitoring activities by local research institutions provide extensive *in situ* measurements and historical data records, providing a robust dataset for model validation. Furthermore, the geographical position of the northern Adriatic Sea ensures consistent coverage by Sentinel-3 OLCI satellite imagery throughout the year. The regional expertise available through collaboration with local marine research institutes and environmental monitoring agencies further supports the comprehensive analysis of water quality parameters in this area.

5.2.2. Secchi Disk Depth

The Secchi disk is a common instrument used to measure the clarity of water by determining the depth at which the disk disappears from view. It is a simple and effective method in which a white disk is lowered into the water and the depth at which it is no longer visible is recorded. The depth at which the Secchi disk is visible is measured in meters (m).

Secchi disk measurements were collected from three different sources. For the Gulf of Trieste, the data came from the National Institute of Biology, Marine Biology Station in Piran. Measurements for the Croatian coastline were sourced from Legal entity for water management - *Hrvatske vode* and the Secchi Disk Project website [210], which includes volunteer-contributed data. The measurements from the National Institute of Biology and *Hrvatske vode* followed standard protocols for official monitoring, ensuring their reliability

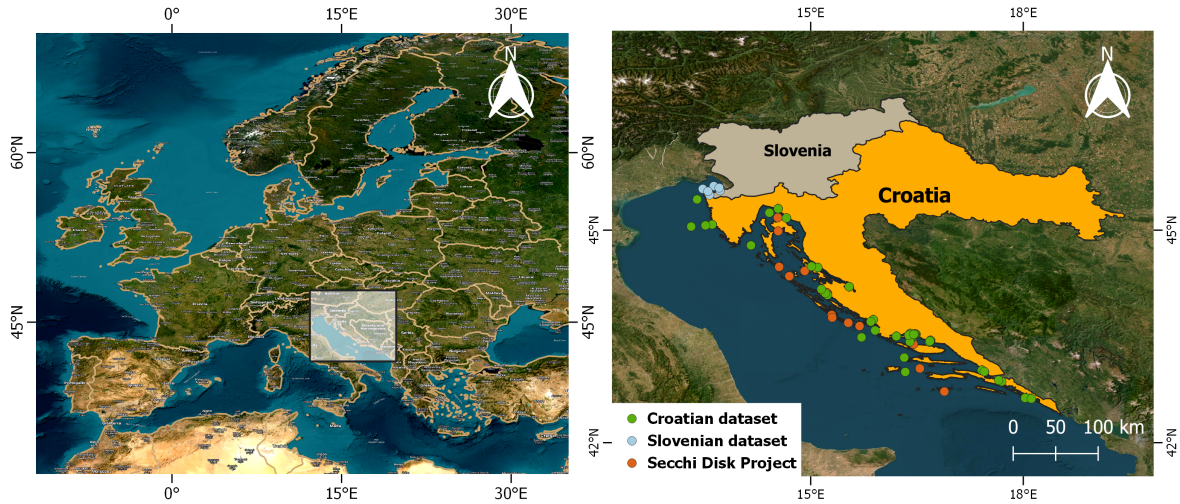


Figure 5.1. Map of the Adriatic Sea showing the locations of Secchi disk depth measurements (shown in Plate Carrée projection; adapted from [141])

and accuracy. It is important to note that the Secchi disk measurements of the Secchi Disk Project were provided by volunteers and that the specific measurement conditions for these data points are not available. Nevertheless, the volunteer dataset shows a similar range and distribution to the official measurements, so it was used without modification. Not all measurements were directly included in the study. Satellite data processing techniques were used to ensure the quality and accuracy of the dataset. The data points were adjusted based on the satellite data when clouds were detected over the measurement locations to reduce the influence of cloud interference. This was necessary due to the inability to control the conditions of the *in situ* measurements.

The spatial distribution of all *in situ* measurements, which comprise a total of 589 measurements between May 2016 and September 2021, is shown in Figure 5.1. This figure shows the locations of the monitoring sites, with some sites having multiple measurements taken during the monitoring period. The Croatian dataset consists of 452 measurements, the Slovenian dataset of 118 measurements and the Secchi Disk Project dataset of 19 measurements. Figure 5.2 illustrates the probability distribution of the Z_{SD} values from the Croatian, Slovenian and Secchi Disk Project datasets. This distribution helps to understand the range, concentration and variability of Z_{SD} values within each dataset.

In addition, Figure 5.3 shows the statistical details of the *in situ* measurements for each year. This includes the number of samples (count), the minimum and maximum Z_{SD} values presented as vertical black lines (whiskers), the standard deviation (σ) which indicates how much the Z_{SD} values vary from the average, and the mean value which gives the average Z_{SD} across the dataset shown as a blue horizontal line .

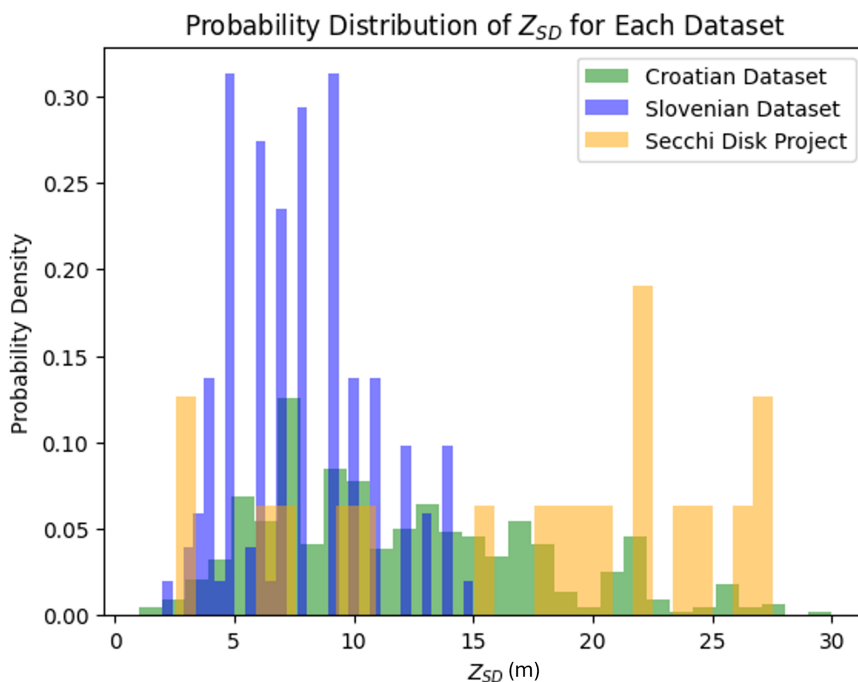


Figure 5.2. Probability distribution of the measured Z_{SD} values in the Croatian, Slovenian, and Secchi Disk Project datasets (Adapted from [141])

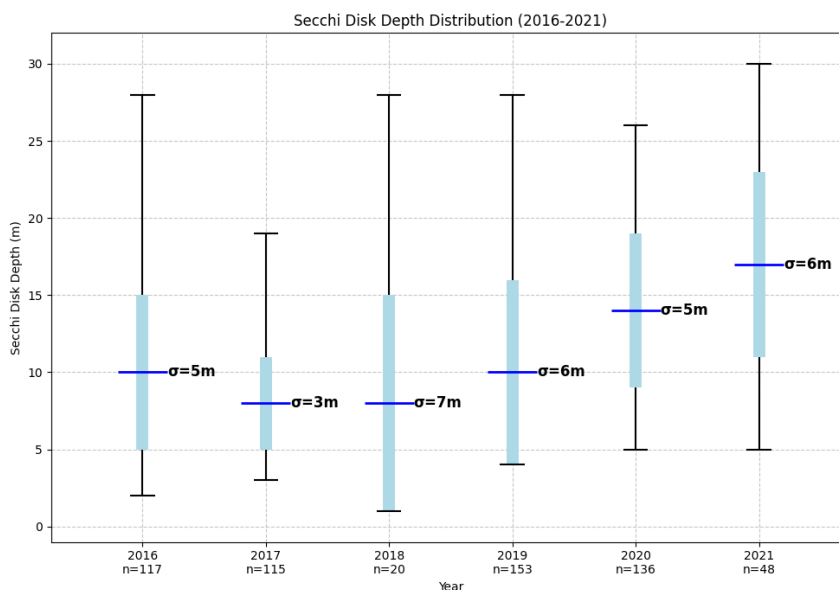


Figure 5.3. Statistical characteristics of in situ Z_{SD} (m) through years

5.2.3. Sentinel-3 OLCI Data

The Sentinel-3 mission, jointly operated by the European Space Agency (ESA) and the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), provides regular ocean and land observation services. This mission includes the Sentinel-3A satellite, which was launched on February 16, 2016, and the Sentinel-3B satellite, which was launched on April 25, 2018. The main instruments on these satellites are the Ocean and Land Colour

Instrument (OLCI), the Sea and Land Surface Temperature Radiometer (SLSTR) and the Synthetic Aperture Radar Altimeter (SRAL). OLCI, an imaging spectrometer with 21 spectral bands ranging from 390 nm to 1040 nm, offers a spatial resolution of 300 meters and revisits the same location every day, making it suitable for studying the Adriatic Sea. OLCI is widely used to study the open oceans and coastal waters, measuring sea surface topography, temperature and color [224].

For this study, 185 Sentinel-3 OLCI Level-1B TOA (Top Of Atmosphere) satellite images were used for the period from May 2016 to September 2021. The images were downloaded via the Sentinel Hub (Credit: Modified Copernicus Sentinel data 2016-2021/Sentinel Hub) [225] [226] using Python. They were selected based on the data from *in situ* Secchi disk depth measurements, analyzing only images from these specific dates and locations. Cloud-based filtering was applied to ensure high data quality.

The Sentinel Hub EO Browser processed the images and converted TOA radiance to TOA reflectance by correcting for atmospheric effects such as Rayleigh scattering, absorption and emission by atmospheric gasses and aerosol effects. This process provided the TOA reflectance values that were used to estimate the Secchi disk depth. The normalized reflectance was calculated using the formula (5.1) from the Copernicus Sentinel-3 OLCI Land User Handbook [227]:

$$R_{TOA} = \pi * (L_{TOA} / E / \cos(\theta)) \quad (5.1)$$

where R_{TOA} represents TOA reflectance, L_{TOA} is TOA radiance, E stands for solar irradiance and θ is the solar zenith angle. For this study, the reflectance values of 21 TOA Sentinel-3 OLCI bands were extracted for each *in situ* measurement. Instead of applying atmospheric correction algorithms such as POLYMER [228], SeaDAS [229] or C2RCC [32] to all 185 different scenes (each scene contains 21 .tiff images of Sentinel-3 OLCI bands), the raw reflectance data from Sentinel-3 OLCI were used directly. Performing atmospheric corrections on such a large dataset would complicate the automated execution of the methodology. In addition, the dynamic and changing conditions of the coastal atmosphere, especially in regions such as the Adriatic Sea, pose a challenge for standard atmospheric correction methods. The authors in [230] point out that no single atmospheric correction algorithm is consistently superior to others in coastal areas. The L1 satellite data was used without applying atmospheric correction, allowing the CNN model to learn and account for atmospheric effects on its own. This approach leverages the ability of machine learning algorithms to handle raw reflectance values directly, bypassing the need for explicit atmospheric correction during preprocessing [231].

5.2.4. Dataset Construction and Preprocessing

To ensure the satellite dataset's quality, the Sentinel-3 OLCI images were pre-processed. Cloud effects in the dataset were removed using band Oa17. Band Oa17 has a wavelength of 865 nanometers in the near-infrared and can be used for atmospheric, aerosol and cloud correction as well as for co-registration of pixels [227, 232]. The typical range of Sentinel-3 OLCI reflectance values for a given band is 0-0.4, however highly reflective surfaces like clouds, can have reflectance values above 1 [226]. The satellite scenes of the study area were examined with the QGIS tool and it was found that the band Oa17 has values above 0.17 for land and cloud pixels. Figure 5.4 shows the mask created by applying a threshold value of 0.17 to the satellite image of the Oa17 band for the area of Kaštela Bay and Brač Channel in Croatia. The yellow color indicates pixels that have a value greater than 0.17, i.e. pixels related to land and clouds. The blue color indicates pixels that have a value of 0.17 or less, i.e. pixels related to the sea. To maintain data diversity in the final dataset, rows in dataset with Oa17 band values greater than 0.17 were retained. Instead of removing them, they were assigned a Secchi depth value of zero, which means that the depth of the Secchi disk below the cloud cannot be estimated.

It was observed that cloud shadows over land could cause some land pixels to be incorrectly classified as sea pixels. Cloud shadows are darker due to the higher proportion of diffuse solar radiation, making them difficult to distinguish from dark surfaces with similar spectral signatures like water [233]. This issue, along with atmospheric correction, is left for the proposed model to learn and address independently to achieve reliable Secchi depth predictions under these constraints.

Finally, each row of the constructed dataset for training and testing is defined as follows:

Oa1	Oa2	Oa3	...	Oa21	Secchi
0.185	0.171	0.141	...	0.025	16

where *Oa1* to *Oa21* represent the band values for each location corresponding to an *in situ* Secchi measurement value, labeled as *Secchi*. To ensure accurate match-ups between satellite and field measurements, the data were aligned based on the same sampling date and the latitude/longitude coordinates of the measurement locations. Although the exact time of the *in situ* measurements was not available, match-ups were made using the measurement date. This approach allowed the integration of satellite band values (*Oa1* to *Oa21*) with the corresponding Secchi depth measurements, forming the final dataset for training and testing purposes.

Figure 5.5 depicts the spectral values of Sentinel-3 OLCI TOA reflectance for various *in situ* Secchi depth measurements. The plot compares the spectral reflectance for the maximum Secchi depth (30 m), the minimum Secchi depth (2 m) under clear skies, and the minimum Secchi depth (1 m) under cloud cover. The spectral curves for the maximum and minimum

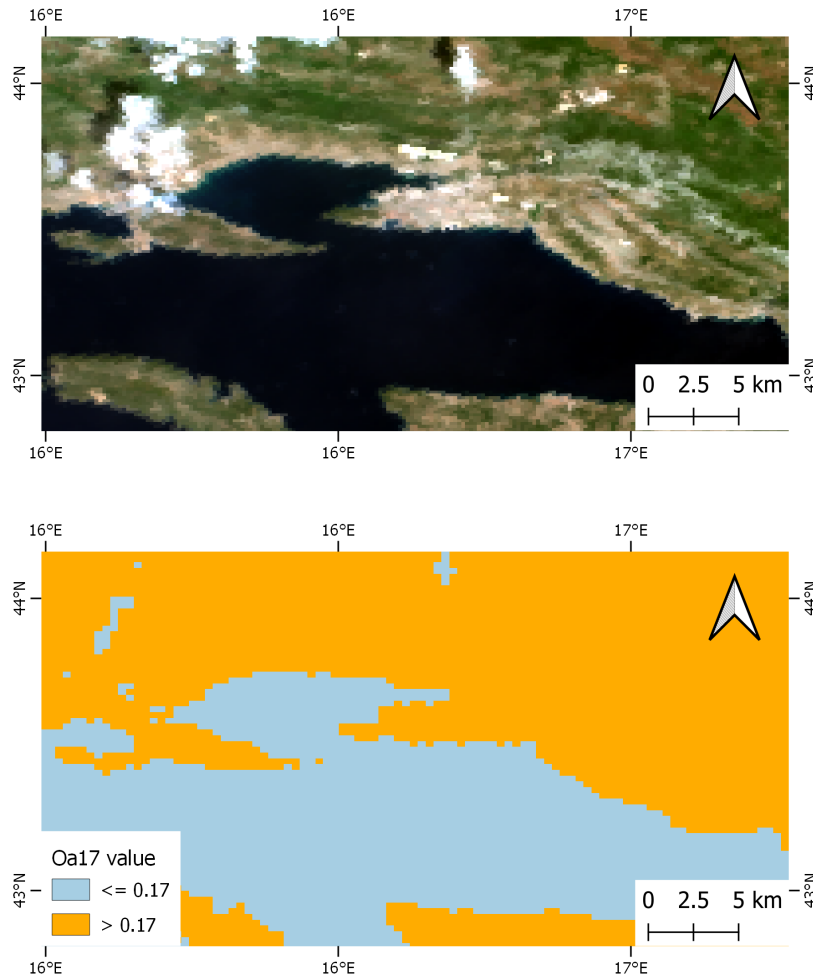


Figure 5.4. Cloud and land mask based on band Oa17 values for the area of the Kaštela Bay and Brač Channel (Croatian coast) on August 2, 2020 (Adapted from [141])

depths in cloud-free conditions follow similar patterns but differ in magnitude, particularly in the visible range of the electromagnetic spectrum. However, the curve for the minimum Secchi depth under cloudy conditions shows significantly higher spectral responses, especially in the infrared region where the Oa17 band is located, which is often used to detect clouds.

5.2.5. 1D-CNN Architecture

Convolution is a mathematical operation used in image processing to extract features from images using certain filters. This study applies 1D convolution to one-dimensional data, such as time series represented by 21 values corresponding to the surface reflectances of 21 bands. This method focuses on the spectral response of the surface pixels. The hypothesis is that the relationship between the different band reflectances captured by the convolution encodes the clarity of the water as measured by Secchi depth. Additional information about CNNs are provided in Section 3.

Figure 5.6 shows a state-of-the-art 1D-CNN architecture developed for the prediction of

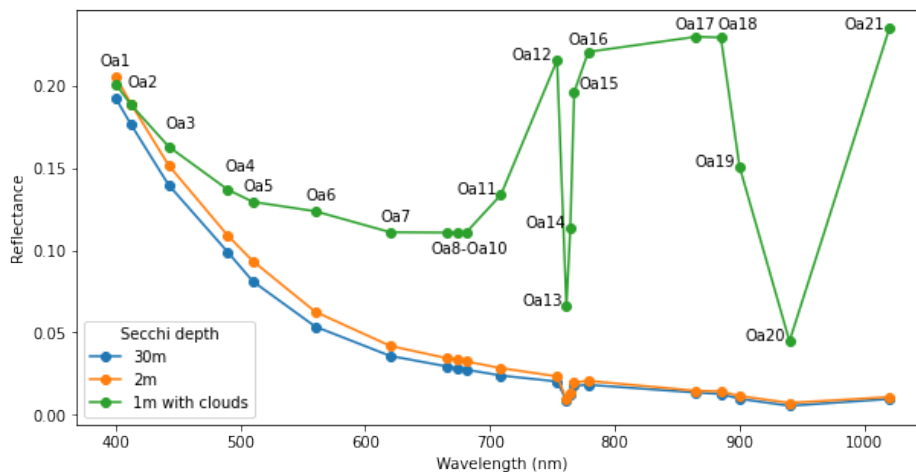


Figure 5.5. One dimensional data example – Sentinel-3 OLCI TOA reflectances corresponding to in situ Secchi values (Adapted from [141])

Secchi disk depth values. This architecture consists of three convolutional layers, a dropout layer, a max-pooling layer and two dense layers. To prevent overfitting, a dropout layer is included that randomly deactivates nodes during training [234]. The Rectified Linear Unit (ReLU) activation function is used in the hidden layers. Unlike Sigmoid and Tanh functions, which compress input values to a specific range, ReLU outputs the input directly if it is positive and zero if it is negative. This nonlinearity allows the model to capture complex relationships and avoid the vanishing gradient problem, resulting in faster and more stable training [235]. The proposed 1D-CNN model was implemented using the Python programming language version 3.8.16 [236] and its libraries Keras (version 2.9.0) [237] and Tensorflow (version 2.9.2) [238].

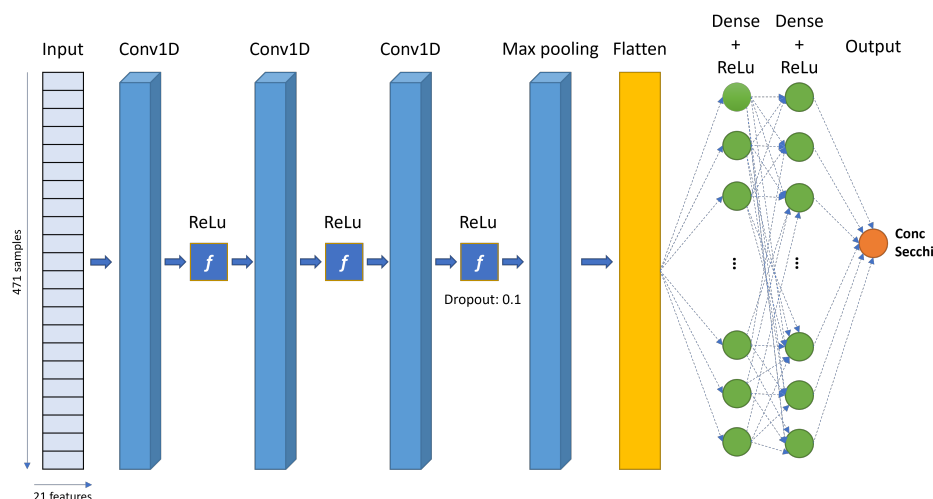


Figure 5.6. Proposed 1D-CNN architecture for the Secchi disk depth prediction (Adapted from [141])

5.2.6. Hyperparameter Tuning

The architecture of the model is set up with hyperparameters that can be tuned during training. These hyperparameters define the architecture of the model, including the number of convolution layers, the kernel size (the size of the convolution filter), and the number of filters. In addition, the training parameters such as the number of epochs, batch size and the learning rate can be selected to improve the model and find the optimal model. The creation of the model also requires the selection of a loss function and an optimizer. Building the model also involves selecting a loss function, which measures the prediction error during training and is defined here as the mean square error (MSE). The optimizer is an algorithm used to find the best learnable parameters to minimize the loss function, and the one used here is Adam [239].

To achieve the best performance, it is important to optimize the settings or hyperparameters of the model. The set of hyperparameters proposed in [240] was used for this tuning process. Figure 5.7 shows the combination of hyperparameters that led to the best results for the 1D-CNN model. Tuning these hyperparameters is often based on trial and error, as the optimal values can vary depending on the dataset and problem. Even small changes, e.g. in the learning rate or batch size, can affect the convergence of the model and its final performance. The chosen hyperparameters were crucial for good results in predicting Secchi disk depth from satellite spectral data. The 1D-CNN model has three hidden convolutional layers (Conv1D) that are designed to capture important spectral features and dependencies. These Conv1D layers recognize patterns in the spectral data and allow the model to extract information needed for accurate predictions. To further process and learn non-linear relationships between the extracted features and the target output, two hidden dense layers were added to the model. The number of hidden layers in a neural network is important for modeling complex relationships in the data. More layers can increase the model's ability to capture complicated patterns, but the effects vary depending on the problem. Adding more layers can improve accuracy in some cases, but it can also lead to overfitting. Finding the right balance between model complexity and generalization is the key to optimal results. The final 1D-CNN model contains 17 009 learnable parameters.

For modeling the 1D-CNN, the data was randomly split into training, validation, and test sets. This random distribution ensured that the samples were evenly divided among the sets, with 60% for training (355 samples), 20% for validation (117 samples), and 20% for testing (117 samples).

5.2.7. Accuracy Assessment

According to [241], the accuracy assessment measures data quality and helps map users to assess how useful a thematic map is for their specific needs. The accuracy assessment involves three main components:

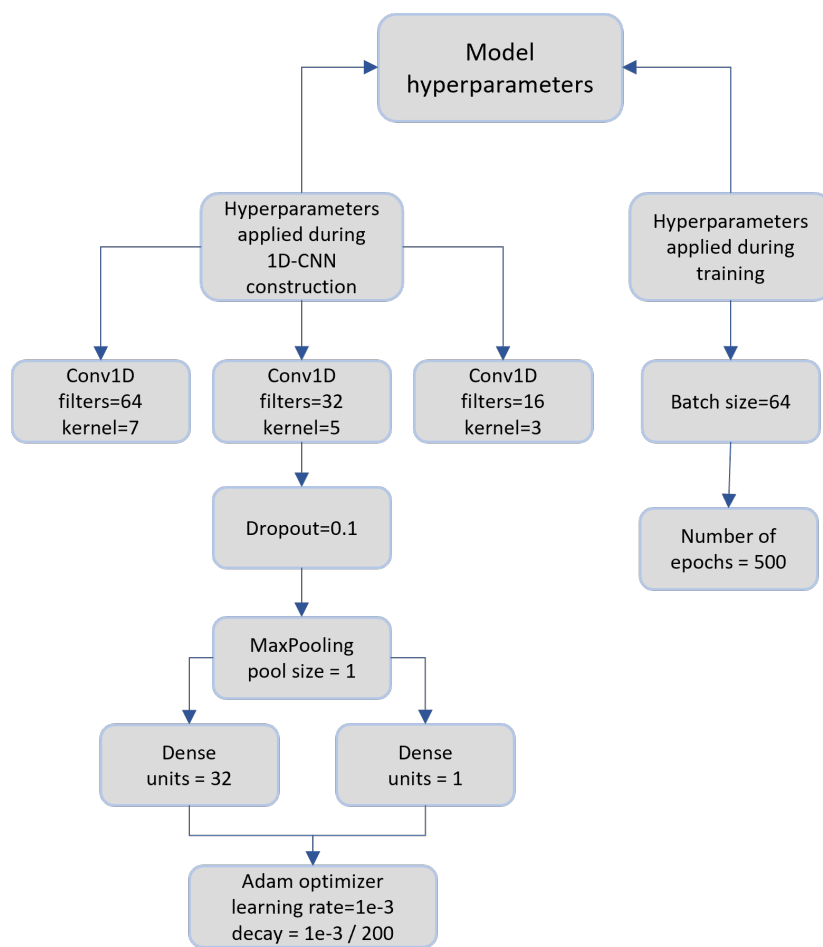


Figure 5.7. Hyperparameters used in the 1D-CNN model (Adapted from [141])

- Sample design – the protocol for selecting reference sample units. In this study, the Secchi disk value serves as the ground truth data for the classification of remote sensing imagery and represents the point sampling unit.
- Response Design – the protocol for determining the land- cover classification for each sampling unit. In this case, the response design for each Secchi sampling unit is a single pixel of the Sentinel-3 OLCI sensor, which has a spatial resolution of 300 × 300 meters. The sampling unit is a specific location on the ground corresponding to the center of the pixel, and the measurement is the Secchi value assigned to that pixel based on the remote sensing data.
- Analysis and estimation protocol – this protocol compares the estimated continuous data with the ground-truth data (analysis) and uses the results to make inferences about the entire population (estimation).

To assess how accurately the 1D-CNN model predicts the continuous variable Secchi, the following analysis techniques and measures from [242] were used in this study:

- Mean Absolute Error (MAE) – measures the absolute differences between the predicted values (y_i) and the measured values (x_i), and takes the average of these differences where N is the total number of data points. A low MAE value indicates that the model is better at predicting the target variable [243]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - x_i| \quad (5.2)$$

- Relative Mean Absolute Error (RMAE) – a unit- and scale-independent version of the MAE, calculated by dividing the MAE by the mean value of the validation dataset [244]:

$$RMAE = \frac{MAE}{\bar{x}} \quad (5.3)$$

- Root Mean Squared Error (RMSE) – the square root of the average of the squared differences between the predicted (y_i) and the actual measured values (x_i). A low RMSE means that the model is more accurate in predicting the measured values [245, 246]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2} \quad (5.4)$$

- Relative Root Mean Squared Error (RRMSE) – a unit/scale-independent measure, calculated by dividing the RMSE by the mean of the validation dataset (\bar{x}). [244]:

$$RRMSE = \frac{RMSE}{\bar{x}} \quad (5.5)$$

- The coefficient of determination (R^2) is a statistical measure that shows how much of the variance in the dependent variable can be predicted from the independent variables in a linear regression model [247]. It indicates how well the model fits the data, with values ranging from 0 to 1; a higher R^2 means a better fit. The formula to calculate R^2 is as follows:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.6)$$

where y_i is the value of the i -th sample, \bar{y} is the mean value of the variable y and \hat{y}_i is the predicted value for the i -th sample

- Pearson's correlation coefficient (r) – measures the strength and direction of the linear relationship between two variables, with values ranging from -1 to 1. The correlation between the variables x and y is calculated using the following formula [248]:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5.7)$$

where y_i are the predicted values, x_i are the measured values and \bar{x} and \bar{y} are the mean values.

5.2.8. Overview of Regression Algorithms for Comparison with 1D-CNN

Another way to demonstrate the strong performance and effectiveness of the proposed 1D-CNN model is to compare it with other commonly used algorithms in GIS and remote sensing [249], including both traditional regression methods (such as linear regression and ridge regression) and algorithms adapted for regression tasks (such as Decision Tree regression, Random Forest regression, Support Vector Machine regression, and Naive Bayes regression).

- Linear regression is a simple algorithm that models the relationship between a dependent variable and one or more independent variables as a linear function. The goal is to find the best fit line that minimizes the difference between the predicted and actual values [250].
- Ridge regression is a type of linear regression that helps prevent overfitting by adding a penalty term to the cost function. This penalty limits the size of the regression coefficients, reducing variance and improving its generalization performance [251].
- Decision Tree regression is a non-parametric algorithm that recursively splits the data into subsets based on the values of the input variables. Each split corresponds to a node in a tree, and the goal is to find the sequence of splits that results in the lowest residual sum of squares [252].
- Random forest is an ensemble learning method that combines multiple decision trees where each decision tree in the forest is trained on a random subset of the data and features. The final prediction is based on the average or majority vote of the individual predictions [253].
- Naive Bayes is a probabilistic algorithm that uses Bayes' theorem to calculate the probability of each class given the input features. It assumes that input features are conditionally independent given the output variable and predicts the class with the highest probability [254].
- Support Vector Machine regression is a kernel-based algorithm that finds a hyperplane in a high-dimensional space with the maximum margin, which separates data points into different classes with minimal error [255].

These regression algorithms were implemented by using the scikit-learn library version 1.0.2 for Python [256]. This powerful library provides a wide range of tools for machine

learning and data analysis including a variety of regression methods. All implemented regression algorithms were applied to a dataset of 21 features and one observation (Secchi disk depth) and used default parameters defined in the scikit-learn library. The results of the analysis are presented in Section 5.3.2.

5.3. Results

In this section, the results of the proposed 1D-CNN model for predicting Secchi disk depth in the eastern Adriatic Sea are presented. A comparison with traditional machine learning regression methods is made to emphasize the superiority of 1D-CNN. Furthermore, based on the implemented model, a map was created showing the distribution of Secchi disk depth values throughout the study area and highlighting any observed spatial patterns or trends. For a comprehensive overview, Figure 5.8 is attached, which illustrates the entire development and evaluation process of the 1D-CNN model.

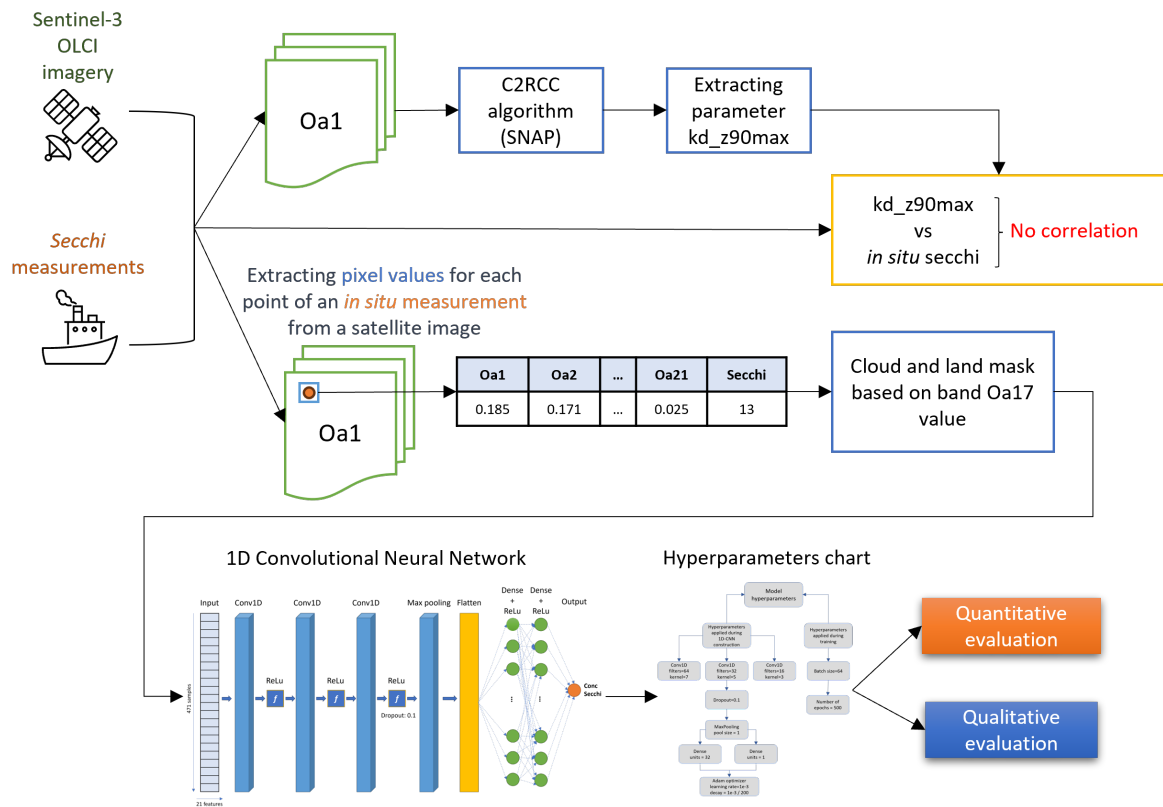


Figure 5.8. Flowchart of the 1D-CNN model development and evaluation process (Adapted from [141])

5.3.1. Quantitative Algorithm Performance

First, a quantitative evaluation of the performance of the 1D-CNN model in predicting Secchi disk depth is presented. The model was trained and evaluated on a system equipped

with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and 16 GB of RAM, with the complete training process requiring approximately 5 hours. Table 5.1 shows how the 1D-CNN model performed on two datasets with Slovenian and Croatian measurements. The Slovenian dataset contains *in situ* data from the National Institute of Biology, Marine Biology Station in Piran, while the Croatian dataset contains data from the Croatian water management authority - *Hrvatske vode*. The model based on volunteer data from the *Secchi Disk Project* is not included in the Table 5.1 because there were fewer than 20 data points, which is not sufficient for training and testing the 1D-CNN model. The *Final dataset* in the table contains all mentioned datasets.

The metrics in Table 5.1 were calculated for the training and test dataset, but not for the validation part of the dataset. The validation dataset was used during hyperparameter tuning and model selection to improve the performance of the model during training. The final evaluation of the model was performed on an unseen, independent test dataset that was not used for the training process. The detailed metrics used to evaluate accuracy (MAE, RMAE, RMSE, RRMSE, R^2 and r) are shown in Table 5.1 for the Slovenian, Croatian and Final datasets. These metrics were calculated for both training and test datasets, with the test dataset providing an unbiased assessment of the model's ability to generalize to new, unseen data. The results in this table show that including more data from different sources, even volunteered data without quality control, results in a model that generalizes better and predicts Secchi disk depth more accurately.

For the Slovenian and Croatian datasets, the coefficient of determination (R^2) ranges from 66% to 83% on the test dataset, suggesting that the model can explain a considerable percentage of the variance in the Secchi depth estimates. In the test dataset, the root mean square error (RMSE) is greater than two meters, indicating some prediction errors. The predictions of the model in the Slovenian and Croatian test datasets deviate on average by about 30% from the actual Secchi depth values, corresponding to the relative root mean square error (RRMSE).

The difference between the training and test performance metrics suggests that the 1D-CNN model may be overfitting the Slovenian and Croatian datasets. While the model performs well during training, with low RMSE and RRMSE values, its performance on new, unseen data (test dataset) is slightly less accurate, leading to higher errors. Despite this, the 1D-CNN model shows promise in predicting Secchi depth, as indicated by the high R^2 values (train: 0.894, test: 0.890) and relatively low MAE (train: 0.014, test: 0.014), RMSE (train: 0.024, test: 0.023), RMAE (train: 0.167, test: 0.167), and RRMSE (train: 0.288, test: 0.276) in the final dataset. These results suggest that the model is capable of explaining a substantial portion of the variance and making accurate predictions on unseen data.

Figure 5.9 provides a visual representation of the prediction errors of the model for the final test dataset. The graph shows the differences between the predicted and actual Secchi depth values in meters (m). The strong linear relationship between the predicted and

Table 5.1. Comparison 1D-CNN performance for different datasets (Adapted from [141])

Dataset		MAE	RMAE	RMSE	RRMSE	R ²	r
Slovenian measurement dataset	Train	0.012	0.241	0.019	0.361	0.815	0.904
	Test	0.012	0.239	0.022	0.499	0.666	0.844
Croatian measurement dataset	Train	0.015	0.187	0.020	0.305	0.915	0.929
	Test	0.021	0.193	0.028	0.325	0.830	0.850
The final dataset	Train	0.014	0.167	0.024	0.288	0.894	0.949
	Test	0.014	0.167	0.023	0.276	0.890	0.944

measured values is evident from the plotted line that closely follows the data points. This indicates a good correlation between the predictions of the model and the actual Secchi depth values and reflects a high accuracy. For a more accurate assessment of the prediction accuracy, statistical metrics such as the mean absolute error (MAE) and root mean square error (RMSE), as given in Table 5.1, should be considered.

The error histogram in Figure 5.10 illustrates the distribution of errors between the 1D-CNN model's predictions and the *in situ* Secchi disk depth measurements. The errors follow a Gaussian distribution centered around zero, indicating that the model's predictions are generally accurate and consistent across the dataset.

Figures 5.9 and 5.10 further demonstrate the model's strong performance in predicting Secchi disk depth for the final test dataset, which includes both Slovenian and Croatian measurements. The model shows low prediction errors, with a Root Mean Squared Error (RMSE) of 0.024 for the training data and 0.023 for the test data, indicating that it fits the training data well and generalizes effectively to new, unseen data. The Mean Absolute Error (MAE) is also low, at 0.014 for both datasets, meaning that the model's predictions are very close to the actual values on average.

The 1D-CNN model demonstrates a good fit with an R-squared value of 0.894 for the training data and 0.890 for the test data. The similarity in these R-squared values suggests that the model is not overfitting and can generalize well to new data. The strong correlation between the predicted and measured values is further supported by Pearson correlation coefficients of 0.949 for the training data and 0.944 for the test data.

Overall, the model performed well in terms of accurate predictions with low error and was able to generalize well to new data. The fact that the values of the individual metrics are similar for the training and test datasets indicates that the model does not overfit the training data and is able to generalize well to unseen data.

5.3.2. Performance Analysis of 1D-CNN and Commonly Used Regression Algorithms

Table 5.2 shows the results of the analysis and compares the accuracy metrics for each algorithm described in Section 5.2.8. The findings indicate that the regression algorithms were

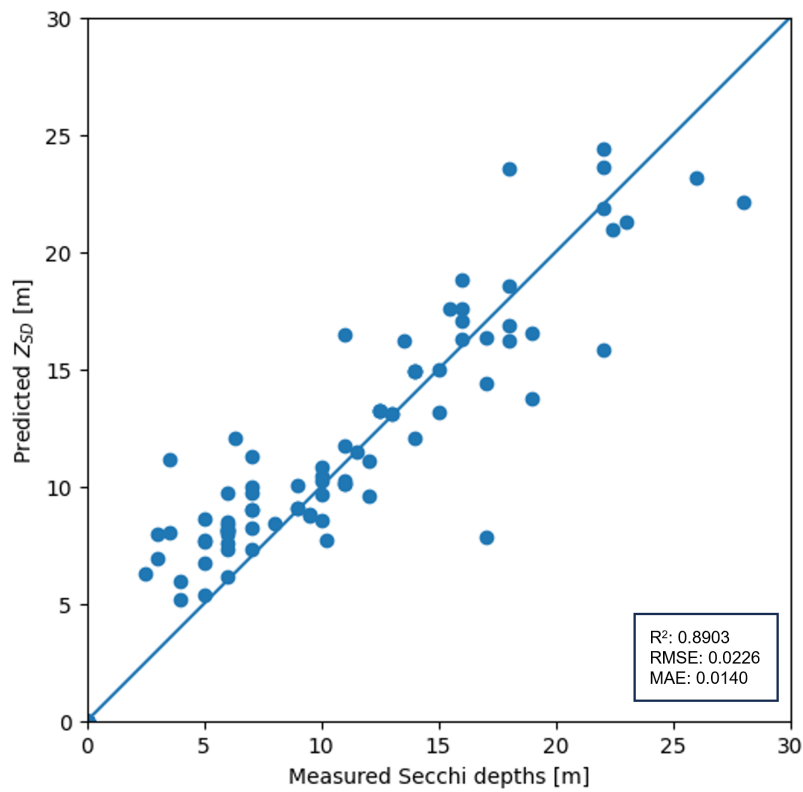


Figure 5.9. A prediction error plot shows the difference between the predicted values and the measured values of Secchi depth in meters (m) on the final test dataset (Adapted from [141])

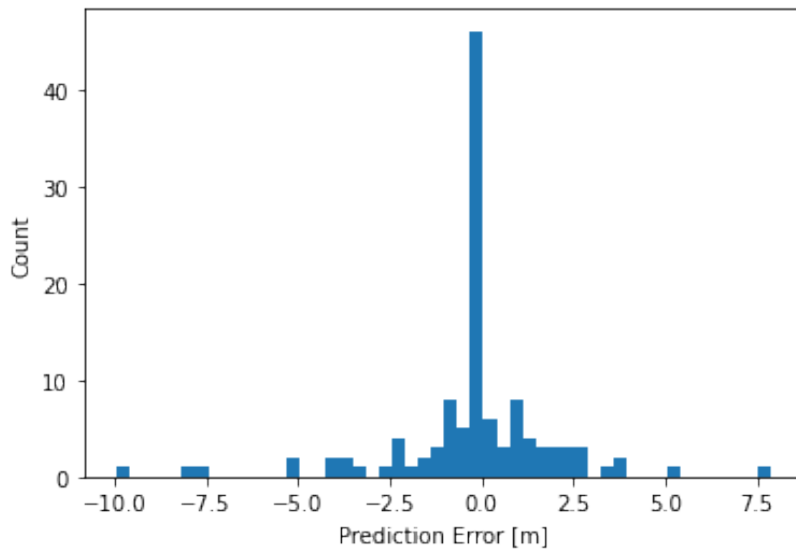


Figure 5.10. Histogram of errors of 1D-CNN model on the final test dataset (Adapted from [141])

less effective in predicting Secchi disk depth compared to the proposed 1D-CNN model, as shown in 5.1.

The Linear, RidgeCV and Bayesian Ridge Regression models have difficulty finding the complex relationships within the data. The MAE, RMSE and R^2 metrics suggest that they cannot effectively predict Secchi disk depth. The Decision Tree model outperforms on the training data with an R^2 of 0.999, while it fails on the test data with an R^2 of 0.597, suggesting that it became too complex during training and could not generalize to new data. The Random Forest model, which is an ensemble method and can usually improve prediction accuracy, could not outperform the 1D-CNN model. This could be due to the fact that it reaches its limits when capturing finer patterns in the data. Support Vector Regression performed particularly poorly with a negative R^2 in the test dataset, suggesting that it was difficult to fit to the data or was not the best choice for this particular problem. Overall, the 1D-CNN model performed better than the regression algorithms as it was able to extract meaningful features and accurately model the input-output relationship, especially given the complexity of predicting Secchi disk depth. This analysis highlights the importance of feature extraction and the ability of deep learning models to effectively process complex data patterns.

None of the regression models used to compare 1D-CNN could achieve an R^2 of more than 0.8, with the exception of the Decision Tree model but only for the training dataset. The RMSE value for both data sets (train and test) was greater than 3 m for all regression models, while the RMSE value of the 1D-CNN model was less than 2.5 m for both data sets. 1D-CNN is not only better correlated with actual measured values, but also predicts values with lower absolute error.

Table 5.2. Comparison of regression algorithms metrics (Adapted from [141])

Model	Dataset	MAE	RMAE	RMSE	RRMSE	R^2	r
Linear	Train	0.029	0.334	0.039	0.448	0.727	0.853
	Test	0.030	0.372	0.041	0.500	0.638	0.809
RidgeCV	Train	0.041	0.462	0.052	0.593	0.514	0.718
	Test	0.038	0.472	0.049	0.615	0.476	0.698
Decision Tree	Train	0.000	0.002	0.002	0.023	0.999	0.999
	Test	0.022	0.327	0.043	0.635	0.597	0.795
Random Forest	Train	0.029	0.315	0.044	0.482	0.643	0.802
	Test	0.029	0.417	0.044	0.673	0.584	0.769
Bayesian Ridge	Train	0.029	0.342	0.039	0.451	0.724	0.851
	Test	0.029	0.350	0.039	0.474	0.665	0.819
Support Vector Regression	Train	0.064	0.730	0.073	0.827	0.039	0.754
	Test	0.063	0.760	0.071	0.865	-0.106	0.745

5.3.3. The Spatial Distribution of Secchi Disk Depth

While *in situ* measurements are performed at specific locations depending on the sampling strategy and Z_{SD} presents discrete values for one location only, a model trained on *in situ* samples can be extrapolated to predict Z_{SD} for a large scene of the study area. This enables us to investigate the spatial distribution of Secchi disk depth in the Adriatic Sea using a 1D-CNN model applied to remote sensing imagery. Figure 5.11 shows a map of the study area created using the 1D-CNN model in Python and visualized with the open source software QGIS 3.10.10-A Coruña [257]. The results of the model show that the Secchi disk depth is greatest in the areas farther from the coast and lowest in the northern region and near the coastline. The reason for the lowest Secchi disk depth values in the northern region of the Adriatic is the Po, Adige and Isonzo/Soča rivers, which contribute to the low water transparency in this region. They influence the concentrations of dissolved nutrients, biological productivity, salinity and stratification in this part of the Adriatic Sea [258]. For this reason, the sea is very shallow with an average depth of 35 m. In addition, Figure 5.11 shows that the coastal areas surrounded by certain islands and sheltered regions such as bays also have lower water transparency. These areas are of particular concern for water quality, especially near mariculture facilities and bathing areas. Several numerical models incorporate Secchi depth into equations related to the die-off of pathogenic microorganisms [259–262]. In addition, a sensitivity analysis [263] has shown that Secchi depth is an important environmental factor in determining the decay of *Escherichia coli* in coastal waters. The improved spatial and temporal resolution provided by the 1D-CNN model will increase the accuracy of water quality modeling.

To prove the effectiveness of the proposed 1D-CNN model and overcome the limitations of the sparse and expensive *in situ* measurements, the model is compared with the well-known Case-2 Regional CoastColour (C2RCC) algorithm [32]. Figure 5.12 shows the map of the kd_z90max product created with version 1.0 of the C2RCC processor within the Sentinel Application Platform (SNAP) software. The kd_z90max product is an indicator of water clarity that is often used as a proxy for Secchi disk depth. It represents the depth of the water column from which 90% of the radiation emitted from the water is derived. The product kd_z90max corresponds to $1/K_d_min$, where K_d_min is the average attenuation coefficient of the irradiance in the three bands with the lowest K_d . The unit of measurement for kd_z90max is meter (m) [264,265]. In addition to kd_z90max , the C2RCC processor can also retrieve other important oceanic ecosystem parameters, such as chlorophyll concentration (*conc_chl*), total suspended matter (*conc_tsm*), and two parameters related to Gelbstoff (also known as yellow substances or colored dissolved organic matter, CDOM): the absorption of organic detritus and Gelbstoff at a wavelength of 443 nm (*iop_adg*), and the Gelbstoff absorption coefficient at the same wavelength (*iop_agelb*). These parameters provide valuable information about the composition and behavior of the ocean [265].

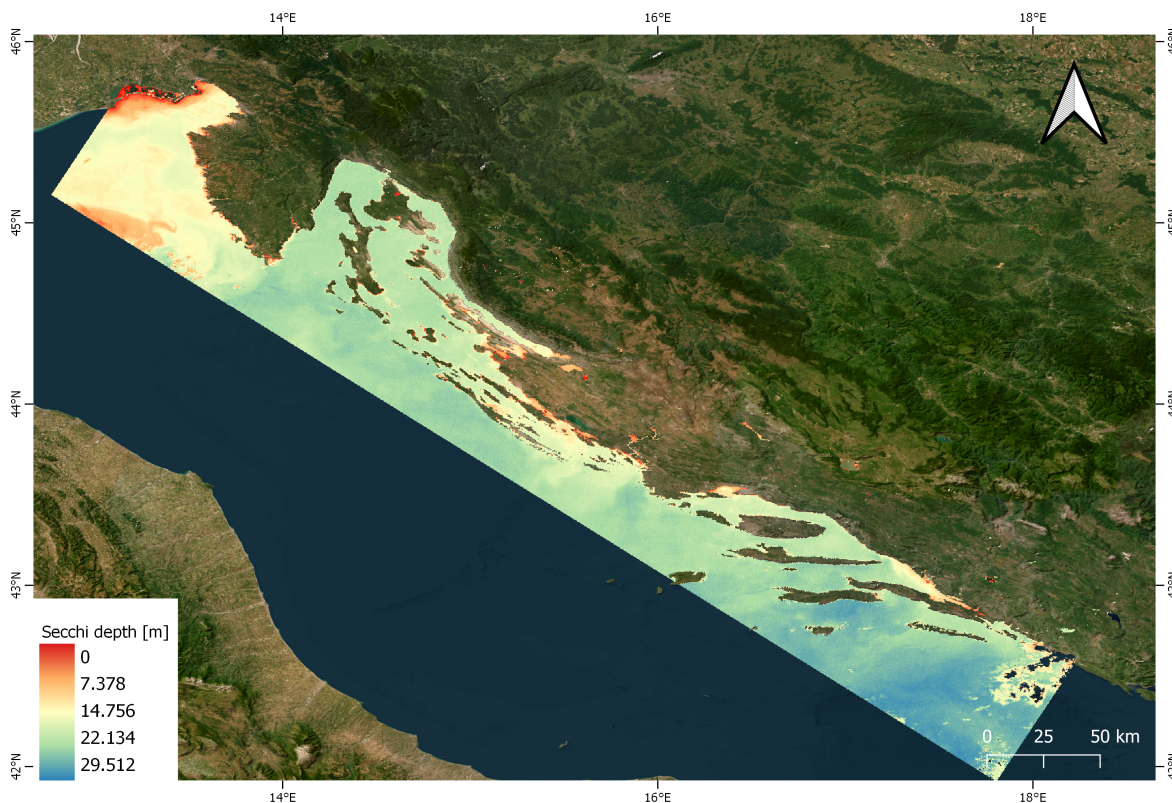


Figure 5.11. The distribution of Secchi disk depth in the study area on September 3, 2021 (Adapted from [141])

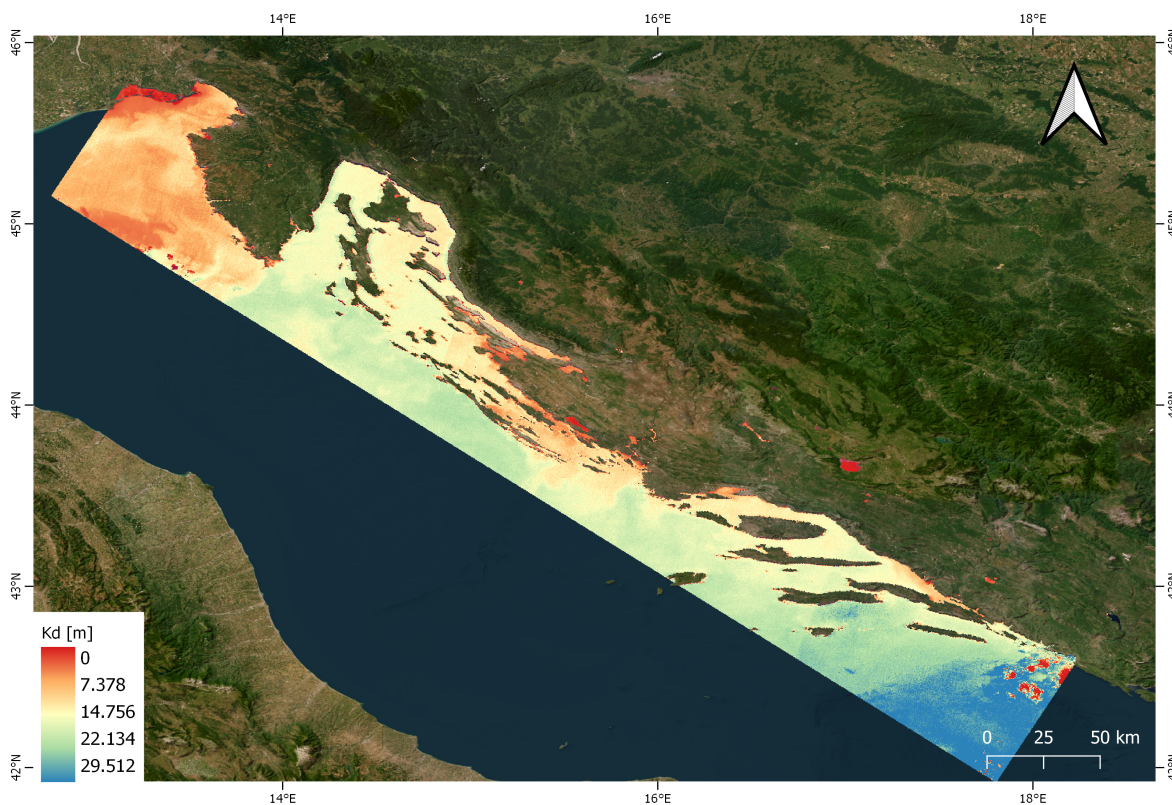


Figure 5.12. The distribution of kd_{z90max} in the study region on September 3, 2021, created using the C2RCC processor in SNAP (Adapted from [141])

The comparison between the maps of the spatial distribution of Secchi disk depth created with the 1D-CNN model (Figure 5.11) and the kd_{z90max} product created with the C2RCC processor (Figure 5.12) shows similar patterns of variation. The similarity can be seen in the regions of the northern Adriatic, the Zadar archipelago and Kaštela Bay. Both maps illustrate a consistent spatial variation of Secchi disk depth and kd_{z90max} , even though the absolute values of the parameters may differ in these regions.

The satellite images from September 3, 2021 were selected for comparing the Secchi disk depth measurements and the kd_{z90max} values due to optimal atmospheric conditions. Due to the limited availability of measurement data on this date, the closest available data points were selected. These points represented diverse environmental conditions, including different water turbidity values, depths, and distances from the coast, making them suitable for evaluating the predictive performance of both the 1D-CNN model and the C2RCC processor. Table 5.3 lists the available points for September 3, 2021, showing sampling dates, locations, *in situ* Secchi disk depth, and corresponding predictions from both C2RCC's kd_{z90max} and the 1D-CNN model. The 1D-CNN model predicting Secchi depth outperformed the kd_{z90max} parameter of the C2RCC processor with a maximum error of 4 meters, compared to an error of over 10 meters for kd_{z90max} . Figure 5.13 illustrates the difference in prediction accuracy between the two models and plots the Secchi disk depth and kd_{z90max} predictions of both models alongside the *in situ* measurements. The curve of the 1D-CNN model closely follows the *in situ* measurements, indicating a higher accuracy, while the curve for the kd_{z90max} predictions of the C2RCC processor shows a significant deviation, indicating a lower accuracy. Results show the 1D-CNN model predicts Secchi disk depth more reliably than C2RCC's kd_{z90max} parameter.

Table 5.3. Comparison of *in situ* secchi depth measurements with kd_{z90max} predicted values from C2RCC processor and Secchi disk depth from 1D-CNN model (Adapted from [141])

#	Date	Lat	Lon	C2RCC [m]	1D-CNN [m]	In situ [m]
1	2021-09-01	45.173	14.657	11	18	22
2	2021-09-02	45.247	14.417	14	24	26
3	2021-09-03	44.787	14.157	15	21	23
4	2021-09-06	45.435	13.397	11	15	18

5.4. Discussion

In this section, the accuracy, limitations, and potential uses of the proposed model for predicting Secchi disk depth with Sentinel-3 OLCI data is discussed. It also examines possible future applications for monitoring long-term changes in Z_{SD} .

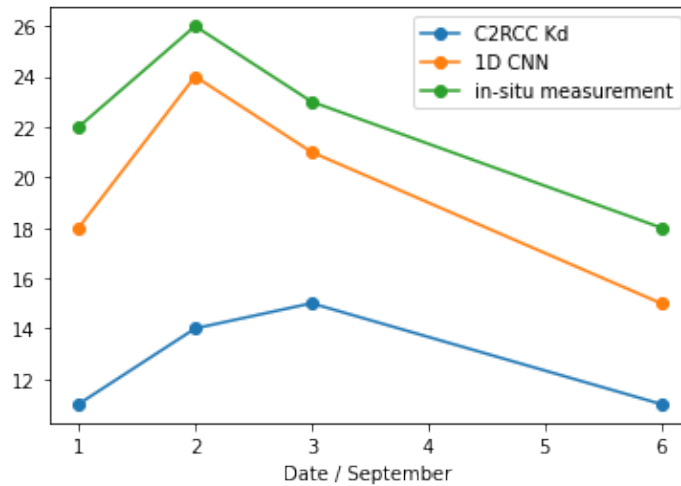


Figure 5.13. Comparison of Secchi disk depth predictions using in situ measurements, 1D-CNN model, and C2RCC processor based on data from Table 5.3 (Adapted from [141])

5.4.1. Accuracy of the 1D-CNN Model

The 1D-CNN model's performance was thoroughly assessed using metrics such as R^2 , RMSE, RRMSE, MAE, and RMAE. It shows good accuracy, especially with multi-spectral data. However, there are limitations when using specific datasets. Although R^2 values show a strong correlation between predicted and measured Z_{SD} values, metrics like RMSE, RRMSE, MAE, and RMAE reveal some variability and differences from actual values. This variability is due to complex interactions between water quality factors, environmental conditions and the limitations of the modeling approach [266]. The model was also compared with regression algorithms and the kd_{z90max} parameter of the C2RCC processor, both quantitatively and qualitatively, demonstrating its effectiveness in predicting Z_{SD} and highlighting its advantages over other methods.

5.4.2. Limitations

Secchi measurements depend on sufficient light to penetrate the water column and being reflected back to the measuring device. Since these measurements are made near the surface, knowing the reflectance at the bottom of the atmosphere (BOA) can help to determine the Secchi depth more accurately. Sunlight travels through the atmosphere to the earth's surface, after which is reflected and then passed through the atmosphere again before reaching the instrument. The first pass of sunlight can affect Secchi depth observations. Level-1 (L1) data includes reflectance values from both passes, and atmospheric conditions may alter the reflectance measured by the instrument.

Using 1D-CNNs for atmospheric correction might help adjust for atmospheric absorption and focus on features related to Secchi depth. However, further research is needed to confirm this. CNNs work by transforming reflectance values into weighted sums and differences

through multiple layers, but the complex nature of these transformations can make it difficult to understand their effects on atmospheric absorption. Therefore, further studies are needed to explore the potential of CNNs for atmospheric correction in remote sensing, although this is not the focus of this research.

Despite its promising performance, the proposed 1D-CNN model has some limitations. The accuracy of predictions can be affected by the quality and availability of ground truth data used for training and validation. Even though statistical analysis shows no significant deviation between measurements taken by citizens and official measurements, Secchi depth remains a subjective measure that cannot be validated completely independently. Additionally, the model's performance might differ in various aquatic environments due to the diversity in water types and optical properties. The model's sensitivity to outliers and noise, especially during atmospheric correction, could impact its predictive performance. Like any machine learning model, there is a risk of overfitting to the training data [267]. Therefore, managing hyperparameters carefully and using effective strategies to prevent overfitting is crucial. Addressing those limitations will improve the model's robustness and accuracy in predicting Secchi depth across different scenarios and locations.

5.4.3. Applicability to Other Regions

The model can be applied to other data sets that have a similar structure to the data set used in this study. Thus, the 1D-CNN model is also suitable for other regions if they have similar water quality characteristics and spectral properties as the Croatian and Slovenian parts of the Adriatic Sea. Furthermore, the Sentinel-3 OLCI imagery are freely accessible and can be downloaded for different geographical areas, as the instrument has global coverage and frequent repetitions. However, applying the model to very different water bodies might require re-training and adjustments to account for regional differences [268]. Nevertheless, the methodology described here for the northern Adriatic Sea can be adapted and used in other study areas and potentially provide valuable results.

5.4.4. Future Implications for Long-term Monitoring

Predicting Z_{SD} using remote sensing data provides an opportunity for ongoing monitoring of water quality and environmental change. Regular tracking of Z_{SD} provides valuable information about the state of the aquatic ecosystem, reveals trends in water clarity, and helps identify factors or conditions that may be causing disturbance. Combining Z_{SD} data with other environmental indicators, such as chlorophyll-a concentration and water temperature, allows for a thorough assessment of water quality. Long-term monitoring of Z_{SD} can also reveal changes in water transparency, which is crucial for understanding the impact of climate change, biogeochemical processes and human activities on the aquatic environment [269].

5.5. Conclusion

This chapter presents a novel approach to address the main challenges in satellite remote sensing for water quality monitoring. The focus is on the development and evaluation of a 1D Convolutional Neural Network (1D-CNN) for estimating Secchi disk depth (Z_{SD}), a key indicator of water quality parameters derived from multispectral Sentinel-3 OLCI imagery.

The 1D-CNN model was selected due to its effectiveness in processing high-dimensional spectral data, capturing non-linear relationships and performing precise pointwise estimation. This approach significantly reduces the need for manual feature processing by automatically extracting complex patterns from the spectral responses of pixels, making it well suited to the task of estimating Z_{SD} . To further refine the model, a comprehensive dataset was created by merging data from two official sources (one from Croatia and one from Slovenia) with contributions from a citizen science project. This dataset spans six years and covers a large area along the northern Adriatic coast, providing a rich basis for training the deep learning algorithm.

The evaluation of the 1D-CNN shows that it is able to achieve high accuracy in Z_{SD} estimation, outperform existing methods (C2RCC) and avoid the problems of overfitting that are common in other machine learning models. In addition, the ability of 1D-CNN to generalize well to unseen data increases its potential for wider application in different geographic regions with similar water quality characteristics. The results show that this approach not only improves the accuracy but also the efficiency of water quality monitoring by remote sensing, thus making an important contribution to the field.

6. EVALUATION OF 2D-CNN MODEL

This chapter describes the application of a 2D convolutional neural network (2D-CNN) for road detection using volunteer OpenStreetMap (OSM) data and multispectral satellite imagery (Sentinel-2). By using remote sensing and a well-trained model, it is possible to detect roads in a more time and cost-effective way compared to traditional field-based monitoring. The problem of road detection represents a *target recognition* task and can be considered as a binary classification. Road detection is crucial for urban planning and development, as road data can sometimes be outdated or unavailable. Moreover, road construction often requires the removal of environmentally important green spaces, so this approach can help monitor such changes.

The chapter is organized as follows. The introduction provides an overview of the literature on deep learning models applied to remote sensing data for road detection. The following section describes the study area, data collection and processing, the architecture of the proposed 2D-CNN model that performed best during hyperparameter tuning, and a summary of the model evaluation metrics. In the results section, metrics and graphs of 2D-CNN performance in detecting roads are provided. Moreover, the model's accuracy and its limitations are discussed. Finally, conclusions are drawn.

6.1. Introduction

Roads are man-made objects that presents important classes which can be useful for urban area monitoring, planning and detection, change detection of environment (e.g. deforestation), human population estimation, and infrastructure analysis for transportation and telecommunication networks [270]. Traditional methods for updating spatial data, such as road networks, are inefficient due to their high cost and time demands. In response to that, advanced remote sensing and deep learning, particularly convolutional neural networks (CNNs), offer a more practical and economical solution for extracting road types from high spatial resolution images.

According to [271] over the past decade deep convolutional neural networks in optical remote sensing imagery have proven to be highly effective for road extraction and automatic feature extraction. The focus has been on three main elements: road areas, centerlines, and boundaries. The most used approaches to accomplish it are based on CNNs, Fully

Convolutional Networks (FCNs), U-net architectures and Generative Adversarial Networks (GANs). These models offer a more practical and economical solution for extracting road types from high spatial resolution images compared to traditional mapping methods.

Patch-based CNNs using image patches centered around pixels are utilized for feature extraction and they are frequently used. However, road extraction remains challenging due to the complexity of images, the presence of various road types, and occlusions (e.g., shadows, cars, and buildings) that resemble road features in color and texture. While CNNs excel at processing high-resolution imagery, mixing pre- and post-processing techniques is recommended to further improve the accuracy of predictions. Additionally, the model's effectiveness depends on factors such as data quality, architecture, and hyperparameters. CNNs trained on a specific dataset may struggle with different scenes if the training data isn't comprehensive enough [272].

Classifying individual pixels is challenging due to insufficient features, making the incorporation of contextual information crucial. To address this, [273] introduced a GPU-based deep convolutional neural network (DCNN) that leverages a larger image context and predicts small patches of labels, improving accuracy and reducing computational costs. The model is using a sliding window approach, overlapping patches to generate a global map of the image. Its architecture features an encoder based on well-known backbones (e.g., VGG, ResNet) and includes post-processing to enhance result smoothness. Similarly, [274] employed a single patch-based deep CNN for extracting roads and buildings from high-resolution satellite imagery (HRSI), but with post-processing to improve performance.

A semantic segmentation approach using the Fully Convolutional Networks (FCNs) architecture derived from the VGG convolutional neural network model was used on the Massachusetts road and building datasets which is high spatial resolution aerial imagery [275]. It was used to predict classes "road" and "building" achieving precision rate around 71% for roads extraction and 78% for building extraction. Authors in [276] also used the Massachusetts road dataset to perform semantic segmentation. They proposed a Road Structure Refined Convolutional Neural Network (RSRCNN) model for road extraction. It is also based on the VGG CNN model, with additional convolutional, deconvolutional, and fusion layers to provide structured output for road regions. The RSRCNN approach outperformed several state-of-the-art methods, including other CNN-based approaches, in terms of precision, recall, F-score, and accuracy for road extraction. The experiments showed that the proposed road-structure-based loss function led to faster convergence and better performance compared to using standard cross-entropy loss. The authors demonstrated that the good performance was due to the road-structure-aware loss function, not just from dataset balance.

In [277], the authors utilize high-resolution satellite imagery from the Pleiades-1A and Geoeye satellites. They first train a convolutional neural network using thousands of 32×32 pixel patches. After training, they apply the CNN to the satellite images, examining each pixel to determine if it belongs to a road. To refine and connect the identified road network,

they employ Line Integral Convolution (LIC), enhancing the accuracy and connectivity of the extracted roads. The model achieved an accuracy of over 95% on the test data.

The studies mentioned above use data with a spatial resolution of 2 m or less to create detailed maps of buildings or roads. However, such data is often expensive and limits the ability to update the maps frequently. On the other hand, satellite data can be accessed for free and have continuous imaging, but have limited spatial resolution (often greater than 10 m) to recognise complex features such as buildings and roads [278, 279]. In [280], the authors propose models based on a U-Net architecture for detecting hardly visible roads in low-resolution Sentinel-2 satellite images. They used OpenStreetMap vector data as ground truth data. They applied the model to both single images and time series data. The results showed that the models can detect large and medium sized roads at a low spatial resolution (10×10 metres) of the Sentinel-2 data. Small roads are hardly visible in such data. On the other hand, the authors in [281] used SAR data obtained from the Sentinel-1 satellite. They implemented a semi-automatic approach for the extraction of trails and off-roads with a correctness of 68% and a completeness of 52%. The authors of [282] fused Sentinel-1 SAR and Sentinel-2 data (at 10 m) with OpenStreetMap to detect roads and buildings. They implemented a model based on Fully Convolutional Networks (FCNs) and U-Net architecture with ResNet-34 encoder. The fusion of Sentinel-1 and Sentinel-2 data improved the detection of roads and buildings compared to the use of individual data sources. The Intersection over Union (IoU) metric for roads improved from 0.44 when using Sentinel-2 data to 0.60 after the fusion.

Given the challenges associated with using low-resolution Sentinel-2 data, this chapter presents a 2D-CNN model specifically applied to patches from Sentinel-2 imagery and OpenStreetMap data. The method focuses on identifying the presence of roads within each patch, simplifying the problem while still leveraging deep learning's ability to manage spatial relationships in satellite data and successfully detecting objects. This approach is expected to enhance performance in detecting roads from Sentinel-2 data without requiring complex multi-source data fusion. By enabling faster road detection, this method could facilitate further analysis and improve the accuracy of generated maps.

6.2. Materials and Methods

6.2.1. Study Area

The study area is located in Split-Dalmatia County, Croatia, and is marked as a red square in Figure 6.1. A specific part of Split-Dalmatia County was selected due to its strategic importance. This area includes both urban regions characterized by high road density and rural areas where road density and material vary. It also includes the A1 highway, the largest and most modern highway in Croatia, which has four lanes. This approach ensures that a

wide range of road types are considered.

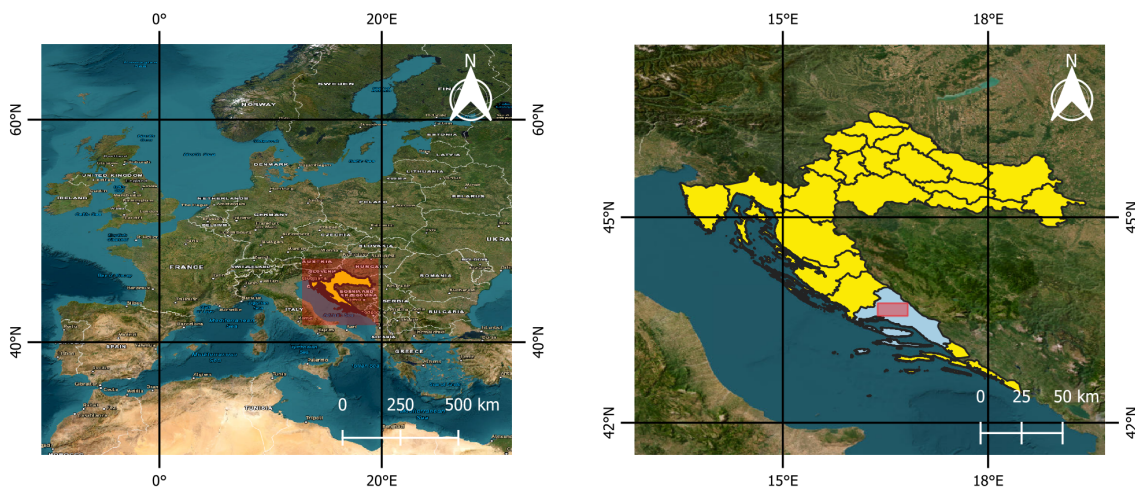


Figure 6.1. Study area in Split-Dalmatia County, Croatia (shown in Plate Carrée projection)

6.2.2. Dataset Construction and Preprocessing

The dataset was created by integrating road data from the OpenStreetMap (OSM) project and Sentinel-2 satellite imagery. The OpenStreetMap project was founded in 2004 and is a free, open geographic database of Volunteered Geographic Information (VGI) on the Internet [283]. The OSM data used in this study was retrieved on June 23, 2022, while the Sentinel-2 imagery was collected throughout 2023 from the Copernicus Data Space Ecosystem Repository (Copernicus Sentinel data [2023] [284]) to ensure that the imagery is up to date to capture all roads collected by the OSM project. Collecting imagery throughout 2023 also ensures a diversity of data that captures different weather conditions and natural variations around the roads.

The OSM data is represented as vectors where roads are presented as polylines. The dataset includes 14 consolidated transportation categories, organized into main road hierarchy (from motorways to tertiary roads), pedestrian infrastructure, special purpose ways (like bridleways and cycleways), and other road types. Since roads are much wider and these lines often do not align precisely with the center of the road, a shapefile was created with a 30-meter buffer around the lines to more reliably extract patches from Sentinel-2 images that include roads. Roads passing through tunnels were excluded from the dataset, because roads cannot be recognized on multispectral images due to the obstruction of the satellite's line of sight.

The Sentinel-2 images were downloaded free of charge from the Copernicus repository via the S3 API. All images were Level-2A, i.e. they were atmospherically corrected. The dataset included bands B01-B09, B8A, B11 and B12. Since roads are relatively narrow

objects for satellite detection, a spatial resolution of 10 meters was used for all Sentinel-2 bands, so the bands that were not originally at 10 m resolution were resampled.

In addition, the Scene Classification Layer (SCL) band was used for each observation to ensure that areas covered by clouds were excluded during patch creation based on the following values for pixel classifications [285]:

- 1 - SATURATED_DEFECTIVE,
- 3 - CLOUD_SHADOW,
- 7 - CLOUD_LOW_PROBA / UNCLASSIFIED,
- 8 - CLOUD_MEDIUM_PROBA,
- 9 - CLOUD_HIGH_PROBA,
- 10 - THIN_CIRRUS.

The patch size was set to 32x32 pixels, which corresponds to an area of 320x320 meters. To ensure a balanced dataset, an equal number of patches were extracted for the classes "road" and "no road" across all bands for the year 2023. As this is a binary classification task, the class "road" was labeled as "1" and the class "no road" was labeled as "0". A total of 850,606 patches were extracted. Because of very small patch size, initially selected study area exceeded the size of memory which could be processed. Therefore, the number of patches for each band was limited to 400 and the maximum number of patches for each date could be 9600, including both classes. The distribution of patches by month, based on Sentinel-2 imagery for the year 2023, is shown in Figure 6.2. For each date in the month for which Sentinel-2 imagery was available, the same number of patches was extracted for both classes. At the top of each bar in the figure is the total number of patches for each month. The pixel values ranges from 0 to 1 which were normalized using the MinMaxScaler technique.

6.2.3. Model Implementation

In order to find the most appropriate model, Keras Tuner with the Hyperband algorithm was used for model tuning, an effective approach to find optimal hyperparameters such as the number of filters, kernel size, dropout rate, and learning rates with minimal computational costs. Figure 6.3 depicts a sequential 2D convolutional neural network designed for road detection as a result of model tuning. The input dataset is composed of grayscale images that indicate the detection area and "road" or "non-road" objects. The dimension of the input data is 32x32x1, where 32x32 indicates the size of an individual patch, and 1 indicates one band.

The model starts with a 2D convolutional layer that has 32 filters, a 5×5 kernel, and uses the ReLU activation function. It is followed by batch normalization, which stabilizes

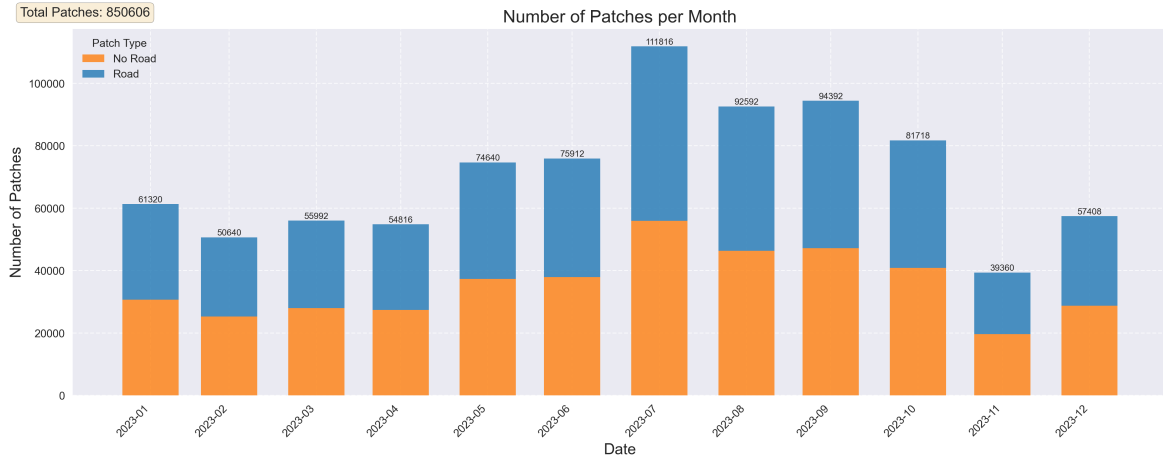


Figure 6.2. Monthly distribution of "road" and "no road" patches in Sentinel-2 dataset

and speeds up the training process. A 2×2 Max pooling layer is then used to halve the spatial dimensions. This is followed by two identical blocks, each containing a Conv2D layer, Batch Normalization, and Max pooling. The second Conv2D layer has 64 filters with a 3×3 kernel, and the third one has 256 filters with a 3×3 kernel. All Max Pooling layers have the same 2×2 kernel size. These convolutional layers extract higher-level features from the input and increase feature abstraction. A Global Average Pooling (GAP) layer is used instead of a flattening layer, which significantly reduces the number of parameters and captures more important features across the entire spatial dimension. The model then passes through a fully connected layer (Dense) with 128 units that uses the ReLU activation function. A Dropout layer with a 40% dropout rate is used to prevent overfitting. Finally, a Dense output layer with a Sigmoid activation function is used for binary object classification.

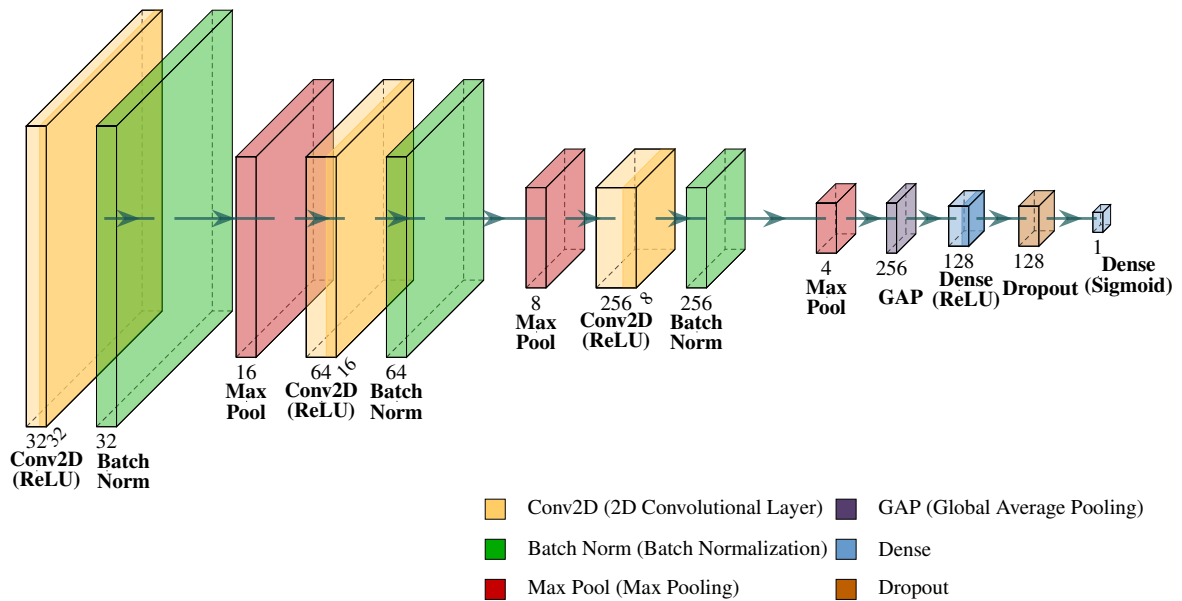


Figure 6.3. Proposed 2D-CNN architecture for road detection

The model was trained on a system with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz

and 16 GB of RAM, without GPU acceleration. Given the memory constraints and processing power limitations, training on a larger dataset was not possible during the early phases. Due to these computational constraints, only a subset of the training dataset, approximately 10,000 patches, was used in the initial training stage of the model. The initial training phase, configured for 100 trials, took 2 hours and 16 minutes to complete, stopping at trial 91 due to early stopping mechanism. Once the best-performing model was identified, it was trained on the entire training dataset and validated on 20% of this dataset, after which it was evaluated on the test dataset. The total number of model parameters is 201,473 (trainable: 200,769 and non-trainable: 704). It is compiled with a binary cross-entropy loss function and optimized using the Adam optimizer.

6.2.4. Evaluation Metrics

Evaluation was performed using classification evaluation measures. For trained CNN model, a confusion matrix was used to present the absolute and relative number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) samples. The test data was classified by the CNN model, and each sample was categorized into one of the following subsets [286]:

- True Positive (TP): the model correctly predicts the road as present (1), and the actual measurement confirms this.
- True Negative (TN): the model correctly predicts the road as absent (0), and the actual measurement confirms this.
- False Positive (FP): the model incorrectly predicts the road as absent (0), while the actual measurement is present (1).
- False Negative (FN): the model incorrectly predicts the road as present (1), while the actual measurement is absent (0).

This categorization (TP, TN, FP, FN) is useful for calculating the following metrics to evaluate tuned CNN model [287]:

- **Precision** - indicates the proportion of true positives relative to the total number of positive predictions made by the model (TP + FP). It measures how many of the predicted positives are actually positive.

$$Precision = \frac{TP}{TP + FP} \quad (6.1)$$

- **Recall** - reflects the proportion of actual positives that were correctly identified by the model. It measures how many of the true positives were detected.

$$Recall = \frac{TP}{TP + FN} \quad (6.2)$$

- **F₁-Score** - the harmonic mean of precision and recall, providing a balanced measure of both metrics.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.3)$$

- **Accuracy** - represents the proportion of correctly classified samples (both true positives and true negatives) relative to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.4)$$

- **ROC (Receiver Operating Characteristic) Curve** - a graphical representation that shows the trade-off between the true positive rate and the false positive rate across different threshold settings. It helps visualize the model's performance at various classification thresholds.
- **Area Under the Curve (AUC)** - measures the overall ability of the model to discriminate between positive and negative classes. It represents the area under the ROC curve, with a value between 0 and 1. A higher AUC indicates better model performance.

6.2.5. Overview of Common CNN Architectures for Comparison with 2D-CNN

To evaluate the performance of the proposed 2D CNN architecture, it will be compared with well-known CNNs, namely: DeepLab v3, ResNet-50 and U-Net.

DeepLabV3 architecture has a typical encoder-decoder structure. The encoder part of the network innovatively integrates the deep separable convolution Xception as the backbone network to extract the initial feature information of the image. It then employs the Atrous Spatial Pyramid Pooling (ASPP) module, replacing the general basic network, to expand the receptive field through continuous downsampling operations. The ASPP module achieves a wider field of view without compromising image resolution or performing additional calculations. This module uses global average pooling, 1 x 1 convolution, and the dilation rate (expressed as rate) of the combination of atrous convolutions of 6, 12, and 18 to encode image context information. The multi-scale feature maps are then concatenated and merged in the channel dimension, the number of output channels is adjusted to 256, and channel compression is achieved using 1 x 1 convolution. At this point, the resolution of

the feature map is 1/16 of the original image. The decoder uses bilinear interpolation to up-sample the feature tensor output by the encoding stage four times and then associates it with the feature map of the corresponding layer in Xception. It uses skip connections to capture detailed information from shallow features to enrich the semantics and details of the image. After performing two 3×3 convolutions to refine the features, the feature map size is gradually restored to the original image size using the four-fold upsampling method of bilinear interpolation. This approach helps to reduce the loss of feature information caused by excessive sampling steps [288]. DeepLabV3 was originally developed for semantic segmentation, but in this work it is adapted to handle binary classification tasks.

ResNet-50 or a deep residual network with 50 layers (48 convolutional layers, 1 max-pooling layer and 1 fully connected layer) was introduced by Microsoft researchers in 2015 [289]. Deep neural networks usually face the problem of vanishing gradient, which is responsible for poor network performance as the networks get deeper, especially architectures with more than 20 to 30 layers. To avoid this, the deep residual network uses Residual Blocks, which are the core units of ResNet, allowing the gradient to bypass some layers and provide smoother training even with increased depth. ResNet-50 has proven to be very effective in image classification, object detection and segmentation. Moreover, it achieved strong performance on the ImageNet dataset. Also, it uses a bottleneck architecture (1 x 1, 3 x 3, 1 x 1 convolutions) designed specifically for image recognition to reduce computational costs while maintaining performance. After each convolutional layer and batch normalization layer, the ReLU function is applied to ensure that only positive values pass through [290].

U-Net is a convolutional neural network architecture originally designed for segmentation of biomedical images, but for purpose of this work it is modified to do a binary classification. The network has 23 convolutional layers and is composed of two main parts: a contraction path (encoder) and an expansion path (decoder). The structure of architecture is U-shaped, hence the name. Encoder is built up of repeated convolution blocks, where each block contains two 3×3 convolution layers followed by a ReLU activation function. After each block, a 2×2 max pooling operation with stride of two is applied for downsampling, which also doubles the number of feature maps. The decoder is an inverted version of the encoder, built from repeated applications of upsampling the feature map followed by a 2×2 convolution that halves the number of feature maps after each block. Feature maps from the encoder are concatenated with corresponding cropped feature maps from the decoder. This is followed by two 3×3 convolution layers, each with a ReLU activation function. In the final layer, a 1×1 convolution is used to map the feature vector to the desired number of classes. At the end, the output segmentation map contains only the pixels for which the full context is available in the input image [291].

6.3. Results

In this section, the results of the proposed 2D-CNN model for road detection in Sentinel-2 image patches are presented. To evaluate the performance of the model, it is compared with well-known CNN architectures: DeepLabV3, ResNet-50 and U-Net. The performance of all models is evaluated using the following metrics: Accuracy, Precision, Recall and F_1 -Score. The comparison is performed on both the training and test datasets, providing insights into the strengths and limitations of each model for road detection in Sentinel-2 imagery.

6.3.1. 2D-CNN Model Performance

To evaluate the performance of the 2D CNN model for road detection in Sentinel-2 imagery patches, 50 training epochs were used. Figure 6.4 shows the accuracy (left plot) and loss (right plot) curves of the model for the training and validation sets. It can be seen from the accuracy curve that the training accuracy shows a stable increase, starting at around 75% and reaching over 93% at the 50th epoch. This indicates that the model is gradually increasing its accuracy and ability to classify road patches in the training set. On the other hand, validation accuracy is more volatile compared to training, but it has also upward trend. The validation accuracy was around 65% at the beginning and fluctuated between 80% and 90% during most of the training period.

The loss curves help to gain insight into the learning process of the model and to determine whether the model converges or overfits for a batch of examples. The graph shows that the loss curve of the training data has a rapidly decreasing loss and looks like a logarithmic curve. The loss decreased from about 0.55 to less than 0.30 in the first 20 epochs and slowly decreased until the 50th epoch, when the loss was about 0.25. The validation loss curve is more variable, but shows an overall decreasing trend, starting at around 0.70 and ending at around 0.33. The fluctuations in validation losses, especially the peaks observed around epochs 5, 40 and 45 indicate periods in which the model was unable to achieve good generalization to the validation set.

In order to gain insight into the performance of the 2D-CNN model, an accuracy assessment was made for the training and test datasets. Figure 6.5 presents the results of the metrics described in Section 6.2.4 for the training dataset. The model showed high performance across all metrics. The confusion matrix shows a similar ratio of correctly classified patches with no road (323,446) and patches representing a road (315,496). Also, the number of misclassified patches for both classes is less than 25,000, or 8% for each class. Due to the large number of correctly classified true positives (TP) and true negatives (TN), it is not surprising that the model showed a balanced performance in terms of the precision metric, namely 0.93 for "no-road" patches and 0.95 for road patches. The recall metric is similarly balanced and shows 0.95 for "no-road" patches and 0.93 for "road" patches. Therefore, the

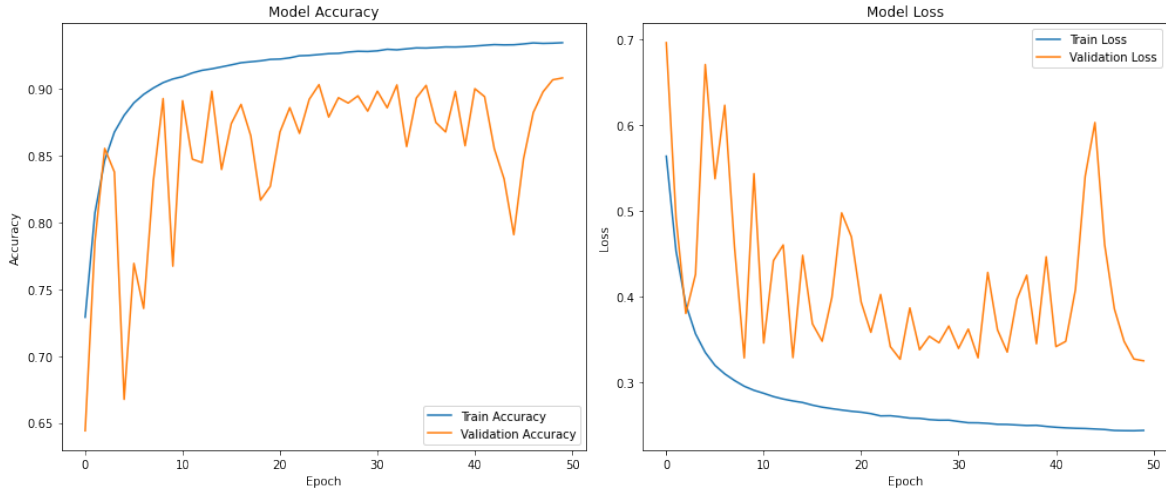


Figure 6.4. Loss and accuracy curves for the training and validation datasets over epochs for the 2D-CNN model used for road detection

model indicates good qualitative (high precision) and quantitative (high recall) performance on the training set. The overall accuracy of the model on the training set was 0.94, while the F₁-Score, which is primarily designed for binary classification, also achieved a high score of 0.94 for both classes. The ROC curve with an AUC of 0.99 shows that the model is capable of distinguishing the classes at different threshold settings. Similarly, the Precision-Recall curve shows a high AUC value of 0.99 suggesting that the model returns accurate results.

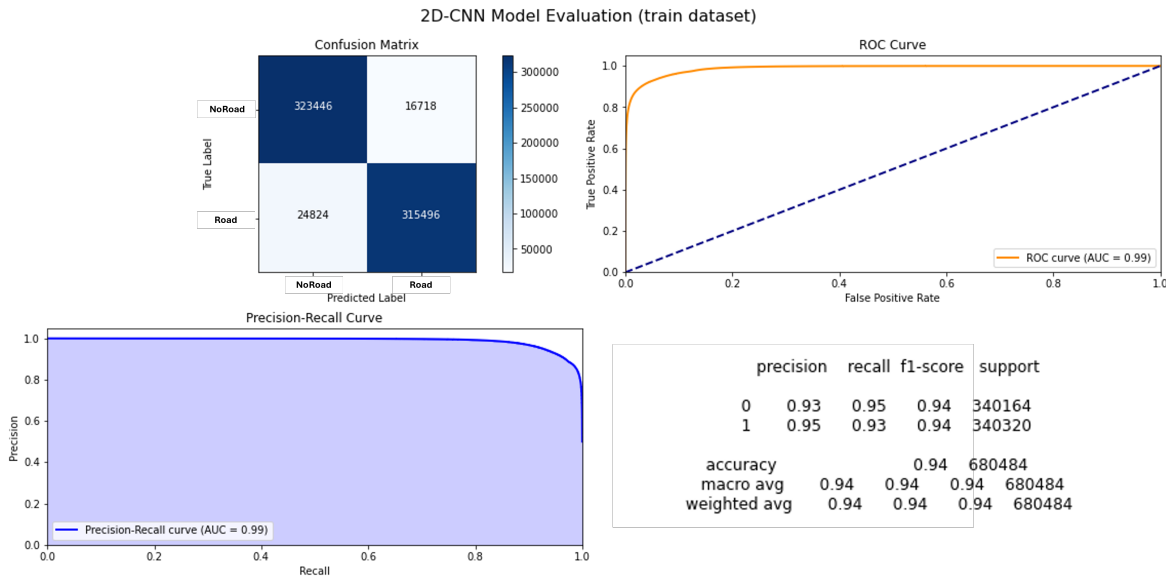


Figure 6.5. 2D-CNN model evaluation for road detection on train dataset

Figure 6.6 shows the performance of the 2D-CNN model on the test dataset. It can be observed that the model maintains a robust performance, albeit with a slight drop compared to the training set. The confusion matrix shows a similar number of correctly predicted patches, namely 78,760 for the "no-road" class and 75,891 for the "road" class. The number of incorrectly predicted patches is about 10% for the "road" class predicted as "no-road"

class and less than 10% for the "no-road" class predicted as "road" class. Therefore, the precision of the model was 0.92 for predicting the road class and 0.90 for the no-road class. For recall, the model scored 0.89 for the "road" class and 0.93 for the "no-road" class. The recall of less than 90% for the "road" class is to be expected given the increase in the number of incorrectly predicted patches in the "road" class as "no-road" class. The overall accuracy and F₁-Score for both classes were 0.91. The ROC curve with an AUC value of 0.97 shows the strong discriminative power of the proposed 2D-CNN model. The Precision-Recall curve also shows high performance with an AUC value of 0.97, indicating that the model maintains good precision across different recall levels.

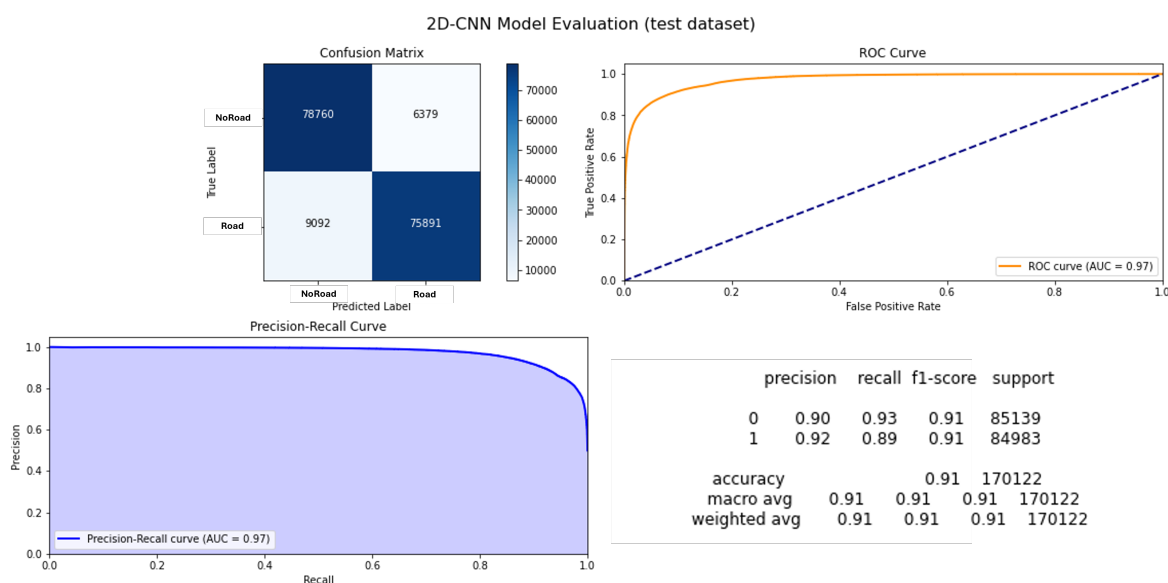


Figure 6.6. 2D-CNN model evaluation for road detection on test dataset

The slight difference in the performance metrics from the training to the test dataset is expected and is not significant given to the number of sample patches. The model generalizes well to unseen data and shows balanced performance for both classes, suggesting that the model is equally effective in identifying "road" and "no-road" patches in Sentinel-2 imagery. Overall, the model's robust performance on both the training and test datasets makes it a reliable tool for road detection in remote sensing applications.

6.3.2. Comparison with DeepLabv3, ResNet-50 and U-Net

To find out how well the 2D-CNN model performs, it was compared with the basic CNN architectures DeepLabV3, ResNet-50 and U-Net in both the training and test datasets. Since these architectures are primarily designed for image segmentation, they were adapted for binary classification by modifying their output layers to include a Sigmoid activation function. The models' performance is assessed using accuracy, F₁-Score, precision, and recall metrics. Table 6.1 represents the results over the training dataset whereas Table 6.2 reflects the results over the test dataset.

Based on the results of the tables, one can deduce that the proposed 2D-CNN model presents the best performance by showing the higher results in terms of accuracy, F₁-Score, precision, and recall of both the training and testing datasets. As discussed in the previous section, the model generalizes well and stays robust even on unseen data. Additionally, the ROC-AUC and precision-recall curves suggest good discriminative power between classes. However, the model has a small weakness; this was noticed by looking at the little difference between recall values in train and test datasets, which signifies that there are partial misclassifications of road patches in the test dataset.

Model	Accuracy	F ₁ -Score	Precision	Recall
2D-CNN	0.9389	0.9382	0.9497	0.9271
DeepLabV3	0.5861	0.3471	0.8223	0.2200
ResNet-50	0.6187	0.4654	0.7789	0.3319
U-Net	0.6841	0.6689	0.7030	0.6380

Table 6.1. Comparison of models on the train dataset

Model	Accuracy	F ₁ -Score	Precision	Recall
2D-CNN	0.9091	0.9075	0.9225	0.8930
DeepLabV3	0.5857	0.3450	0.8207	0.2184
ResNet-50	0.6167	0.4618	0.7732	0.3292
U-Net	0.6820	0.6663	0.7001	0.6356

Table 6.2. Comparison of models on the test dataset

DeepLabV3 model works effectively in terms of precision, especially on the training dataset where the precision is above 0.82. This means that the model correctly reproduces the road patches during classification. However, other metrics such as accuracy, F₁-Score and recall show low values, especially on the test dataset where the recall is 0.218 and the F₁-Score is 0.345. The low recall value means that DeepLabV3 does not detect a large proportion of true positives and therefore does not effectively detect a large number of patches of the "road" class.

The ResNet-50 architecture has a similar precision score on both train and test datasets, hence, it is consistent in detecting "road" patches across different datasets achieving around 0.77 to 0.78 score. However, despite outperforming DeepLabV3 in terms of F₁-Score and general performance, it still shows mediocre results overall. The relatively low recall score around 0.32 means that this model cannot recognize many patches related to the "road" class. The F₁-Score of ResNet-50 is higher compared to DeepLabV3, but it is still less than 0.5, which means generally poor performance.

The performance of the U-Net architecture is well-balanced, with an accuracy of 0.6841 on the training dataset, and 0.6820 on the test dataset. Precision, F₁-Score, and recall are relatively stable in all the datasets, with precision 0.7001 and recall 0.6356 in the test dataset.

These reflect that U-Net has a better generalization capability than DeepLabV3 and ResNet-50, but it does not compete with the 2D-CNN; even though it demonstrates a more stable performance across different metrics.

6.4. Discussion

This section describes the performance, limitations, and applicability of the proposed 2D-CNN model. It discussed the most important features related to accuracy, regional applicability, and long-term monitoring, providing insights for future research.

6.4.1. Accuracy of the 2D-CNN Model

It can be seen that the accuracy of the 2D-CNN model is very high for both the train and test datasets. The overall accuracy in the training dataset is 93.89%, in the test dataset 90.91%, which allows a very good distinction between "road" and "no-roads" patches in Sentinel-2 imagery. This slight decrease of about 3% between the train and test datasets indicates a good generalization to unseen data without significant overfitting. These results show that the 2D-CNN network has learned the detection of roads in Sentinel-2 images quite well. The final test accuracy of 91% supports its strong performance on unseen data. However, the persistent discrepancy between train and test metrics as well as the fluctuations in validation performance suggest that the model could still be optimized in terms of its generalization capabilities. This aspect is very important when it comes to real-world applications where the model is expected to reliably recognize road features on unseen satellite imagery. Compared to the baseline models, such as DeepLabV3, ResNet-50 and U-Net, the proposed 2D-CNN outperformed them in terms of accuracy for both datasets. In addition, the proposed 2D-CNN outperformed the basic models in other metrics as well, such as F_1 -Score, precision and recall.

6.4.2. Limitations

Despite these promising results, some limitations of the 2D-CNN model have to be considered. First, the Sentinel-2 imagery resolution (10 m) limits the capability of the model to recognize narrow roads or fine details. Thus, it could fail in identifying the smaller road structures. The present model considers only "road" and "no-road" classes. This simplification might turn out to be not enough in complex urban landscapes, since it may lead to some misclassifications of similar features, such as parking lots or large pedestrian areas. Furthermore, the types of roads, like highways, local roads, and even unpaved roads, are not segregated, which again restricts the effectiveness of carrying out a detailed analysis with this transport network. Degradation in model accuracy may be observed with poor image qual-

ity due to sensor problems or atmospheric interference like cloud cover or seasonal changes affecting the image quality. Despite all efforts to prevent this, there is always a certain risk of overfitting to the specific characteristics of the training dataset.

6.4.3. Applicability to Other Regions

The 2D-CNN model was trained on a specific geographical area which covers different types of roads, from motorways to the pedestrian pathways. While this variety offers robustness in detecting different road types, landscapes and road structures can differ significantly across regions and affect model's ability in making accurate predictions. To see how well the model performs in different regions and landscapes, it should be tested, which is possible due to the global availability of Sentinel-2 imagery. In addition, different countries have different standards for urban construction, which may also affect the model's performance in detecting roads. However, this should be considered and explored as part of future work.

6.4.4. Future Implications for Long-term Monitoring

The developed 2D-CNN model shows the potential for long-term monitoring of road networks by using Sentinel-2 imagery. Thus, in the future, it may be possible to use this model to track changes in the road network over time, which would be useful in urban development studies and infrastructure monitoring. If the results of the model are integrated with other geospatial data such as population density or economic indicators, or with data from other satellite sensors, it could also provide comprehensive insights into urbanization patterns and detect roads under different conditions. Moreover, the model could be used to regularly update road maps, which could lead to more accurate geographic information systems.

6.5. Conclusion

The 2D-CNN model performed very well in predicting the presence of roads within Sentinel-2 remote sensing imagery, achieving a classification accuracy of 91%, despite the low spatial resolution of the Sentinel-2 bands, which were resampled to 10 m for a uniform dataset. This success highlights the model's capability to effectively detect spatial relationships, accurately distinguishing between "road" and "no-road" patches.

Notably, this relatively simple 2D-CNN architecture outperformed more complex models such as DeepLabV3, ResNet-50, and U-Net. This finding emphasizes the importance of selecting an appropriate model architecture tailored to the specific task, often more critical than relying on advanced but potentially less suitable models.

Furthermore, while the 2D-CNN model is designed to classify image patches as either containing roads or not, its performance can serve as a valuable preliminary step for more

detailed road segmentation. By first filtering out patches without roads, the model can focus subsequent segmentation efforts on patches where roads are present, thus optimizing resource allocation and improving the efficiency of the segmentation process. This approach can reduce computational overhead and enhance the precision of detailed segmentation models applied to road patches.

The model's success aligns with the core objective of demonstrating the practical application of CNN architectures in remote sensing tasks. It underscores the potential of deep learning as a powerful tool for geographic information systems (GIS) and spatial analysis. The 2D-CNN's balance of metrics precision, recall and generalization makes it a valuable approach for addressing spatially related problems in remote sensing.

While these results are promising, they are exploratory in nature. Further research is likely to refine these findings and expand their applicability to diverse global landscapes and infrastructures.

7. EVALUATION OF 3D-CNN MODEL

This chapter describes an approach that combines advanced deep learning techniques and multispectral remote sensing data to predict fire propagation potential (FPP) [292]. This task represents *pixel classification* problem in remote sensing. A 3D convolutional neural network (3D-CNN) is proposed as a deep learning method that can successfully predict FPP based on Sentinel-2 satellite images. This chapter describes the implementation and characteristics of the 3D-CNN, as well as a comparison with other deep learning models. Successful prediction of FPP could aid in monitoring and prevention measures that are part of effective forest fire management.

The structure of the chapter is organized as follows. In the Introduction, an overview of the literature on the potential of fire propagation and existing tools that were developed for this purpose is given. Also, an overview of previous works that used machine learning methods, as well as works that used deep learning methods and applied them to remote sensing data was described. The Materials and Methods section describes the research area, the construction and preprocessing of the data set used, and the architecture of the 3D-CNN model, as well as the 1D-CNN and 2D-CNN models that were used for comparison. The Results section presents the performance metrics of the 3D-CNN model and compares it to the 1D-CNN and 2D-CNN models, including a qualitative evaluation of the FPP prediction maps. The Discussion discussed the accuracy of the proposed 3D-CNN model, its limitations, and directions for future research. The chapter ends with a conclusion on the implications of this work on forest fire management and risk reduction.

7.1. Introduction

Forest fires pose a major threat to many ecosystems. They can cause great damage to forests and soil, and pose a threat to human lives. During the last few decades, it has been shown that forest fire seasons have become longer and more intense, especially in the summer months. Some of the causes of this are global warming, rapid industrialization and human activities that can be accidental, negligent or intentional [293]. In addition to the fact that forest fires destroy vegetation, during their activity significant amounts of greenhouse gases are released, which contribute to climate change. Based on the above, there is a need for actions that could lead to the implementation of accurate and reliable models for the purpose of preventing or

predicting fires.

Fire propagation models that use fire spread simulators such as FlamMap, Farsite and Wildfire Analyst or machine learning techniques play a significant role in the prediction of forest fires behavior. The mentioned models simulate the spread of fire and estimate the burned area based on the observed environmental factors. In this way, important insights essential for designing effective fire prevention and management strategies are obtained [294].

Fire risk comprises fire danger, exposure and vulnerability [295]. Given the scale of the problem, countries worldwide have dedicated substantial resources to fire suppression, prevention, and post-fire recovery, which often has an economic impact in terms of cost [296]. Fire risk assessment includes estimating the probability of a dangerous fire event, but not all ignitions cause danger. If a fire does not spread, it is not considered dangerous. However, once a wildfire has spread, it is very difficult to control and extinguish [297]. Fire propagation potential is quantitative measure used for distinguishing those two situations.

Traditional techniques often rely on physical, qualitative and statistical analysis to understand and predict wildfires and often fail to accurately capture the intricate and nonlinear nature of fire dynamics [298]. The aforementioned traditional techniques attempt to simplify the complex interactions between different factors in the environment, leading to limited predictive power. The emergence of various machine learning (ML) algorithms has greatly changed the way of spatial prediction of forest fire susceptibility compared to traditional methods. Machine learning has enabled researchers to analyze large data sets and discover hidden patterns within them that may indicate fire danger [299]. Examples of ML algorithms that have been successfully used to generate wildfire susceptibility maps are artificial neural networks (ANNs), random forests (RF) and support vector machines (SVMs) [300–302]. These algorithms capture complex relationships and interactions between input variables, such as topography, vegetation types, and weather conditions, which provides more reliable results in fire-prone areas. Also, an increasing number of studies provide a comparison of the performance of different ML algorithms in predicting forest fires susceptibility [292, 303, 304]. In this way, they provide an insight into the strengths and weaknesses of ML algorithms in different contexts, and emphasize the importance of choosing the most appropriate algorithm based on the characteristics of the studied research area and the available data.

However, despite their advantages, many of these ML models are limited by their shallow pixel-based architectures, which cannot fully exploit the spatial patterns present in the data. This limitation emphasizes the need for more advanced approaches such as convolutional neural networks that can extract and exploit representative features from the input data. Recent studies have shown that deep learning is making significant progress compared to traditional ML methods. In [305], the authors introduced a CNN to predict fire susceptibility in the Chinese province of Yunnan and achieved an accuracy of 82% in their validation set. In continuation of this work, the same authors compared different CNN and

multilayer perceptron (MLP) architectures globally [306], analyzing fire susceptibility across seasons, and found that a CNN with two convolutional layers and three fully connected layers achieved extremely high AUC score (0.956 - 0.982). Complementing these studies, the authors in [307] implemented a deep neural network with three hidden layers and achieved an AUC of 0.894. In [308], an ensemble model combining two deep learning networks for forest fire susceptibility in two regions in Chile is proposed and achieves an AUC of 0.953.

This chapter presents a 3D-CNN model implemented for the purpose of predicting fire propagation potential using remote sensing data. In addition to remote sensing data, data on past fires and short interventions by firefighters are used. Data on past fires represent a high fire risk labeled as FPP=1, while data on short interventions by firefighters serve as a reference for non-spreading fires labeled as FPP=0. In order to evaluate the proposed model, 1D and 2D convolutional neural networks were implemented to compare with the proposed 3D-CNN model.

7.2. Materials and Methods

7.2.1. Study Area

Figure 7.1 shows study area located in the Republic of Croatia, specifically within the Split-Dalmatia County. Some data points geographically belong to the Šibenik-Knin County due to the proximity to the county border between these two counties. All data were collected for the period from 2017 to 2021.

Firefighter Interventions Database (in Croatian "Upravljanje Vatrogasnim Intervencijama - UVI") is official database of records of every intervention in Croatia and includes information about location and resources used in intervention. In the scope of this research, interventions that lasted less than 2 hours and less than 3 firefighters were engaged were considered situations with low propagation potential. Those records in the Split-Dalmatia County are represented as point data, marked in yellow on the map depicting the study area. These points indicate the location of fires but not their size, since they did not spread significantly. In contrast, data extracted from the European Forest Fire Information System (EFFIS) database include both fire occurrences and their sizes. These are displayed on the map as polygons, with the size of each polygon corresponding to the scale of the fire event. EFFIS provides support to services responsible for forest fire protection in the European Union (EU) and neighboring countries, while also offering updated and reliable information on wildfires in Europe to the European Commission services and the European Parliament [309].

This research area was selected due to its sensitivity to forest fires and based on long-term monitoring of the frequency of forest fires in the area, which are triggered by a combination of climatic conditions, vegetation type and human activities.

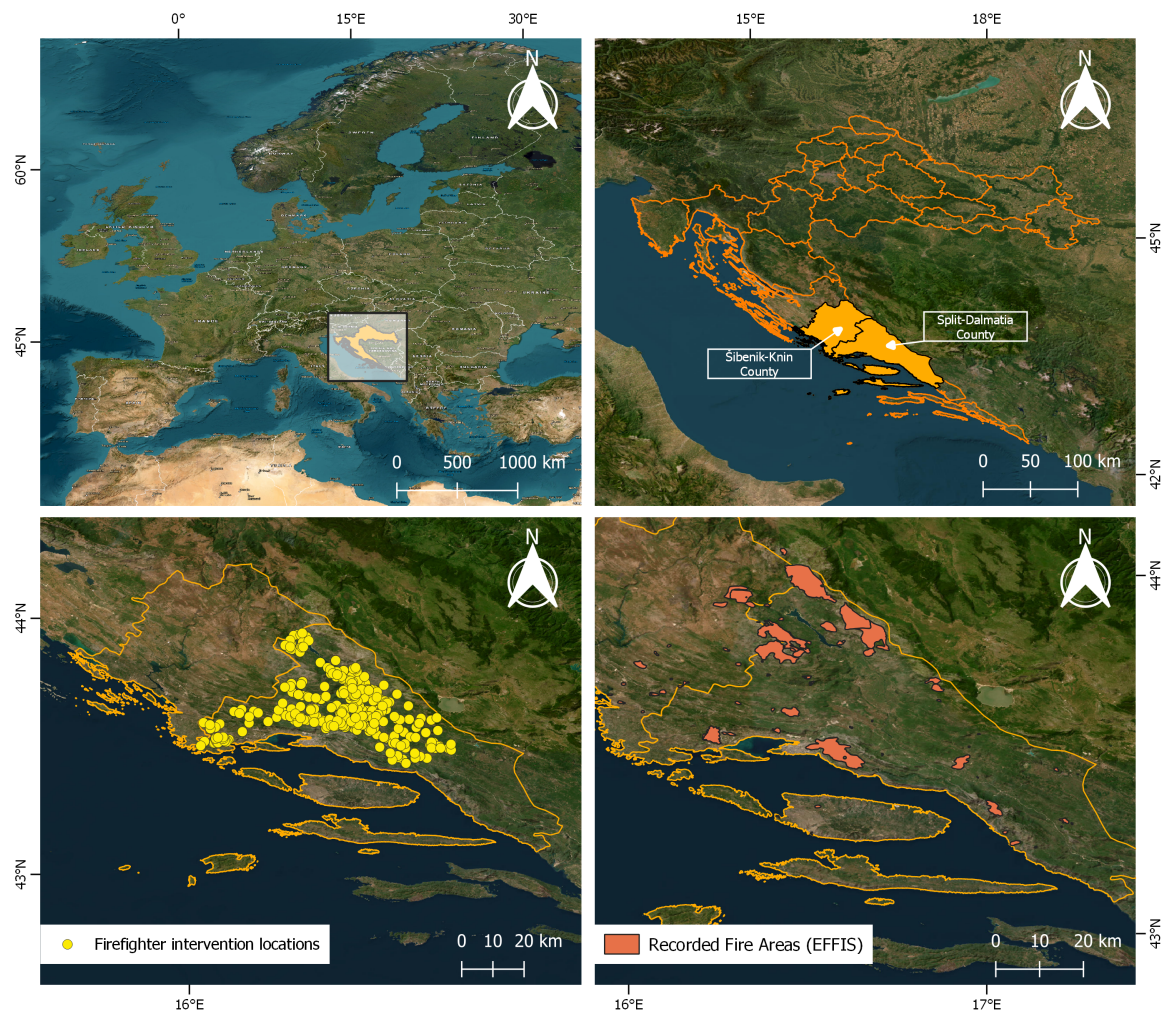


Figure 7.1. Study area in Split-Dalmatia County, Croatia. Points represent small fire locations, while polygons represent larger fires from EFFIS database (shown in Plate Carrée projection)

7.2.2. Dataset Construction and Preprocessing

In order to train and implement a reliable model for detecting fire propagation potential, a spatial dataset consisting of Sentinel-2 satellite imagery, fire data from the European Forest Fire Information System (EFFIS), and incident points from smaller fires handled by firefighters was used.

The Sentinel-2A/B images were retrieved from the Copernicus Data Space Ecosystem via the S3 API for the period from 2017 to 2021. This data is open access (Copernicus Sentinel data [2017-2021] [284]) and was provided in .jp2 format at Level-2, meaning the images were atmospherically corrected and resampled to a uniform spatial resolution of 60 meters for all bands (B01-B07, B8A, B09, B11, and B12).

Fire intervention points represent incidents where fires lasted less than two hours and required no more than two firefighters. These points are considered to have a low potential of fire propagation and are labeled as $FPP = 0$. In contrast, data from EFFIS, which covers

larger-scale fire events, is considered to represent propagated fires and is labeled as $FPP = 1$.

For each EFFIS fire polygon, data cubes were constructed by dividing the polygons into 1km x 1km patches and associating them with corresponding Sentinel-2 band data. Similarly, for each fire intervention point, a 1km x 1km buffer zone was created around the location, and all relevant Sentinel-2 bands patches were extracted for analysis. On the final dataset downsampling was performed to achieve balanced dataset with similar number of representatives of each class.

A total of 880 data cubes were generated for the study, with 453 cubes corresponding to fire intervention points and 427 cubes representing EFFIS fire polygons. Each data cube has a shape of (17, 17, 11), where the first two dimensions (17x17) represent a 1km x 1km patch (17 pixels, each 60 meters across), and the last dimension (11) stands for the number of spectral bands. To ensure consistent feature scaling and improve the stability of the model during training, the dataset was normalized using `MinMaxScaler`, which scales the data to a range of 0 to 1. This process of generating data cubes and downloading the corresponding Sentinel-2 imagery was fully automated using Python scripts, which ensured consistent and efficient data handling. Sentinel-2 images were carefully selected to capture the environmental conditions as close as possible to the time of each fire incident, while ensuring they preceded the fire event. For both EFFIS data and intervention reports, images were consistently chosen from within a 4-day window before each fire occurrence. Figure 7.2 depicts a box plot for the EFFIS and interventions datasets, showing that the majority of selected images fell within 1 to 2 days before the fire events, with median time differences ranging from 1 to 2 days across all years for both datasets. This narrow temporal window ensures that the satellite data closely represents the pre-fire conditions, providing a robust foundation for modeling fire propagation potential. The consistent approach across years (2017-2021) and between the two fire datasets (EFFIS and interventions) allows for a standardized and comparable analysis of environmental factors contributing to fire risk.

7.2.3. Model Implementations

In this section, the architectures of 1D-CNN, 2D-CNN and 3D-CNN models are described. The same architecture was used for each model, which is adjusted during hyperparameter tuning. For this purpose, Keras Tuner with the `RandomSearch` function was used, which in 100 attempts for each individual model managed to adjust the architecture with the number of layers, filters and other parameters in order to obtain the optimal performance for each CNN. In this way, a fair comparison of the models was ensured, as their initial structures were identical.

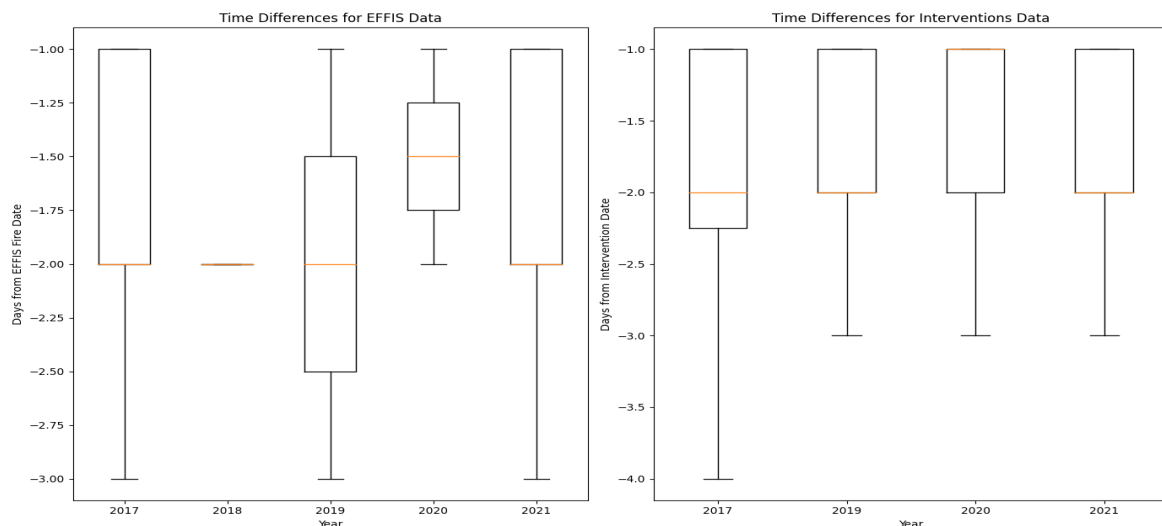


Figure 7.2. Time differences between fire incidents and corresponding Sentinel-2 images (2017-2021). Left: EFFIS data; Right: Intervention data. Negative values indicate that images were taken before fire events.

1D-CNN

Figure 7.3 shows a sequential 1D-CNN model representing the best architecture found by hyperparameter tuning. The model is fitted to Sentinel-2 satellite image data and takes as input a one-dimensional array of 11 spectral values for one pixel related to the bands of the Sentinel-2 satellite. Thus, the shape of the train data in an input layer is (11,1).

The proposed architecture includes three 1D convolutional layers (Conv1D), each followed by a Max Pooling layer with pool size 2 and stride 2 for spatial dimensionality reduction to extract important features. In addition, all Conv1D layers have a kernel size of 3 and ReLU activation functions, but differ in the number of filters: the first and last layers have 128 filters, while the second layer has 64 filters. The output of the last Max Pooling layer is flattened to produce a 1D feature vector, followed by a dense layer with 256 units and the ReLU activation function. To prevent overfitting, a dropout with a rate of 0.3 is applied. The architecture ends with the dense output layer with one unit (binary classification) with the Sigmoid activation function.

The model was compiled using binary cross-entropy as the loss function and optimized using the Adam algorithm. The architecture comprises a total of 83,137 trainable parameters, with no non-trainable parameters. This proposed 1D-CNN model is designed to capture the spectral relationships in the Sentinel-2 data at the pixel level, enabling effective prediction of fire propagation potential.

2D-CNN

Figure 7.4 depicts a sequential 2D-CNN model, representing the optimal architecture identified through hyperparameter tuning for fire propagation potential prediction. The model

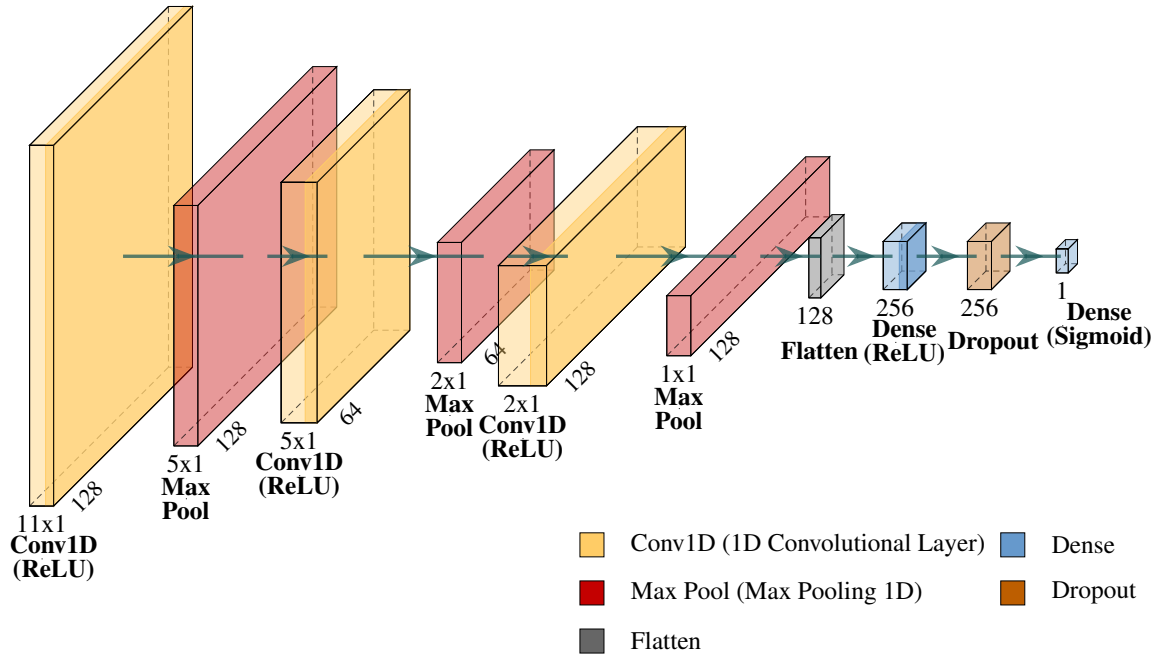


Figure 7.3. Proposed 1D-CNN architecture for Fire Propagation Potential

was developed to process data from Sentinel-2 satellite imagery and analyze spatial features across the two-dimensional patches of spectral data. Each sample fed into the network is a two-dimensional array (17x17 pixels) extracted from a single spectral band, representing the spatial distribution of that band's spectral values. This results in an input shape of (17, 17, 1), where the last dimension indicates that each patch contains data from one spectral band, preserving both the spatial characteristics and spectral information of that specific band.

The architecture includes two 2D convolutional layers (Conv2D). The first layer uses 32 filters, kernel size 3×3 , and maintains spatial dimensions of $17 \times 17 \times 32$, while the second layer also uses 32 filters but with kernel size 5×5 and maintains spatial dimensions of $8 \times 8 \times 32$. Both of these layers use ReLU activation functions. Each convolutional layer is followed by a Max Pooling layer using pool size 2 and stride 2 to reduce the spatial dimensions and extract key features. This first Max Pooling layer reduces dimensions from $17 \times 17 \times 32$ to $8 \times 8 \times 32$, while the second one is reducing the dimensions from $8 \times 8 \times 32$ to $4 \times 4 \times 32$. The output from the final Max Pooling layer is flattened to generate a 1D feature vector of 512 units. Following the flattening operation, the first dense layer comprises 256 units with a ReLU activation function. This is succeeded by a dropout layer with a rate of 0.2 to mitigate overfitting. The output dense layer consists of a single unit with a Sigmoid activation function, facilitating binary classification of fire propagation potential.

For this architecture the binary cross-entropy serves as the loss function, while optimization is performed by using the Adam algorithm. The architecture summarizes 157,537 trainable parameters, hence, all of them are active during learning.

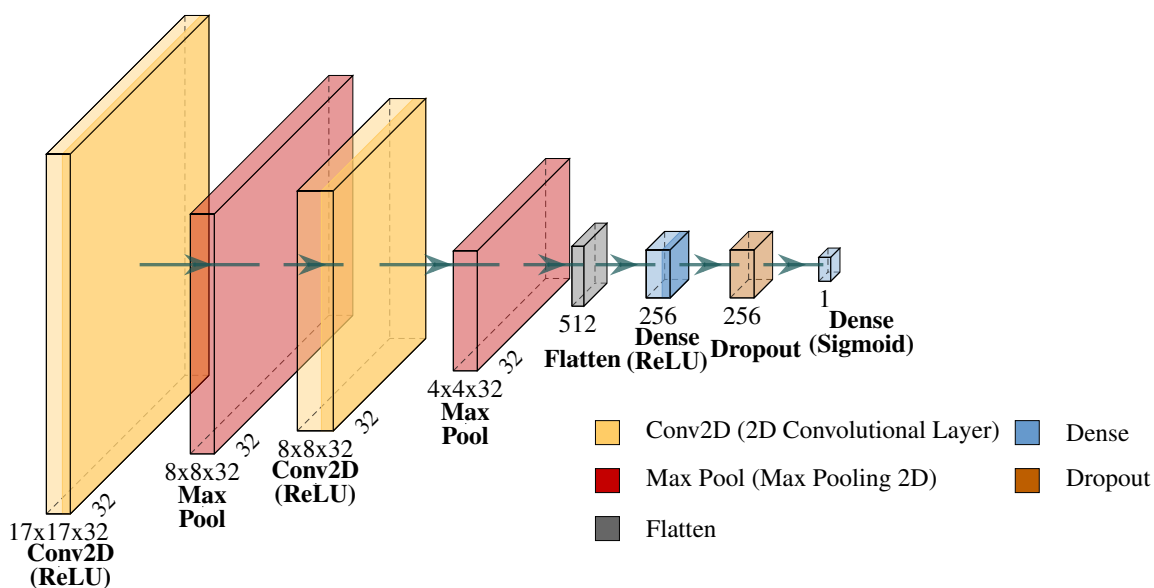


Figure 7.4. Proposed 2D-CNN architecture for Fire Propagation Potential

3D-CNN

Figure 7.5 illustrates the optimized 3D-CNN architecture developed through hyperparameter tuning for fire propagation potential prediction. This model is designed to learn and process spatial-spectral relationships in three-dimensional data cubes derived from Sentinel-2 satellite imagery. Each input sample having dimensions of $17 \times 17 \times 11 \times 1$, representing spatial and spectral information.

The architecture consists of three 3D convolutional layers (Conv3D). The first layer consists of 64 filters with a kernel size of $5 \times 5 \times 5$; thus, the size of the input remains spatial and spectral dimensions of $17 \times 17 \times 11 \times 64$. The second layer contains 32 filters with a kernel size of $5 \times 5 \times 5$, resulting in a dimension of $8 \times 8 \times 5 \times 32$. The last convolution layer applies a kernel size of $3 \times 3 \times 3$ by using 96 filters and generates $4 \times 4 \times 2 \times 96$ data cubes. All Conv3D are followed by ReLU for activation. After every convolutional layer, there is a 3D Max Pooling layer to reduce the spatial dimensions with a pool size of $2 \times 2 \times 2$ and stride $2 \times 2 \times 2$. The first Max Pooling layer reduces dimensions to $8 \times 8 \times 5 \times 64$, the second further reduces to $4 \times 4 \times 2 \times 32$, and the last performs a final reduction to $2 \times 2 \times 1 \times 96$. The output from the final Max Pooling layer is flattened to create a one-dimensional feature vector of 384 units. This is followed by a Dense layer with 192 units and ReLU activation function. To mitigate overfitting, a dropout layer with a rate of 0.2 is implemented. The output layer is a Dense layer with one unit and Sigmoid activation function, facilitating binary classification of fire propagation potential.

In the proposed 3D-CNN architecture, as in the two previously described architectures, the binary cross-entropy was used as the loss function and the Adam algorithm was used for its optimization. The total number of trainable parameters is 421,249, with no non-trainable parameters.

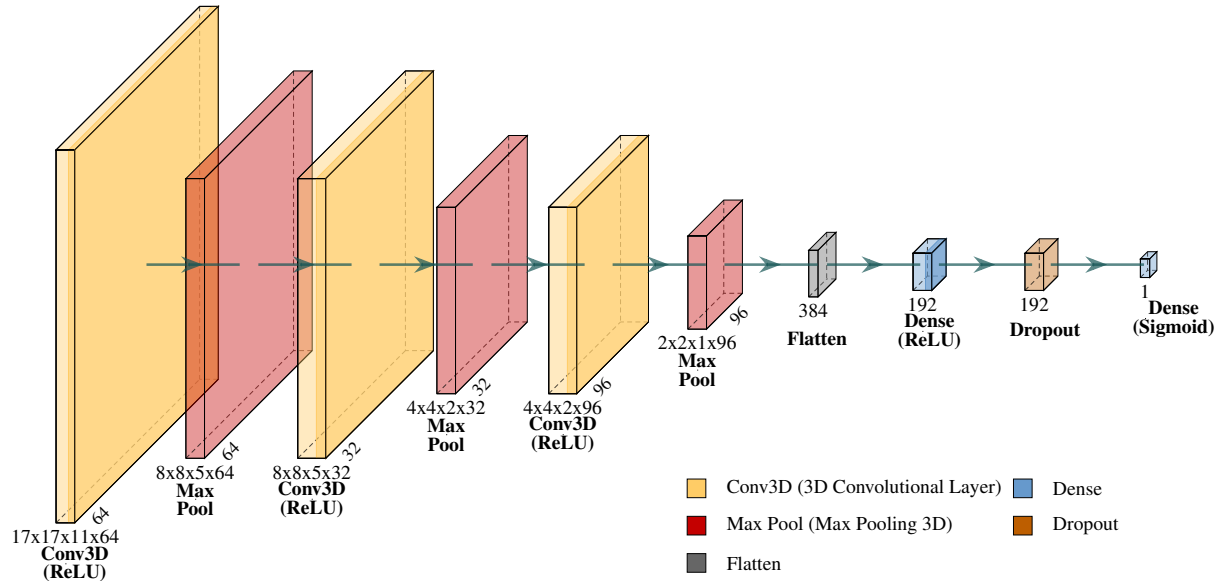


Figure 7.5. Proposed 3D-CNN architecture for Fire Propagation Potential

7.3. Results

This section describes the results of the proposed 3D-CNN model for predicting fire propagation potential based on Sentinel-2 imagery. The proposed model is compared with 1D-CNN and 2D-CNN models whose principle is described in Section 3.5. Accuracy, Precision, Recall and F₁-score metrics, confusion matrix, ROC curve and Precision-Recall curve were used to evaluate the mentioned models. The definitions of used metrics can be found in Section 6.2.4. The comparison was made on the train and test datasets, and provides an insight into the advantages and limitations of each model. All models are trained and tested on a system with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and 16 GB of RAM, without GPU acceleration. Each model completed all 100 trials despite implemented early stopping with patience of 5. The training times varied significantly: 1D-CNN required 51.5 hours, 2D-CNN took 2.3 hours, and 3D-CNN needed 16.4 hours to complete the training process.

7.3.1. 3D-CNN Model Performance

The 3D-CNN model was trained over 50 epochs to predict fire propagation potential using Sentinel-2 satellite imagery. As illustrated in Figure 7.6, the model's performance was evaluated by examining accuracy (left plot) and loss (right plot) curves for both the training and validation datasets.

The accuracy curve on the training set (blue line) shows an increase in accuracy through the epochs where it reaches high values of 97-98% by the end of the training period. Based on the curve increasing trend, it can be noticed that the model effectively recognizes and learns patterns from the training data and thus minimizes errors during the epochs. On the

other hand, the accuracy curve on the validation set (orange line) shows some degree of accuracy fluctuation across epochs, especially in the early epochs, with values oscillating between 85% and 96%. This suggests that the model exhibits some instability in terms of performance and generalization. Despite the above, the validation accuracy, although variable, achieves values close to the training accuracy, which suggests that the model can accurately classify FPP on the validation set under certain conditions.

The loss curve on the training data set (blue line) decreases consistently across all epochs. The model reaches low loss values close to 0.05 during training, which indicates effective error minimization during training. However, the loss curve for the validation dataset (orange line) is irregular and fluctuates frequently, similar to the accuracy curve. This is particularly noticeable in the first epochs. Although, over time the validation loss decreases approaching the training loss values.

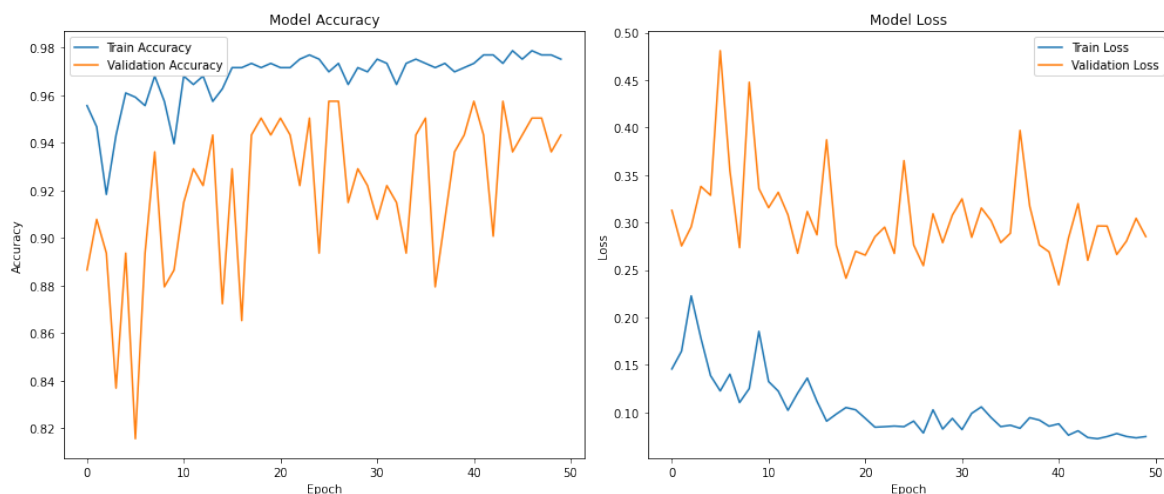


Figure 7.6. Loss and accuracy curves for the training and validation datasets over epochs for the 3D-CNN model used for fire propagation potential

Figure 7.7 shows that the 3D CNN model performs strongly on the training dataset. The confusion matrix shows the high accuracy of the model. The model predicted 348 true negative and 335 true positive satellite image patches, with only 18 false positive and 3 false negative patches. This indicates a strong classification performance during training. The ROC curve shows excellent discrimination ability with an AUC of 0.99. The curve rises steeply and quickly approaches the upper left corner, indicating a large number of true positive rates and a small number of false positive rates at different thresholds. The precision-recall curve also shows strong performance with an AUC of 0.98. The curve maintains high precision values over a wide range of recall values, indicating that the model successfully balances precision and recall. The effectiveness of the model is supported by the high values of the precision, recall and F_1 -Score metrics, which achieve results of 0.95 or more in both classes. The overall accuracy of model is 0.97. The described metric results show that the 3D-CNN model has a successful and balanced performance on the training dataset, thereby

learning to effectively distinguish between areas with and without fire propagation potential.

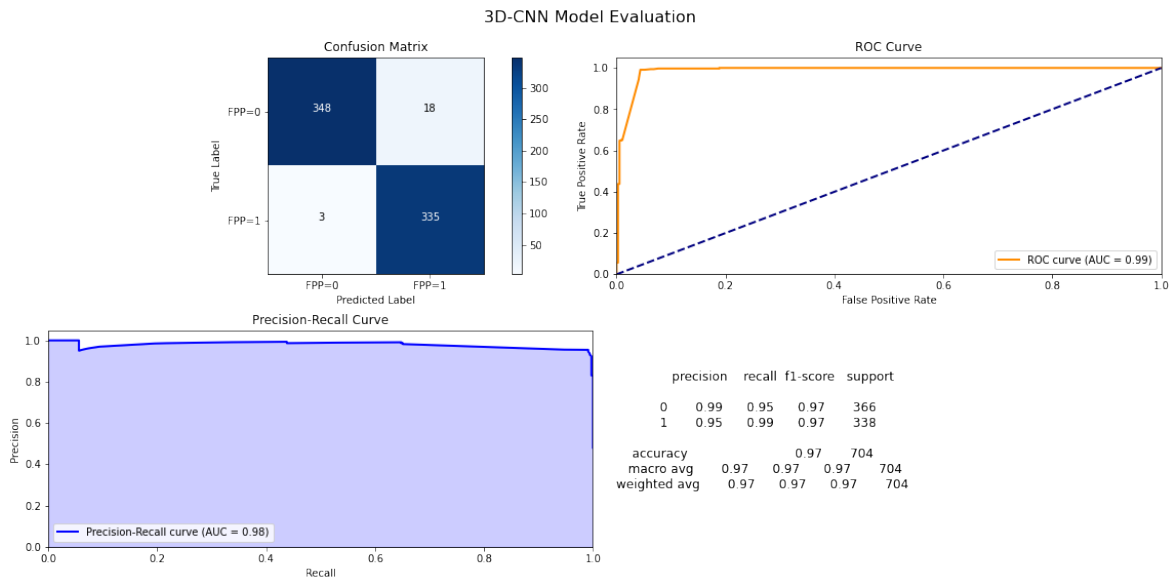


Figure 7.7. 3D-CNN model evaluation for fire propagation potential prediction on train dataset

The 3D-CNN model also shows strong performance on the test dataset as can be seen in Figure 7.8. The confusion matrix shows high accuracy, where the model predicted 84 true negative, 85 true positive, 3 false positive and 4 false negative satellite image patches. The low number of false positive and negative patches indicates that the model generalizes well to unseen data. The ROC curve shows an AUC of 0.97 indicating the excellent ability of the model to distinguish between classes, and the ROC curve itself rises sharply and maintains a high positive rate over various false positive rate thresholds. The ROC curve is located near the upper left corner, which indicates excellent model performance. Also, the Precision-Recall curve shows robust performance with an AUC of 0.94, and high precision over a wide range of recall values.

7.3.2. Comparison with 1D-CNN and 2D-CNN Models

In order to evaluate the effectiveness of the proposed 3D-CNN model architecture in predicting fire propagation potential from Sentinel-2 images, it was compared with the 1D-CNN and 2D-CNN models. Each of the mentioned models processes the input data in a different way, capturing the spatial and spectral information present in the satellite images. The models were trained and tested on the same datasets, and their performance was compared using the following metrics: accuracy, F₁-Score, precision and recall. Tables 7.1 and 7.2 present the results for the specified models for the train and test datasets.

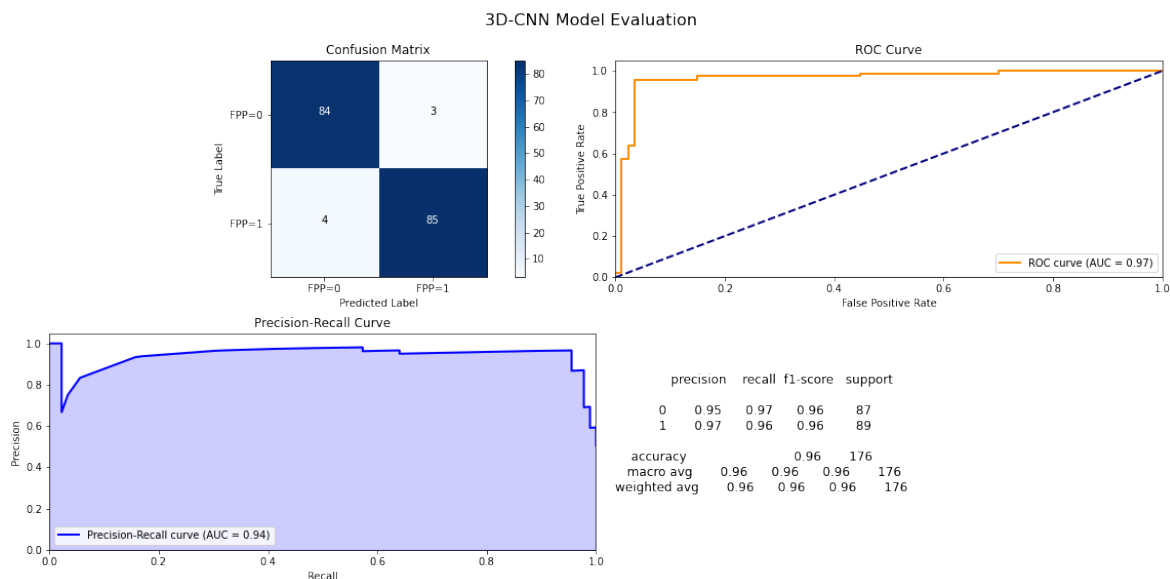


Figure 7.8. 3D-CNN model evaluation for fire propagation potential prediction on test dataset

Model	Accuracy	F ₁ -Score	Precision	Recall
1D-CNN	0.9140	0.9157	0.8934	0.9392
2D-CNN	0.9531	0.9523	0.9313	0.9742
3D-CNN	0.9702	0.9696	0.9490	0.9911

Table 7.1. CNN model comparison for FPP prediction on training dataset

Model	Accuracy	F ₁ -Score	Precision	Recall
1D-CNN	0.9101	0.9124	0.8893	0.9368
2D-CNN	0.9112	0.9127	0.9072	0.9183
3D-CNN	0.9602	0.9382	0.9605	0.9551

Table 7.2. CNN model comparison for FPP prediction on testing dataset

The 1D-CNN model receives a one-dimensional array of spectral values of Sentinel-2 satellite images as an input data. Spectral bands represent an array of features characteristic for each individual pixel of the observed area. The disadvantage of this approach is that it cannot fully capture the spatial relationships between neighboring pixels. The proposed 1D-CNN model shows consistent performance across both datasets with accuracy and F₁-Score around 0.91. As for the precision, it achieves a relatively lower value compared to the previous two metrics, i.e. 0.8934 for the train dataset and 0.8893 for the test dataset. This suggests more false positive predictions of potential fire propagation. Recall achieves a higher score compared to precision of approximately 0.93 for both datasets suggesting good sensitivity of the model in detecting areas of potential fire propagation.

The 2D-CNN model treats Sentinel-2 images as a stack of 2D images, where each individual image represents a spectral band. In this way, the model can capture the spatial

patterns and textures within each individual band. 2D-CNN shows improved performance over 1D-CNN in the training dataset where it achieves an accuracy of 0.9531, while for the testing dataset it achieves similar accuracy to 1D-CNN indicating that its performance degrades in the test dataset. The same is for the F_1 -Score parameter, while in the training data set it reached a value of 0.9523, in the test data set its value decreases reaching a value of approximately 0.91 similar to 1D-CNN. Lower values of the metrics in the test dataset indicates a certain degree of overfitting to the training data. On the other hand, the precision is improved compared to 1D-CNN on both datasets, indicating better precision in positive predictions. However, recall has a high value on the training set of 0.9742, while on the test set it drops to 0.9183 again indicating a certain degree of overfitting.

The 3D-CNN model extends the input data considering the spatial and spectral dimensions of the Sentinel-2 data as a 3D volume. The proposed model is successful on both datasets in all metrics. It achieves the highest accuracy value of 0.9702 for the training dataset and 0.9602 for the test dataset. It also achieves the highest F_1 -Score, the highest precision and the highest recall among all models on both datasets. High precision indicates very few false positives predictions. The high score on the test dataset indicates good generalization and reliability of the 3D-CNN model in predicting fire propagation potential.

7.3.3. Qualitative Evaluation of FPP Prediction Maps

In addition to the quantitative analysis described in the previous section, a visualization of the predicted fire propagation potential values was also made. For this purpose, Sentinel-2 satellite images taken on March 8, 2022, were used, and the fire that broke out in the area of Dugi Rat in Split-Dalmatia County, Croatia, on March 10, 2022, was analyzed. The fire covered an area of 352 hectares and lasted for approximately two days. The maps were generated using 1D, 2D, and 3D-CNN models in Python, and the visualization was performed in QGIS. Since 1 km x 1 km patches were used, for better visualization, the central pixel of each planned section was taken, after which Inverse Distance Weighted (IDW) interpolation was performed.

Figure 7.9 shows the fire propagation potential prediction map generated by the 1D-CNN model. This model correctly predicts high fire propagation potential in the region of the Dugi Rat fire, marked by the blue polygon, but it can be noted that the map shows a high degree of fire propagation potential across the whole map. There is little variation in the prediction; i.e., the FPP is 1 for most areas, which is highlighted in red. It can be concluded that the model overestimates fire risk and has limited ability to distinguish between different levels of FPP in the landscape, leading to many false positives.

Figure 7.10 shows the fire propagation potential prediction map generated by the 2D-CNN model. It can be noted that the mentioned model shows a very low degree of fire propagation potential on the entire map, which is indicated by a predominantly yellow color.

Only a few predicted sections indicate areas of high fire propagation potential, marked in red, but they do not correspond to the actual location of the fires in Dugi Rat. Therefore, the model underestimates the fire risk, which can lead to many false-negative results.

Figure 7.11 shows the fire propagation potential prediction map generated by the 3D-CNN model. While the model indicates varying levels of fire propagation potential in the Dugi Rat fire area (outlined in blue), it shows a mixed pattern of both high (FPP = 1) and low (FPP = 0) risk predictions within this zone. The distribution of fire propagation potential across the entire map appears more balanced between the two classes compared to the previous models. However, the spatial resolution of the predictions (limited by the Sentinel-2 resolution) presents challenges for detailed local-scale risk assessment. The model provides a broader overview of potential fire risk patterns rather than precise location-specific predictions.

Compared to the generated maps, it can be observed that the 3D-CNN model offers more balanced predictions, avoiding overestimation of the FPP, which occurs with the 1D-CNN model, or underestimation of the FPP, as is the case with the 2D-CNN. A likely reason for the better predictions of 3D-CNN lies in its ability to simultaneously process both spatial and spectral information from the input data.

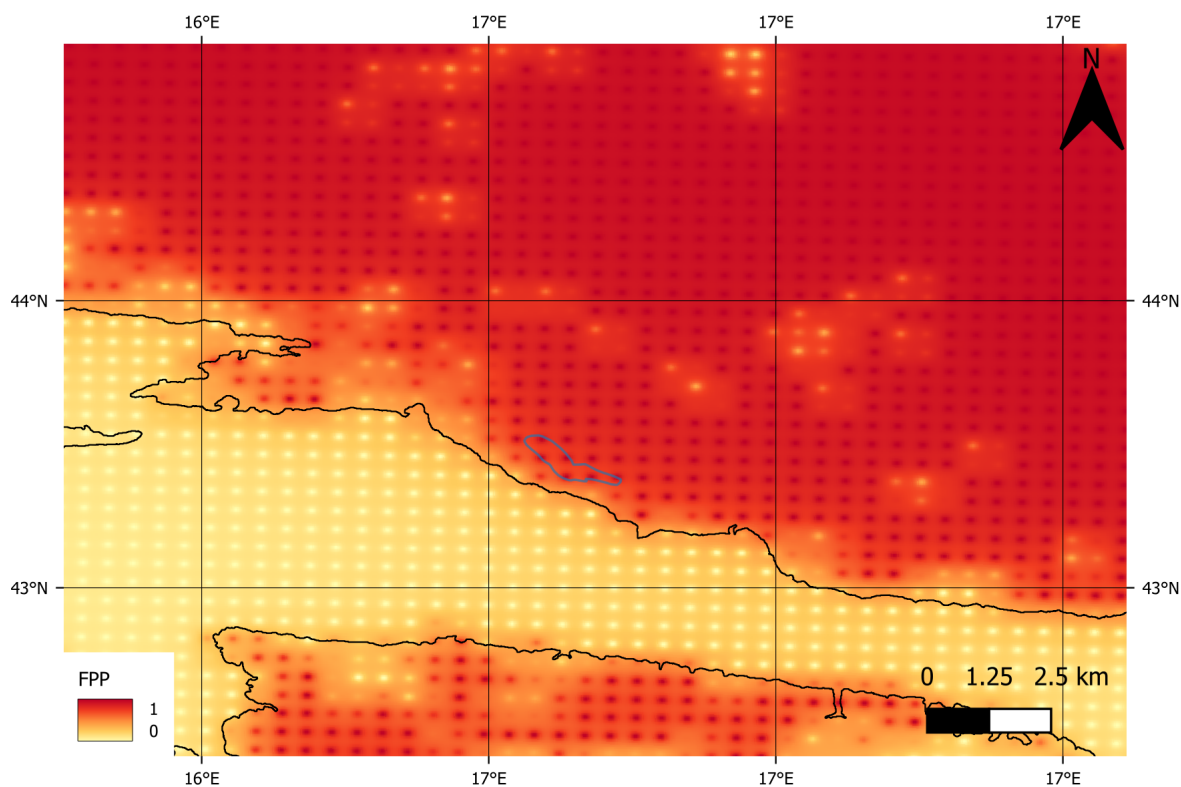


Figure 7.9. Fire propagation potential prediction map using 1D-CNN model

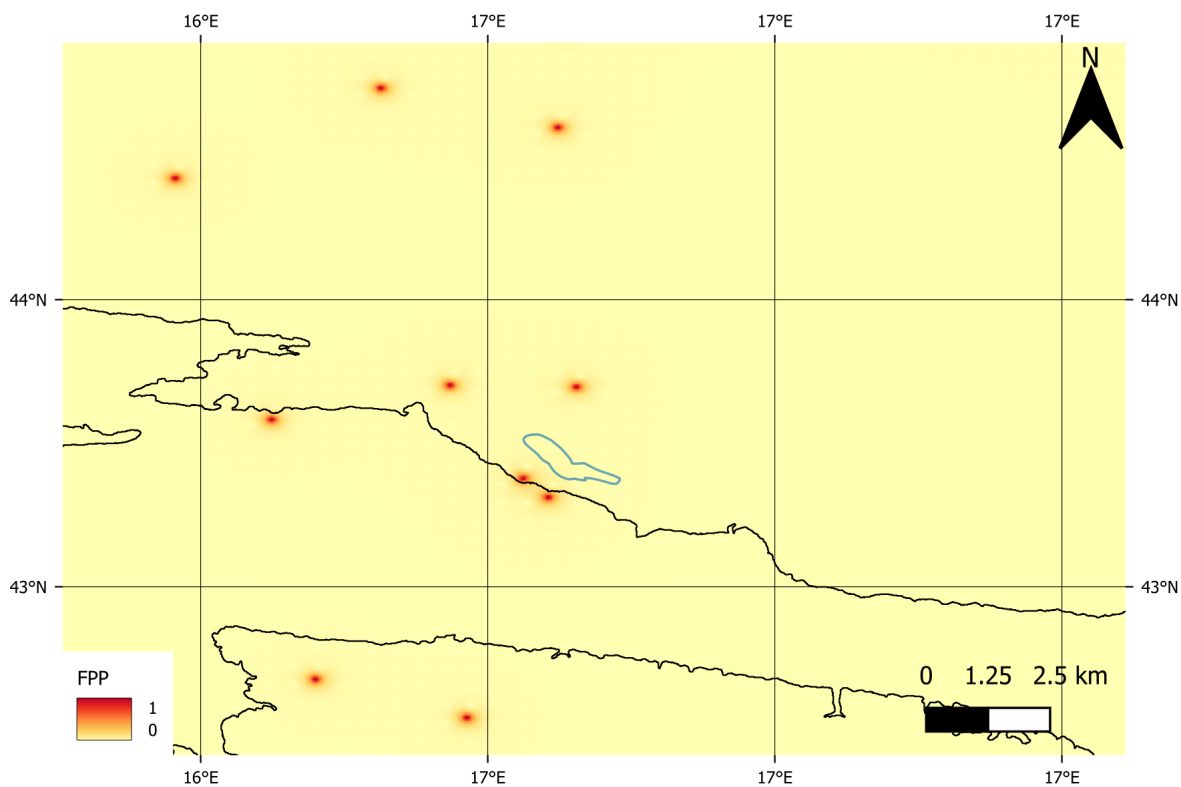


Figure 7.10. Fire propagation potential prediction map using 2D-CNN model

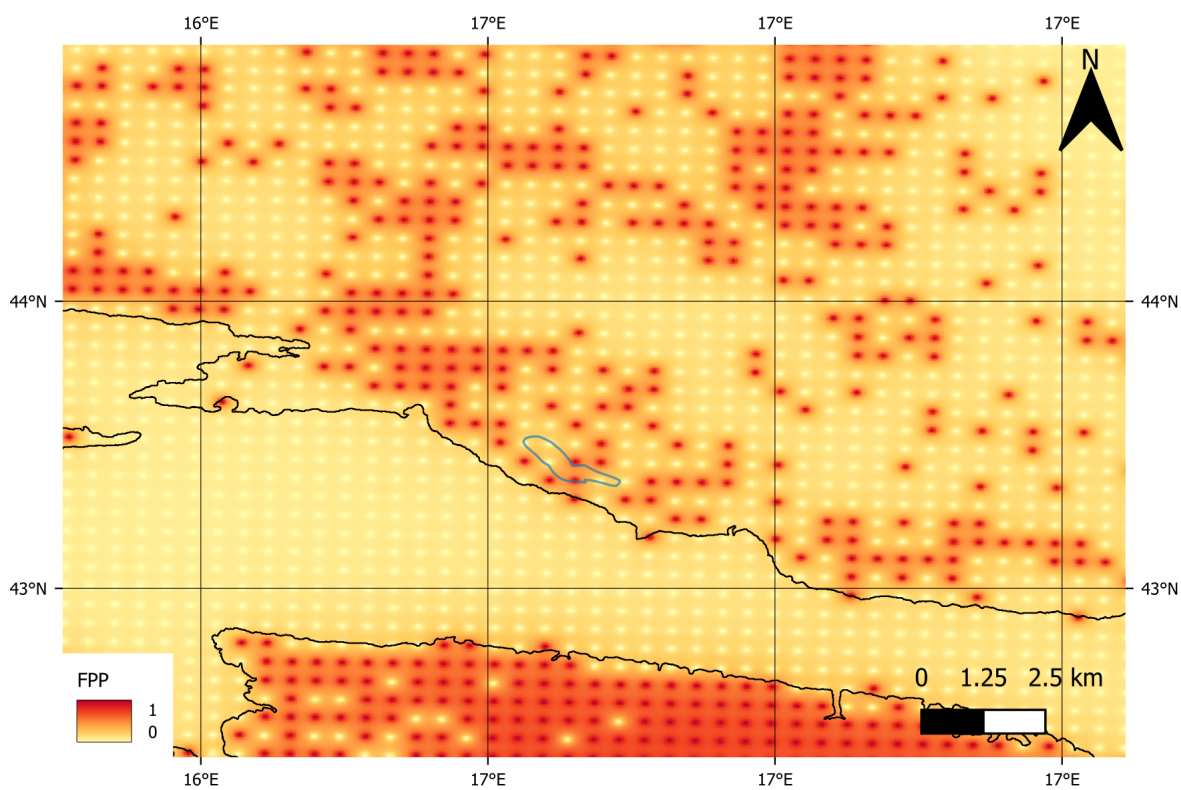


Figure 7.11. Fire propagation potential prediction map using 3D-CNN model

7.4. Discussion

This section will present a discussion of the proposed 3D-CNN model for the purpose of predicting fire propagation potential and will compare it with 1D and 2D-CNN models. Limitations of this approach will also be described and directions given on what could be changed in future research to achieve a better and more optimal model.

7.4.1. Accuracy of the 3D-CNN Model

The 3D-CNN model achieves an overall accuracy greater than 95% on both datasets, with minimal deviation between the training and testing datasets. Also, for other metrics, the model achieves high results above 93% on both data sets. This indicates good generalization, robustness and stability of the model, which successfully applies the learning from training to unseen test data. The ability of the 3D-CNN model to process data as a spatio-spectral volume and find spatial and spectral relationships in it gives it an advantage over 1D and 2D-CNN models.

During a qualitative analysis for a specific fire incident that occurred in Dugi Rat in 2022, it could be observed that the 3D-CNN model better recognizes more complex spatial patterns of fire risk compared to 1D-CNN and 2D-CNN, which make extreme predictions of one of two FPP classes. Although 3D-CNN also shows some degree of overprediction of fire spread potential as high risk in some parts of the observed area, the nuanced predictions provided by the 3D-CNN model likely lead to higher accuracy compared to the other two models.

7.4.2. Limitations

Although the 3D-CNN shows promising results in predicting fire propagation potential based on Sentinel-2 satellite imagery compared to the 1D-CNN and 2D-CNN models, several limitations of this study must be acknowledged. First, the input data are the spectral values of all Sentinel-2 bands, which may be limited to those that influence fire propagation potential. In addition, the bands were resampled to a resolution of 60 m using patches of size 17 x 17 as input data, which corresponds to an area of 1 km x 1 km. An area of one square kilometer can be too big for capturing some relevant information about aspects that impact fire propagation potential.

Moreover, the current model may not adequately capture temporal variations in fire propagation potential, especially since the Sentinel-2 satellite images are available every 5 days. Finally, it should be considered that the performance of the model may vary when applied to regions with different environmental characteristics or fire propagation patterns than those represented in the training data.

7.4.3. Future Research Directions

In future research, several promising options could be explored to achieve a better model of fire propagation potential. First is data, where using smaller patches could potentially capture more factors that influence fire propagation potential. This could be computationally more challenging, as the number of data would increase, however it could lead to significant improvements in the accuracy of the 3D-CNN model. A crucial improvement could be made in the definition of ground truth data, which is currently limited to areas where fires have actually occurred. This leaves many potentially high-risk areas undefined in the training data. Future work could incorporate expert assessments or develop standardized evaluation methods to identify and label high-risk areas that haven't experienced fires, providing more comprehensive training data.

In addition to the above, the input data could be expanded by taking into account other data sources such as weather information (temperature, humidity, wind speed and direction), radar data for vegetation structure, topographic data or other satellite data (e.g. Landsat, MODIS). By expanding the data set in this way, the semantics of the input data is enriched which enables the model to gain a deeper understanding of the complex factors that influence fire propagation potential.

In addition to data, hybrid models combining CNNs with Recurrent Neural Networks (RNNs) or using transfer learning could be taken into account to further improve model performance. Also, in order to achieve the generalization of the model and its applicability in the world, it is necessary to conduct studies in different geographical areas. These are just some suggestions that could improve the proposed 3D-CNN model and thus help in the strategy to reduce the risk of potential fires.

7.5. Conclusion

This chapter describes the implementation and evaluation of the 3D-CNN model for the purpose of predicting the fire propagation potential in the Split-Dalmatia County.

Based on the quantitative analysis, the model showed high accuracy and high results in other metrics such as precision, recall and F₁-Score, which suggests successful learning of the model from the training data. On the other hand, based on the qualitative analysis that was conducted on the basis of the real fire that happened in Dugi Rat in the Split-Dalmatia County in March 2022, the model showed a certain degree of excessive adjustment of the data.

In order to further evaluate the proposed model, it was compared quantitatively and qualitatively with 1D-CNN and 2D-CNN models. Compared to the mentioned models, 3D-CNN showed better performance in both quantitative and qualitative model analysis. The reason for this may be the limitation of the 1D-CNN model that it only takes into account spectral

information, and the 2D-CNN model only spatial information, while the 3D-CNN simultaneously processes both the spectral and spatial characteristics of satellite images. This is presented in the qualitative analysis, where the visualization of the predicted data of these models highlighted their differences. The 1D-CNN and 2D-CNN models had extreme predictions, that is, one class of FPP was dominant on the map. On the other hand, the 3D-CNN model provided a more nuanced map.

It should be emphasized that the current version of the proposed 3D-CNN has its limitations, which could be solved in future work. For example, expanding the data set with weather or topographic data could result in greater accuracy and stability of the model. Also, work could be done on the geographical diversity of the same data, so that the mentioned model could be applicable in different geographical regions.

In conclusion, the 3D-CNN model proved to be a good choice for the purpose of predicting fire propagation potential that has the potential to monitor fire-related risks in real-time that could help fire management agencies.

8. EVALUATION OF 4D-CNN MODEL

This chapter describes the application of a 4D convolutional neural network for the detection of burned areas based on data from the Sentinel-2 satellite. This task is a representative of a *scene understanding* problem and change detection, since it requires putting an event in spatio-temporal context. Different patch sizes and their impact on model performance are compared, with an analysis of the relationship between fire size and the selection of an appropriate patch size.

The chapter is organized as follows. The introduction provides an overview of existing methods, from traditional approaches and machine learning to deep learning, which have been used in the analysis of forest fires. Studies applying convolutional neural networks to remote sensing data are also presented, with a special focus on 4D-CNN. The Materials and Methods section presents the research area, description of the data, their processing and preparation for the model, as well as the implementation of the model itself. This is followed by a Results section, where quantitative and qualitative analyses are presented. In the Discussion section, the accuracy of the proposed model is analyzed, its limitations are identified, and directions for future research are suggested. The chapter concludes with a summary of findings.

8.1. Introduction

Changes in frequency and extend in wildfires create demand for novel tools capable of rapid analysis. Accurate and timely detection of burned areas is important when assessing ecosystem recovery and understanding fire spread after a fire has occurred. Remote sensing data enable biophysical measurements of soil conditions before and after fire. The remote sensing measurements are often used in fire risk mapping, fuel mapping, active fire detection, burned area assessment, burn severity assessment and vegetation recovery monitoring [310].

Traditional methods based on statistics have shown effectiveness in prediction, the occurrence, and risk assessment of forest fires. Examples of the application of traditional methods are linear regression for spatial patterns of fire occurrence [301], Poisson regression for human-caused fires [311] Monte Carlo simulations for fire hazard assessment [312] and spectral index and thresholding [313–315].

Compared to the mentioned techniques, machine learning techniques have proven to be

a powerful tool. Methods such as Multilayer Perceptron (MLP) show promising results in forest fire probability mapping [316, 317]. Also, using algorithms such as Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM) have improved forest fire mapping and burning area labeling using satellite data [318–322].

Deep learning has made a breakthrough over the traditional methods and machine learning techniques primarily due to the efficient processing of time series data such as remote sensing images. Deep learning techniques are used for various tasks of temporal analysis of satellite images, such as crop classification, land cover change detection, and urban area monitoring [323–325]. Also, various deep model architectures have been explored and described in the literature, from sequential recurrent encoders that process time sequences to convolutional neural networks designed for time series analysis [326–328].

In the context of forest fires, these achievements are particularly valuable. Convolutional neural networks have proven to be effective in processing radar and optical data, such as data from the Sentinel-1 and Sentinel-2 satellites [329], enabling near-real-time monitoring of fire progress [330]. Furthermore, burned areas can be precisely mapped through time sequences of satellite images [331, 332]. Deep learning techniques enable simultaneous processing of the spectral, spatial and temporal dimensions of satellite imagery, capturing complex patterns that traditional methods may miss.

1D, 2D, 3D and 4D-CNN models process remote sensing data which are usually spatial, spectral and temporal. 1D-CNNs process only one dimension, which is usually spectral or temporal, while 2D-CNNs often process spatial data, consisting of two dimensions - width and height. 3D-CNN can process multidimensional data but is limited in capturing all relevant information at the same time, because it requires a trade-off between the spectral and temporal dimensions. On the other hand, 4D-CNN extends the capabilities of 3D-CNN by adding an additional dimension, allowing the simultaneous processing of spatial, spectral, and temporal information. Although this method shows significant potential in remote sensing, the number of studies dealing with the application of 4D-CNN in this area is still limited, which indicates the need for further research.

For example, one study describes implementation of 4D U-Net architecture to land cover classification based on Landsat-8 satellite data. The proposed 4D U-Net architecture significantly outperforms lower-order U-Net models (e.g. 2D U-Net, 3D U-Net) and some of the current state-of-the-art methods (e.g. FCN, FCN + LSTM) [86]. Another study introduces a 4D Fractal CNN for predicting the eutrophication state of water bodies, highlighting the advantages of the model in processing spatio-temporal characteristics and interactions between multiple factors [120]. Also, the experimental analysis shows that the 4D-CNN architecture enables effective modeling of spatial, spectral and temporal features, thereby outperforming existing 3D and 2D approaches in the classification of land covers based on multispectral data [89].

In this study, a proposed 4D-CNN architecture is designed for burned area mapping using

multi-temporal Sentinel-2 imagery. The main motivation for using this type of architecture is its ability to process four-dimensional data. This makes proposed architecture suitable for analyzing pre and post fire Sentinel-2 imagery, where changes in spectral signatures across time and space are key indicators of burned areas. The main research questions defined in this chapter are:

- RQ1: How accurately can burned areas be detected using 4D-CNN and multi-temporal Sentinel-2 data?
- RQ2: How does the patch size affect the model's ability to detect and map burned areas of different spatial sizes?

8.2. Materials and Methods

8.2.1. Study Area

The study area includes fires that occurred in Split-Dalmatia County, Croatia. Figure 8.1 shows these fires labeled as Site 1, which is related to the fire marked in orange, and Site 2, which is related to the fires marked in red. Site 1 was used to train the 4D-CNN model, and that fire event occurred on March 13, 2022. Site 2 was used for additional evaluations, and those fire events occurred on March 19, 2022. The fire at Site 1 covered an area of 2304 ha, while the fires at Site 2 were 7 ha, 29 ha, and 68 ha in size.

8.2.2. Dataset Construction and Preprocessing

The datasets were constructed from Sentinel-2 imagery and labeled based on the fire data obtained from the European Forest Fire Information System (EFFIS) database. Sentinel-2 images were downloaded from the Copernicus Data Space Ecosystem (Copernicus Sentinel data [2022] [284]), which provides open and free access. Images were downloaded for dates March 13 and 23, 2022 representing a scene before and after fire event. These scenes were chosen because there was no cloud coverage and they were the closest dates of occurred fire event. Dataset was limited only for these scenes because it provides sufficient information for answering RQs. These are Level-2A images which are atmospherically corrected and with a spatial resolution of 10 m where bands that were not originally at 10 m resolution were resampled. All images are downloaded in .jp2 format. Visible bands (B02, B03, B04), near-infrared (B08), and short-wave infrared (B11 and B12) were used. Visible bands help detect surface changes and assess vegetation health, which is useful in detecting changes caused by fires. The near-infrared band helps distinguish healthy from fire-damaged vegetation, while the short-wave infrared bands help detect thermal anomalies and burned areas. For cooled soil after a fire, these bands will show low reflectance, while for still-hot areas, the reflectance

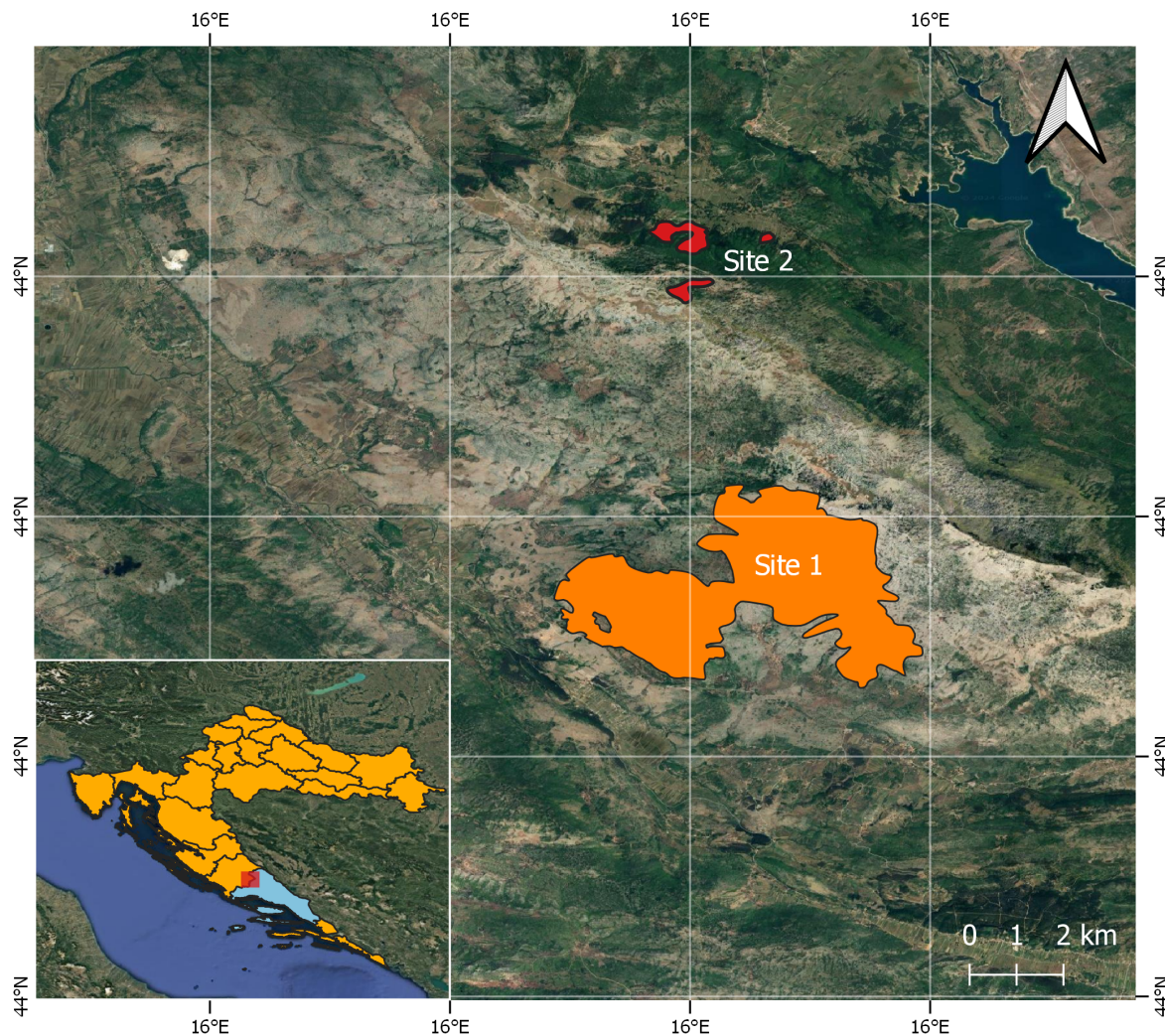


Figure 8.1. Study area representing fires in Split-Dalmatia County, Croatia, March 2022 (shown in Plate Carrée projection)

will be high. Although there are various indices in the literature for detecting fires and their consequences, such as Normalized Difference Vegetation Index (NDVI), Burned Area Index (BAI) and Normalized Burn Ratio (NBR), which are based on specific bands, in this paper we will focus to the use of direct bands [333, 334].

The EFFIS dataset presents the ground-truth values of burned area. Patches extracted from the Sentinel-2 images related to burned area are labeled as 1, while other parts of the Sentinel-2 images are labeled as 0, indicating non-burned area.

In order to evaluate model performance at different patch sizes, there were constructed three datasets with patch sizes of 10×10 , 32×32 and 64×64 pixels. For each fire incident the same number of patches was extracted from pre-fire image and post-fire image, creating 4D data with the shape (number_of_samples, number_of_bands, patch_size, patch_size, number_of_time_frames). For each patch size the shape of the whole dataset was (1012, 6, 10, 10, 2), (1156, 6, 32, 32, 2) and (356, 6, 64, 64, 2). Each sample contains 12 band values in total - 6 bands for each time frame, where the patch size is either 10×10 , 32×32 , or 64

$\times 64$ pixels. The dataset for each patch size was split as 60:20:20, where 60% of data was used for model training, 20% for validation, and 20% for model testing to avoid overfitting.

8.2.3. Model Implementation

At the time of writing this doctoral dissertation, well-known frameworks such as TensorFlow and PyTorch did not support direct 4D convolutions. To overcome this limitation, a pseudo-4D approach was implemented using 3D convolutional neural networks that process multiple time instants sequentially. While traditional 3D-CNNs handle three dimensions (height, width, and spectral bands) at once, this approach is extended by incorporating the temporal dimension through sequential processing of data from different time steps. This method allows four dimensions to be effectively modeled: spatial (height and width), spectral, and temporal characteristics of the data simultaneously. The temporal information is particularly important as changes in the landscape over time are captured.

In this section the proposed architecture of 4D-CNN will be described only for dataset containing 32×32 patch sizes to avoid repetition, because same architecture was used for datasets containing 10×10 and 64×64 patch sizes. Figure 8.2 illustrates the proposed 4D-CNN architecture.

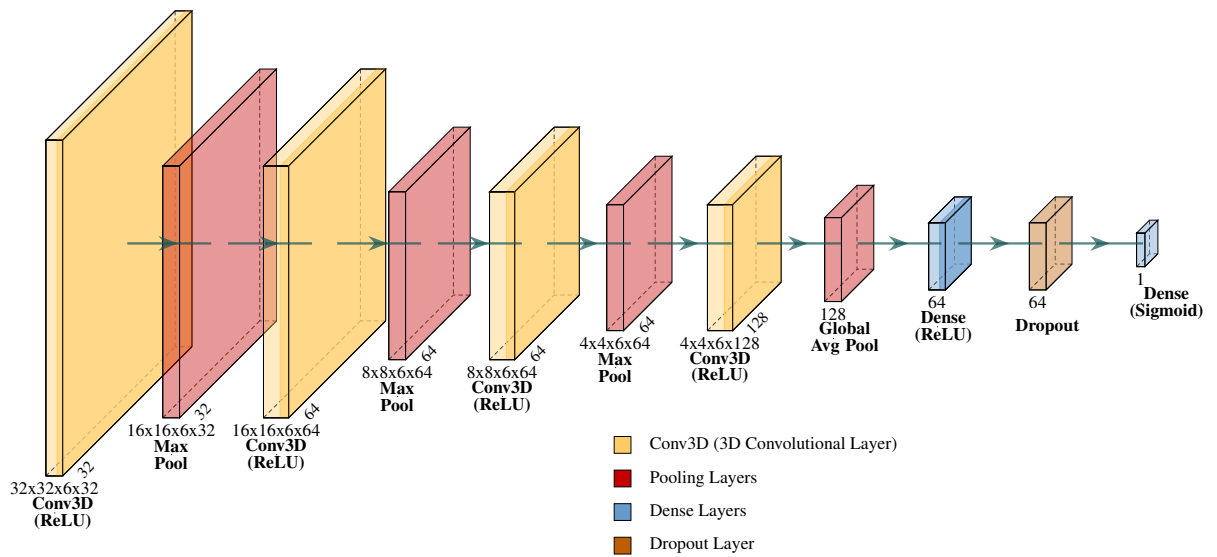


Figure 8.2. 4D-CNN Architecture

This architecture processes input patches of size $6 \times 32 \times 32 \times 2$, representing Sentinel-2 pre- and post-fire images. Dimension 6 represents the number of spectral bands, 32×32 represents the spatial dimension, and 2 represents two time frames (before and after the fire event). As mentioned before, 4D convolutional layers are not available in well-know Python frameworks so 3D convolutional layers are used to extract spectral-spatial-temporal features. The network has a total of four 3D convolutional layers. The first 3D convolutional layer uses 32 filters and a $3 \times 3 \times 3$ kernel with ReLU activation function. This layer is followed by

a MaxPooling layer that reduces the spatial dimension, but at the same time preserves the spectral and temporal dimensions. In following two 3D convolutional layers, the number of filters is increased to 64, which are also followed by a MaxPooling layers that reduces the spatial dimension to 8×8 and 4×4 , respectively. Furthermore, in the last 3D convolutional layer there is an additional increase in the number of filters to 128, which is followed by a Global Average Pooling 3D layer that reduces dimensionality by averaging all spatial and temporal dimensions, resulting in a 1D vector of 128 features. After feature extraction, the model applies a fully connected (dense) layer with 64 units and a ReLU activation function, followed by a dropout layer with a rate of 0.5 to prevent overfitting. In the last dense layer which is also the output layer of one unit, the Sigmoid activation function was used for the binary classification of burned (1) or unburned (0) areas. All convolutional layers use the 'same' padding to maintain spatial dimensions before pooling. The total number of parameters that can be trained in the model is 397,409. In this 4D-CNN, all three key aspects spectral, spatial, and temporal are learned and used for more accurate predictions of burned and nonburned areas.

8.3. Results

This section describes the results of the proposed 4D-CNN model for predicting burned areas based on Sentinel-2 imagery. The metrics described in Section 6.2.4 were used to evaluate the effectiveness of the proposed model, namely Accuracy, F_1 -Score, Recall, Precision, Confusion Matrix, AUC and ROC curve. The comparison was made on the train and test datasets for different size of patches. Model is trained and tested on a system with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and 16 GB of RAM, without GPU acceleration. Due to the relatively small dataset size, consisting of only pre-fire and post-fire images, the complete training and testing process took less than 10 minutes for all patch sizes.

8.3.1. 4D-CNN Model Performance

The proposed 4D-CNN model shows strong performance in detecting burned areas in the Split-Dalmatia County area for different patch sizes. Figure 8.3 shows the confusion matrices for the training and testing datasets for different patch sizes: 10×10 , 32×32 and 64×64 , respectively. The confusion matrix indicates how well the model classified burned and unburned areas. Analyzing the confusion matrices for the training dataset (first column), it can be observed that for the patch size of 32×32 the model had the highest number of false positive and false negative classifications. On the other hand, for the testing dataset (second column), the confusion matrix for the 10×10 patch size shows the highest number of false positive and false negative classifications.

Figure 8.4 shows the detailed performance of the metrics for different patch sizes, as well

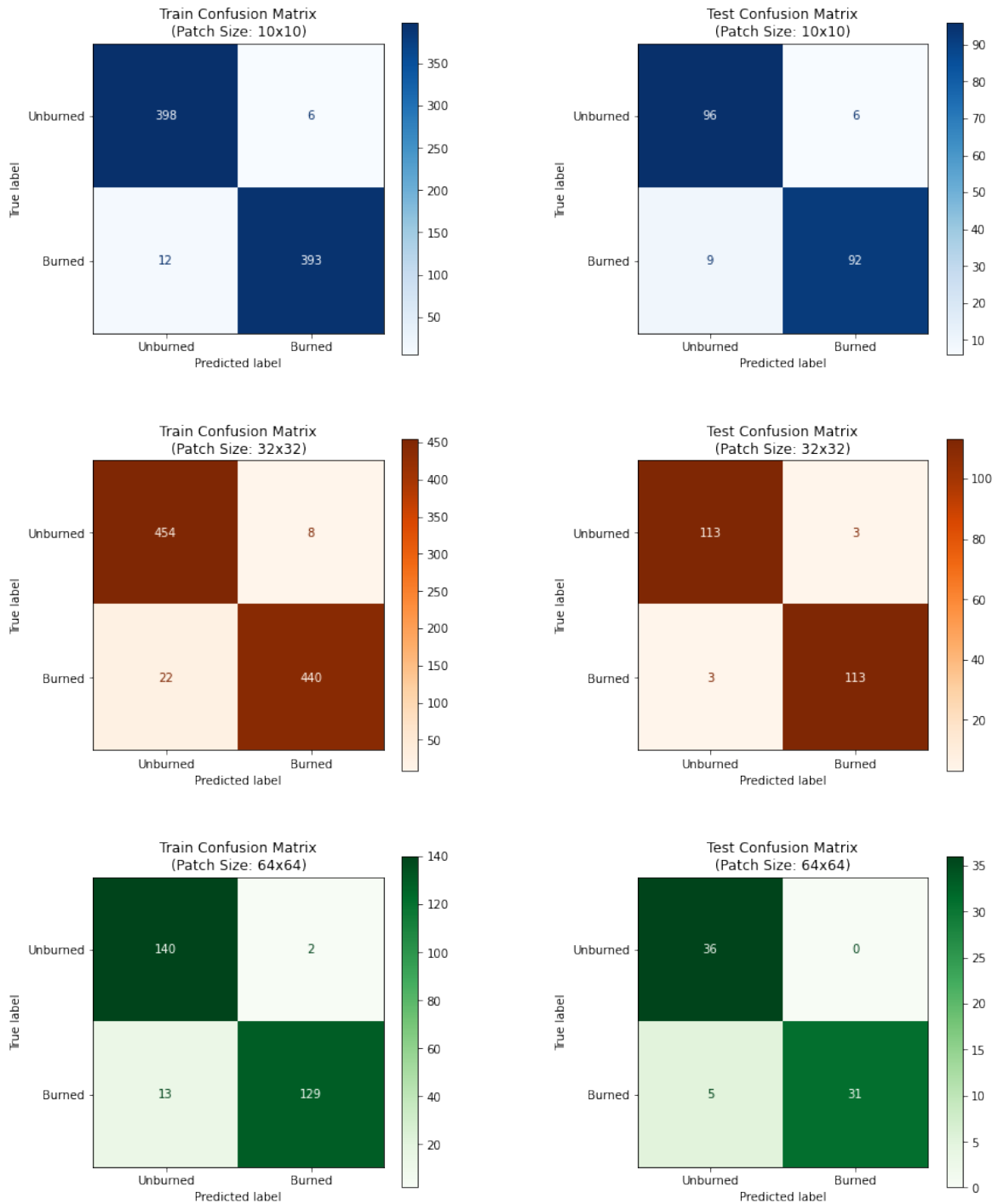


Figure 8.3. Confusion matrices for the 4D-CNN model illustrating classification performance for various patch sizes in detecting burned and unburned areas

as individually for the 'Burned' and 'Unburned' classes, on the training and testing datasets. The model achieved the highest overall accuracy of 0.98 on the training dataset for the 10×10 patch size, while for the 32×32 patch size it achieved the highest accuracy of 0.97 on the testing dataset. It can be observed that a balanced overall accuracy in data classification is achieved for a patch size of 32×32 in both datasets. This optimal configuration allowed the model to maintain a balanced performance in all metrics for a 32×32 patch size, with values

greater than 0.95 achieved for both 'Burned' and 'Unburned' classes. In addition, the model shows consistent performance at different patch sizes, with the performance difference less than 5% between the training and testing datasets.



Figure 8.4. Performance metrics of the 4D-CNN model for various patch sizes on train and test datasets

Further analysis of the model's classification performance was performed using receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values. This analysis was performed for both the train and test datasets to provide a comprehensive evaluation of the model's effectiveness. Figure 8.5 shows ROC curves and AUC values for each patch size.

For the smallest patch size of 10×10 , the ROC curves for both datasets show a sharp rise in the upper left corner, indicating a high ability of the model to discriminate between burned and unburned areas. The AUC values show that the model achieved a perfect score of 1.00 on the training dataset, while it achieved an AUC of 0.97 on the testing dataset.

For the medium patch size of 32×32 , the ROC curves for both datasets have almost

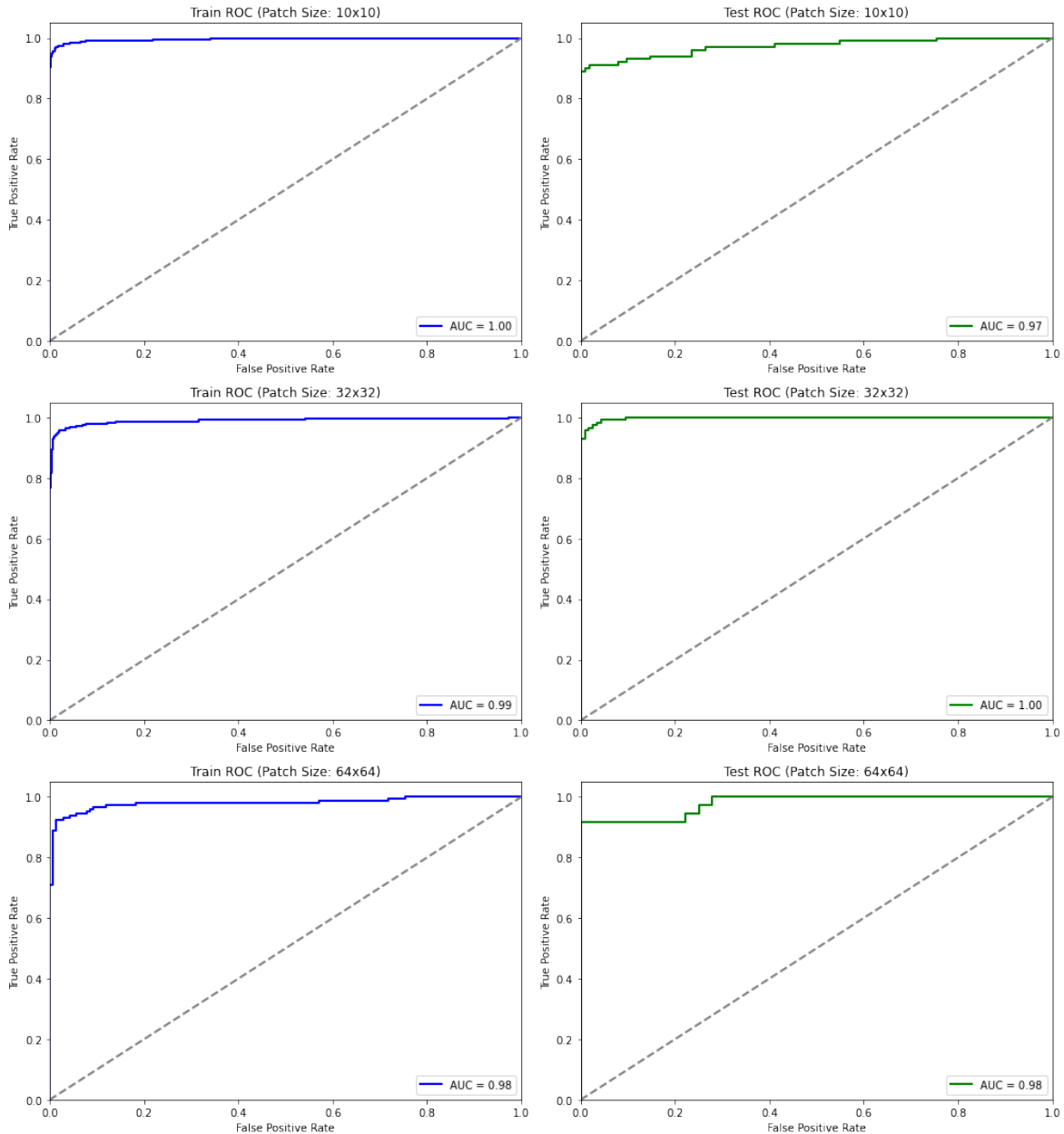


Figure 8.5. ROC curves with AUC values for the 4D-CNN model on different patch sizes for both train and test datasets

identical patterns with a steep rise at low false positive rates, suggesting optimal classification performance. The AUC value on the training dataset was 0.99, while on the testing dataset the model achieved a perfect score of 1.00.

For the largest patch size of 64×64 , the ROC curves show similar patterns to the 32×32 patch size, but with a slightly more gradual increase. This is particularly evident on the test dataset at lower false positive rates. Despite this, the model maintains consistent performance with AUC values of 0.98 for both datasets.

It is worth noting that all patch sizes achieved AUC values greater than 0.97. This indicates excellent classification performance with respect to the spatial resolution of the dataset.

Moreover, this suggests that the model is suitable for detecting burned areas for different patch sizes, allowing it to adapt to different scales of analysis with exceptional results.

8.3.2. Qualitative Analysis

Qualitative analysis of the 4D-CNN model, based on the visualization of predicted values, reveals how well the proposed model predicts burned areas on Sentinel-2 datasets. The analysis is performed for all three datasets with different patch sizes (10×10 , 32×32 , and 64×64), which were used during model training. The ground truth data for actual burned areas are marked as Site 1 and Site 2 with white polylines representing burned areas.

Figure 8.6 shows the map generated from the predictions for the dataset where the patch size was 10×10 . The model successfully detected a large burned area (Site 1), where the model effectively captured the intricate boundary details within the burned area, which relate to areas that were not affected by the fire. At Site 2, the model successfully detected two larger fires, while the smallest one was not detected. For patches of size 10×10 , it is observed that the model creates a fine-grained pattern, evenly distributed, with a certain distance between the patterns. However, such a pattern results in fragmented representations, which is particularly noticeable in more densely burned areas such as Site 1, and may lead to misinterpretation of burned areas. In both areas, Site 1 and Site 2, the model also produces false positive predictions outside the actual burned areas.

Figure 8.7 shows a map generated from the model's predictions for patches of size 32×32 . The model successfully detected the large fire at Site 1, while at Site 2 it failed to detect the smallest fire. Compared to the map generated for 10×10 patches, the model better connects the detected burned areas, which contributes to spatial continuity and more precise boundaries of burned areas. The detection pattern shows improved coherence in the central burned area, with clearer distinction between burned and unburned patches. Thus, the 4D-CNN model for 32×32 patches strikes a better balance between detail retention and generalization, although it still produces a certain number of false positive predictions.

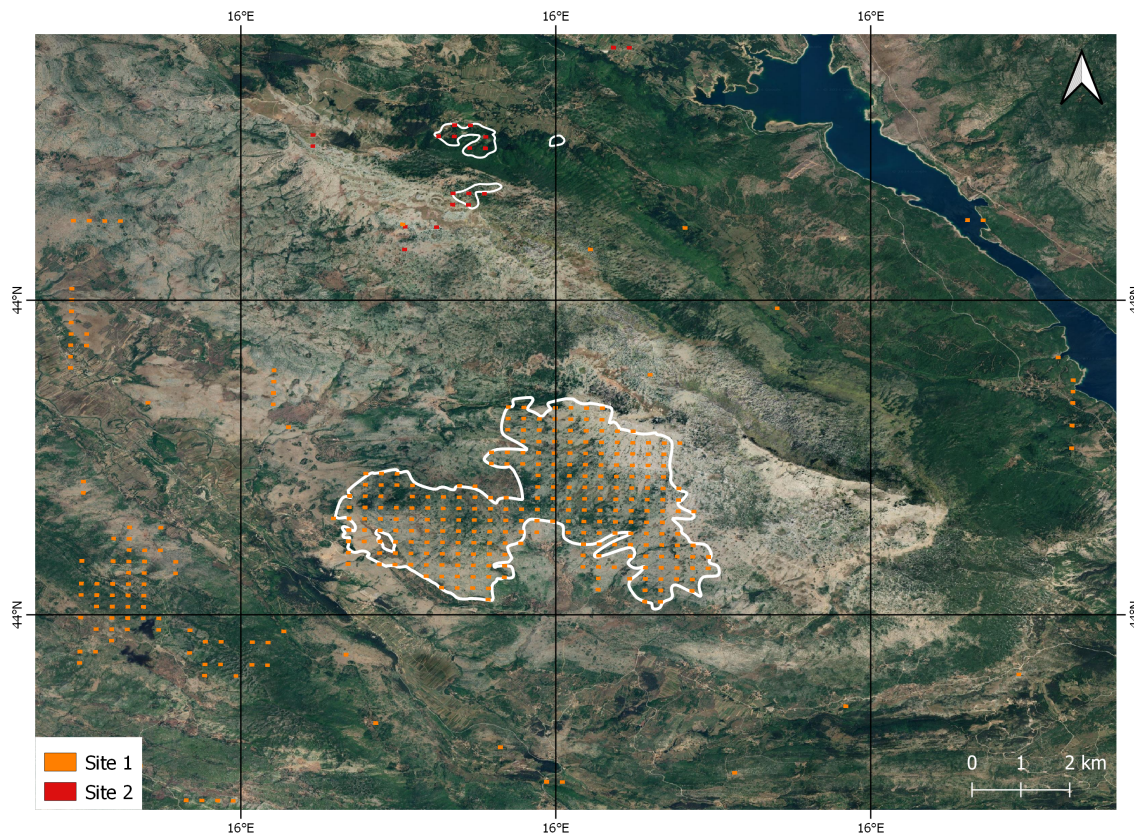


Figure 8.6. Predicted burned area map using a 10×10 patch size

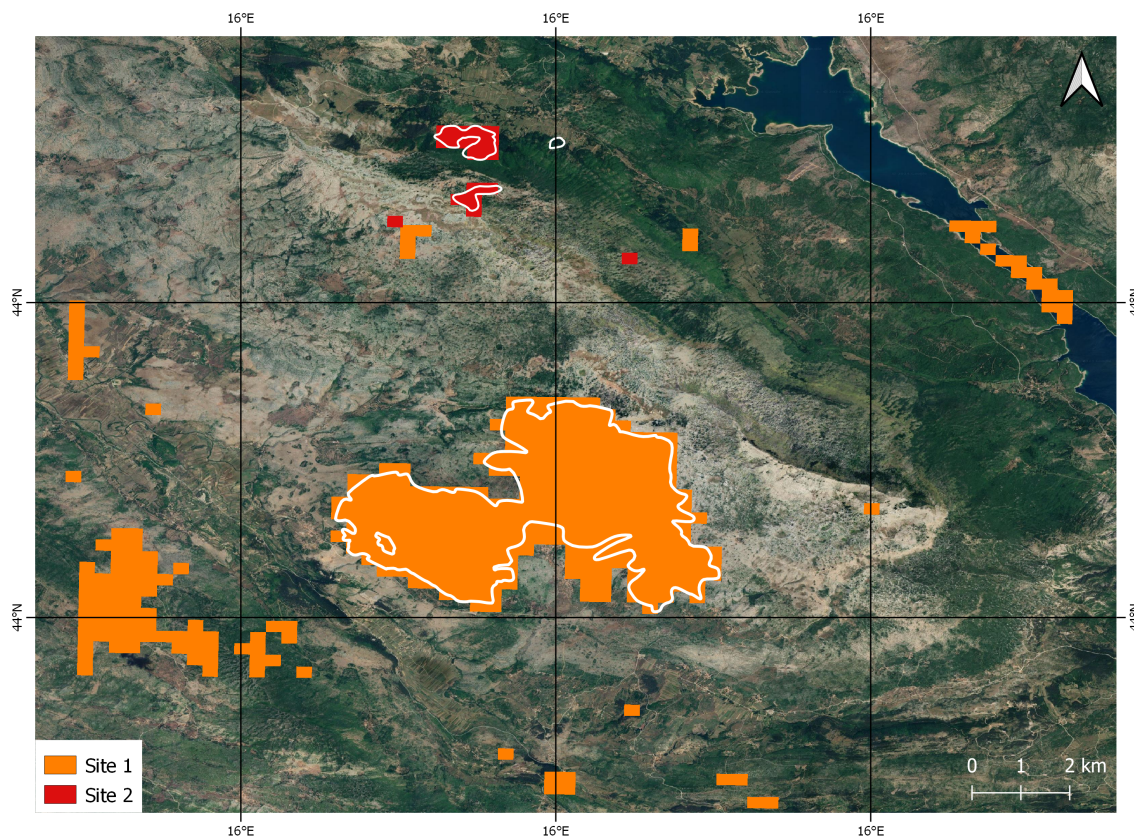


Figure 8.7. Predicted burned area map using a 32×32 patch size

Figure 8.8 shows the map generated for patches of size 64×64 . As for the previous two patch sizes, the model successfully detected all fires except the smallest one located at Site 2. It can be noted that the model for this patch size tends to generalize more, although the predicted values are not as smooth as for the 32×32 patch size, because of their size, they give a blocky, rectangular pattern in the figure. The larger patch size creates more continuous detection areas but may oversimplify complex burn patterns, especially visible in the transition zones between burned and unburned areas. Given the size of the patch, false positive predictions will indicate a larger burned region than it actually is, which may lead to an exaggeration of burned areas, so it is necessary to further verify such results before drawing final conclusions.

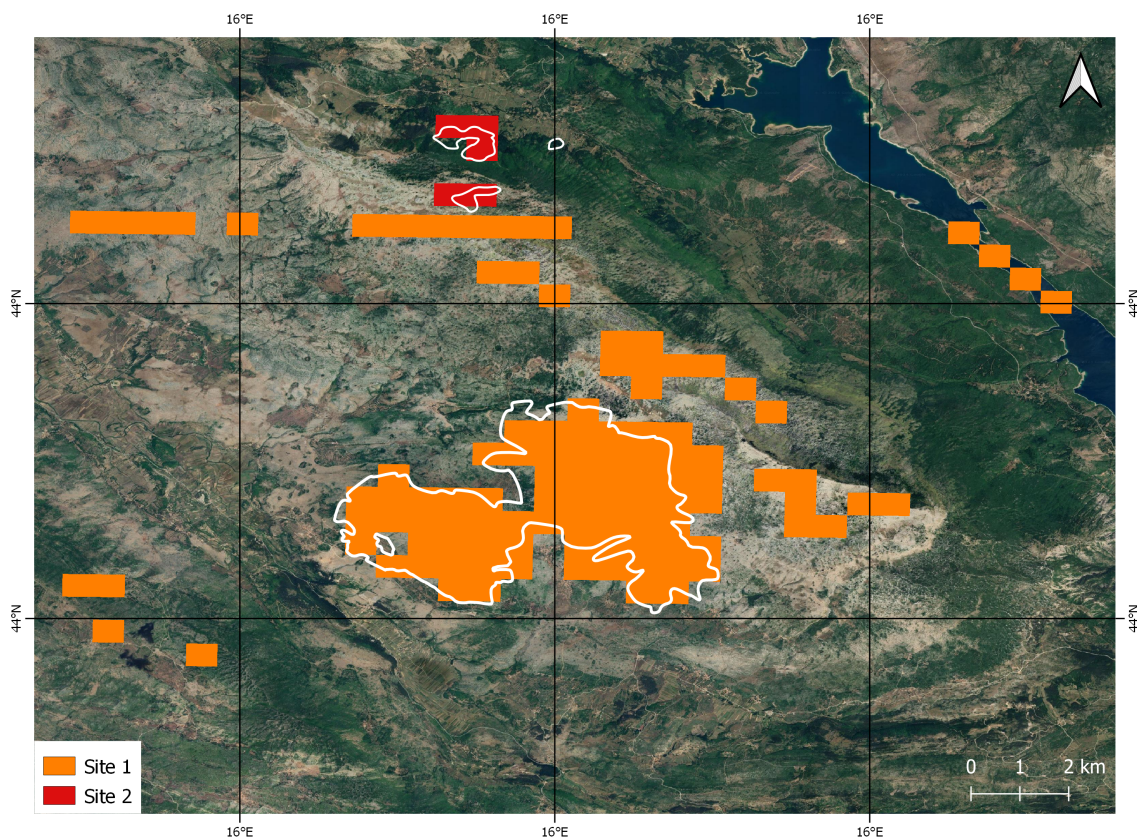


Figure 8.8. Predicted burned area map using a 64×64 patch size

8.4. Discussion

This section presents a discussion of how the 4D-CNN model performs for different patch data sizes and how these affect the spatial patterns of predicted burned areas. Also, the limitations of the model are reviewed and guidelines for future research are proposed, with the aim of further improving the accuracy and efficiency of the model.

8.4.1. Performance and Insights of the 4D-CNN Model

The experimental results demonstrate the significant ability of the 4D-CNN model in detecting burned areas using multi-temporal Sentinel-2 data. In the following, the questions raised in the introduction of this chapter will be discussed, providing insights into the model's performance (RQ1) and the impact of patch size selection (RQ2).

RQ1: *How accurately can burned areas be detected using 4D-CNN and multi-temporal Sentinel-2 data?*

The 4D-CNN model showed exceptional performance in detecting burned areas using deep learning and multi-spatial-spectral-temporal data from the Sentinel-2 satellite. Using Sentinel-2 data with its time resolution of 5 days between images of the same scene proved to be a good choice. This timeframe is too short for significant recovery of the burned area which enables the model to accurately make a change detection.

The proposed model achieved very high scores in all evaluation metrics and dataset configurations, with minimal difference in performance (less than 5%) between the training and testing datasets. The model resulted in the best performance at a patch size of 32×32 , where misclassifications were reduced to only three instances for the categories of false positive and false negative results on the test dataset. The model achieved balanced metrics of 0.97 for Precision, Recall, F1-Score and Accuracy on both datasets (training and testing). Additionally, the model achieves an AUC value of 0.99 on the training dataset and 1.00 on the testing dataset, confirming its robustness in distinguishing between burned and unburned areas.

Compared to existing methods in the literature shown in Table 8.1, the proposed 4D-CNN approach demonstrates competitive performance, matching or surpassing the accuracy of recent deep learning approaches such as the U-Net inspired architecture (Seydi et al. [332], 97.07%). Furthermore, it outperforms traditional methods such as spectral index and thresholding, which achieve accuracies between 84% and 96%, and well-established machine learning algorithms like Support Vector Machines and Random Forest, which achieve accuracies between 91% and 95%. However, these evaluations were not performed on the same dataset, so the algorithms' metrics cannot be directly compared.

RQ2: *How does the patch size affect the model's ability to detect and map burned areas of different spatial sizes?*

Quantitative and qualitative analysis reveal a clear relationship between patch size and model mapping precision. The different patch sizes used in this study offer important insights into the model's ability to detect and map burned areas of different sizes and shapes.

For the smallest patch size of 10×10 the model showed high accuracy of 0.98 on the training data set. However, on the test dataset it drops to 0.93, which is still a high score but indicates some variability in the model's generalization on the test dataset. Confusion matrices show an increased number of false positive predictions, especially in border areas of burned regions. The ROC curves for this patch size show a sharp increase in the upper left corner,

Table 8.1. Comparison of Remote Sensing Methods for Burned Area Detection

Paper	Year	Overall Accuracy	Method	Data Source
Petropoulos et al. [318]	2011	95.78%	Support Vector Machines	Landsat TM
Quintano et al. [313]	2018	84%	Combination of Spectral Index and Thresholding	Landsat-8 and Sentinel-2
Lima et al. [315]	2019	96%	Spectral Index Thresholding	Sentinel-2
Roy et al. [320]	2019	92%	Random Forest Change Detection and Region Growing Techniques	Landsat-8 and Sentinel-2
Barboza Castillo et al. [314]	2020	94%	Thresholding using Spectral Index	Sentinel-2
Syifa et al. [321]	2020	92%	Support Vector Machine and Imperialist Competitive Algorithm	Sentinel-2
Ngadze et al. [322]	2020	92%	Random Forest	Sentinel-2
Seydi et al. [319]	2021	91.02%	Random Forest with Spectral and Spatial Features	Sentinel-2
Seydi et al. [332]	2022	97.07%	Deep Learning Approach inspired by U-Net architecture	Sentinel-2
Proposed 4D-CNN	2024	97%	Deep Learning Approach	Sentinel-2

with an AUC score of 1.00 on the training set and 0.97 on the test set, confirming the high discriminative ability of the model. Smaller patches allow model to recognize finer details, but increase sensitivity to noise and local variation, which can reduce model's accuracy.

For the medium patch size of 32 x 32, the model showed consistent performance across both training and testing datasets. The confusion matrices revealed the best distribution of false positive and false negative predictions. For all performed metrics this model achieved score of 0.97 across all evaluation criteria on both training and testing datasets, which suggest strong model's capability for generalization. The ROC curves for this patch size show almost identical patterns with a steep rise at low false positive rates, achieving an AUC value of 0.99 on the training dataset and a perfect score of 1.00 on the testing dataset. The model for this patch size appears to provide sufficient spatial context to accurately interpret the spectral and temporal patterns associated with burned areas.

For the largest patch size of 64 x 64, model showed a tendency for more conservative predictions, as reflected in perfect precision (1.00) but lower recall (0.86) for burned areas on the test dataset. This indicates that larger patches may provide more stable predictions of larger burned areas, but may reduce the ability to detect smaller burned areas and complex burn patterns within them. The slightly flattened ROC curves, with an AUC value of 0.98 for both datasets (training and testing) confirm these findings, indicating some loss of model precision in recognizing fine details due to the large sample size.

Additional validation of these findings was conducted through qualitative analysis at two different locations: Site 1, which contains the large fire used during model training, and Site 2, which includes smaller fires that were not part of the training set. This comparison made it possible to assess the generalization ability of the model through different patch sizes in real conditions. At Site 1, for all patch sizes model successfully detected a large burned

area, with the 32 x 32 patch size showing the best balance between preserving boundary detail and mapping continuity. However, testing at Site 2 revealed limitations of the model in detecting very small burned areas, where the smallest fire was not successfully detected at any patch size. This confirms earlier quantitative findings on the effect of patch size on model's detection ability.

8.4.2. Limitations

There are several limitations of the current study:

- The model's ability to detect very small burned areas - although 10 x 10 patches were used, which represents the lowest spatial resolution available in Sentinel-2 imagery, this was not sufficient to detect the smallest fire at Site 2. This limitation could be particularly significant in areas where small fires are frequent, so the model in its current configuration would not be adequate for application.
- The model was trained and tested in one specific geographic area, which may limit its ability to estimate burned areas in different landscape or fire patterns.
- The time resolution of Sentinel-2 data (5 days), although sufficient for the detection of burned areas, may result in missed changes when detection is required in a shorter period of time. In addition, the presence of clouds on Sentinel-2 images can interfere with continuous monitoring and increase the time until the next usable image.

8.4.3. Future Research Directions

The current study opens up several promising directions for future research:

- Research into a hybrid approach combining different patch sizes could overcome current limitations in detecting fires of different sizes and shapes. In particular, an adaptive mechanism could be developed that would automatically select the optimal patch size based on the characteristics of the analyzed area.
- Different data sources could improve the model's detection capabilities, especially when the data comes with different spatial and temporal resolutions. In addition, this could reduce the limitations caused by cloudiness — in the event that Sentinel-2 captures a cloudy image, another data source could provide a cloud-free image.
- Model validation in different geographical areas with diverse types of vegetation and terrain characteristics. For example, the application of transfer learning techniques could facilitate the adaptation of the model to new areas with minimal additional training.

8.5. Conclusion

In this chapter, a 4D-CNN model for the burned area detection by using time series of Sentinel-2 data is presented. The research focuses on the 4D-CNN model ability to detect differences in spectral ratios between pre and post fire imagery, demonstrating effective capture of spatial, spectral and temporal patterns.

A detailed analysis was performed with three different patch sizes for the study area located in Split-Dalmatia County, which resulted in high model performance, with an overall accuracy greater than 0.93 for all tested sizes. For a patch size of 32 x 32 model provided an optimal balance between detail detection precision and prediction stability, achieving an accuracy of 0.97 on the test dataset. Smaller patches (10 x 10) were effective in detecting small details but were more sensitive to noise. Larger patches (64 x 64) provided more stable predictions for large burned areas but reduced the ability to detect smaller areas. Qualitative analysis at two different locations confirmed the model's ability to successfully detect large burned area, but indicated limitations in detecting very small burned areas.

The proposed approach has certain limitations, but also the potential for further improvements that will be useful for future research and practical application in fire management. These findings may help in the development of more effective fire monitoring and management systems, especially in Mediterranean regions similar to the study area. More precise burned area mapping will help fire analyst to better estimate consequences of fire.

9. A COMPARATIVE ANALYSIS OF CNN MODELS FOR SEMANTIC SEGMENTATION

This chapter systematically investigates and compares multiple CNN architectures (1D-CNN, 2D-CNN, 3D-CNN, and 4D-CNN) for land cover *semantic segmentation* utilizing Sentinel-2 satellite imagery. Segmentation is performed on the top-level classes of CORINE Land Cover (CLC) data. The research evaluates how varying CNN architectures affects segmentation accuracy and prediction quality, providing both quantitative metrics and qualitative visual assessments of the results.

The chapter structure is as follows. In the Introduction section the context for CNN comparison is provided; in Materials and Methods, the dataset and CNN architectures are described. The Results section presents experimental outcomes and Discussion analyzes performance variations. Finally, the Conclusion, summarizes findings and future directions.

9.1. Introduction

Land cover classification and mapping are of great importance for environmental monitoring, urban planning and resource management. Landscapes are continuously changing due to human activities, natural disasters, and changing weather conditions, so making timely information is crucial for decision-makers and researchers. Traditional classification methods are time-consuming and often subjective, which justifies the need for automated and reliable classification techniques [100].

Convolutional neural networks as a part of deep learning have shown exceptional efficiency in processing and analyzing remote sensing data [97]. The availability of high-quality satellite imagery enables the development of different CNN models that apply classification, regression, or segmentation methods depending on the observed problem. Specifically, Sentinel-2 imagery combined with the CORINE land cover database provides an opportunity to develop and evaluate sophisticated segmentation approaches [58].

Various CNN architectures have demonstrated effective results in remote sensing applications, with each dimensional approach offering distinct advantages. The choice between 1D, 2D, 3D, or 4D CNN architectures significantly impacts both segmentation accuracy and computational efficiency - 1D-CNNs excel at spectral feature extraction, 2D-CNNs are effective for spatial features, and 3D-CNNs can extract joint spectral-spatial features [96]. More

details about each CNN can be found in Section 3.5.

This chapter compares the performance of 1D, 2D, 3D and 4D-CNN architectures for land cover segmentation using Sentinel-2 images and CORINE Land Cover data. A systematic comparison of individual models was made using a similar architecture and the same data set, which enables an objective insight into the selection of a suitable model for the observed land cover segmentation problem.

The main objectives of this chapter are:

- evaluate the performance differences between 1D, 2D, 3D and 4D-CNN architectures,
- analyze the impact of CNN architectures on segmentation accuracy,
- provide quantitative performance metrics and qualitative visual assessment of segmentation results.

The comparison aims to provide insights into the strengths and limitations of each architectural approach, helping researchers and practitioners make informed decisions when selecting CNN models for land cover segmentation tasks. The findings from this analysis will contribute to the growing body of knowledge on deep learning applications in remote sensing and guide future developments in automated land cover mapping.

9.2. Materials and Methods

9.2.1. Study Area

Figure 9.1 shows the study area located in Split-Dalmatia County in Croatia. This area was chosen due to the variety of land types, including forests, agricultural areas, settlements and the sea, thus ensuring the heterogeneity of the scene, which is essential for the application of the segmentation technique. The study area, marked by a red rectangle on the map, covers 1311×983 (width \times height) pixels at 10-meter spatial resolution.

9.2.2. Dataset Construction and Preprocessing

The datasets were constructed from Sentinel-2 images, which were labeled based on land use information obtained from the CORINE dataset. Sentinel-2 images were taken from the Copernicus Data Space Ecosystem (Copernicus Sentinel data [2023] [284]). For each month in 2023, one product with the smallest cloud coverage was selected, but for all products cloud coverage didn't exceed 10%. Downloaded images are Level-2A, i.e. they are atmospherically corrected. All images were downloaded in .jp2 format and resampled to a spatial resolution of 10 m.

The CORINE Land Cover dataset is prepared using the European Union's Copernicus Land Monitoring Service information [335]. The CLC product offers a pan-European land

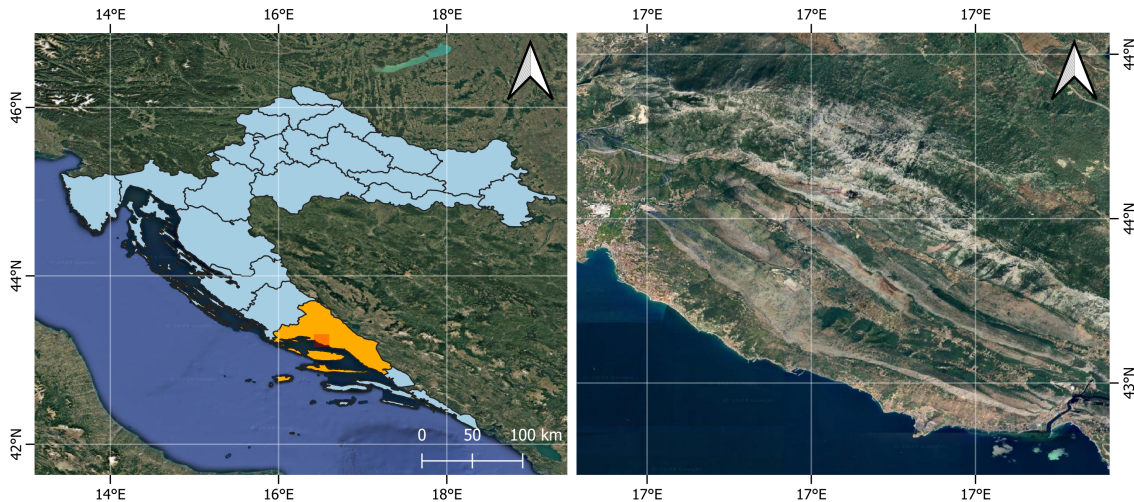


Figure 9.1. Study area in Split-Dalmatia County, Croatia (shown in Plate Carrée projection)

cover and land use inventory. There are a total of 44 thematic classes, from forest areas to individual vineyards. In this work, the dataset for the reference year 2018 was used, and it is available as vector and raster data with a spatial resolution of 100 m. In order to make the CLC dataset consistent with the Sentinel-2 dataset, they were resampled to a spatial resolution of 10 m.

Considering the large number of classes provided by the CLC dataset, it was reclassified into five main groups for simplicity. Each of these main groups is associated with the corresponding CLC classes, which are defined as: 'Urban areas', 'Agricultural areas', 'Woodland and forest', 'Wetlands and other natural areas' and 'Water-related areas' [336]. The labels were one-hot encoded, transforming each class label into a binary vector where the target class is represented by 1 and other classes by 0, which is essential for semantic segmentation tasks.

Datasets were created for each individual CNN, considering their degree. Therefore, for 1D-CNN, a dataset was created that represents the spectral values of pixels of all 12 bands extracted from 12 monthly scenes during the year 2023. The final dataset has a shape of (14 745 600, 12). The dataset for the 2D-CNN model was created to find spatial patterns. The final dataset has a shape of (691 200, 32, 32, 1), where the first value indicates the number of samples, the next two values indicate the size of the 32×32 patch, and the number 1 indicates that it was extracted for every single band. For the 3D-CNN model, a dataset with a shape of (57 600, 32, 32, 12) was created, which represents data cubes with a spatial dimension of 32×32 through 12 spectral bands. For the dataset intended for the 4D-CNN model, four scenes representing distinct seasons (winter, autumn, summer, and spring) were selected to capture the most significant temporal variations in land cover characteristics. The final dataset has a shape of (4, 983, 1311, 12), where the first argument indicates that there are 4 scenes, $983 \times$

1311 indicates the size of the scene (height x width), and 12 is the number of bands.

9.2.3. CNN Architecture

For all models (1D-CNN, 2D-CNN, 3D-CNN and 4D-CNN) similar architecture was used to enable objective comparison between the models. Figure 9.2 shows an architecture that follows a sequential pattern of convolutional blocks. While popular segmentation architectures like U-Net [291] employ skip connections and a symmetric encoder-decoder path, the proposed architecture has a simpler sequential structure adapted to the specific requirements of land cover segmentation.

The proposed architecture maintains a consistent structure in all model variants (1D, 2D, 3D and 4D), with adjustments to the specific requirements of the dimensionality of the input data. The input layer processes different forms of data:

- The 1D model uses the shape [12, 1] which represents 12 spectral values of one pixel,
- The 2D model uses the shape [32, 32, 1], which represents the spatial size of the patch of an individual channel,
- The 3D model uses the shape [32, 32, 12, 1] which represents spatial-spectral data,
- The 4D model uses the shape [4, 32, 32, 12] which represents seasonal spatial-spectral data.

The architecture's main structure contains three convolutional blocks with model-specific filter configurations. The first block implements 32 filters in the 1D model, 64 filters in the 2D model, and 32 filters in both 3D and 4D models. Moving to the second block, the 1D model uses 64 filters, the 2D model uses 128 filters, while both 3D and 4D models use 64 filters. In the final block, the 1D and 2D models employ 128 filters, whereas the 3D and 4D models use 64 filters. The number of filters gradually increased through the blocks, which increased the complexity of the model.

Each convolutional block for 3D and 4D architectures is followed by batch normalization and Max Pooling operations. Also, the kernel sizes are adjusted for each model based on their dimensionality using the 'same' padding. 1D and 2D architectures include a basic decoder path at the end of the convolutional blocks, while 3D and 4D architectures use Global Average Pooling followed by a Dense layer with 64 units and Dropout (0.3). The final layer in all architectures is a Dense layer using Softmax activation function for the classification of five different land cover types, outputting a one-hot encoded prediction vector for each pixel/patch.

The total number of trainable parameters varies across architectures: the 1D-CNN has 111 301 parameters, 2D-CNN has 926 021 parameters, 3D-CNN has 694 981 parameters, and 4D-CNN has 288 517 parameters, reflecting the different complexities of each approach.

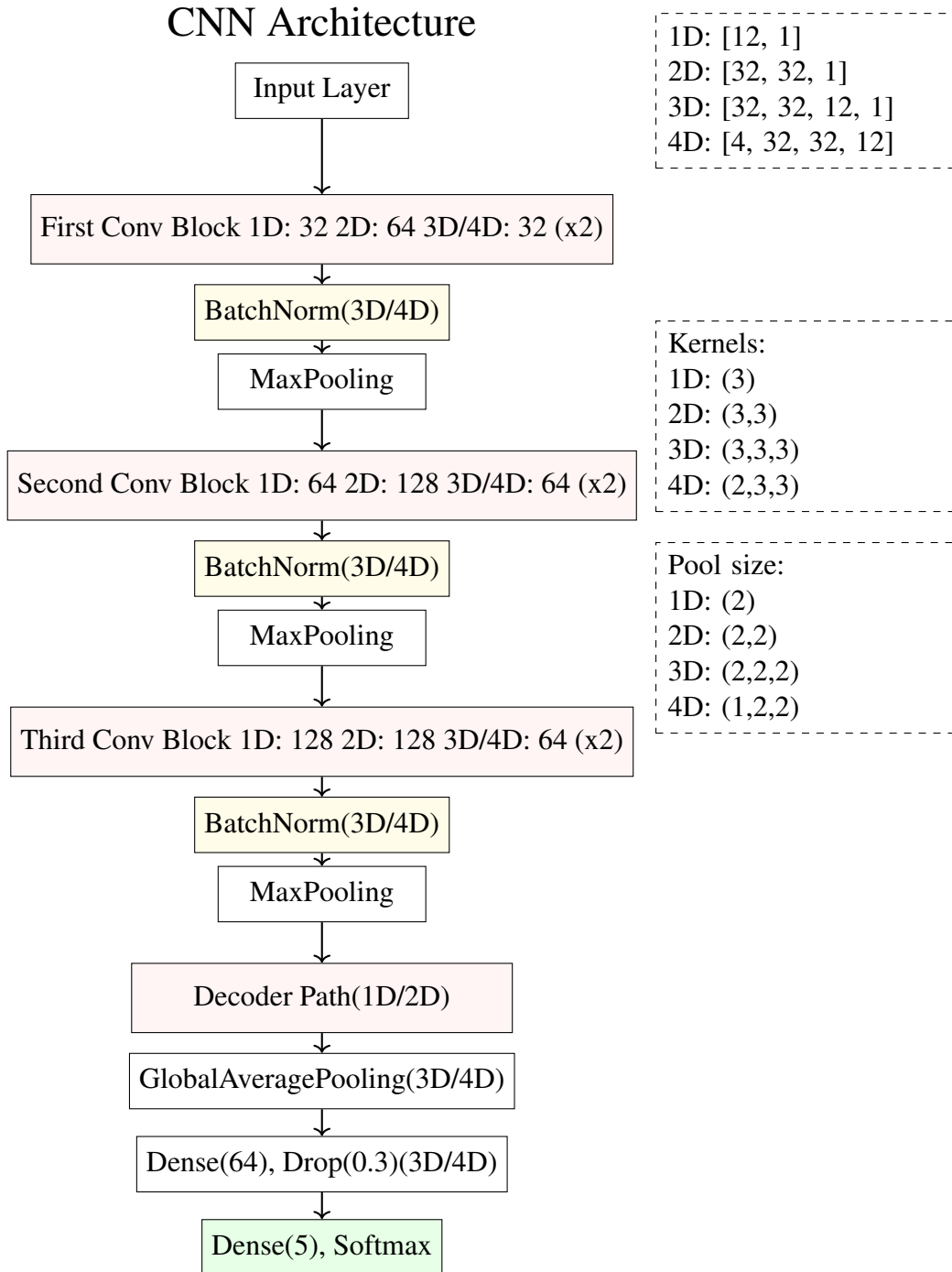


Figure 9.2. Unified CNN Architecture showing all model variants (1D/2D/3D/4D)

9.3. Results

This section presents the quantitative and qualitative results of different CNN architectures (1D-CNN, 2D-CNN, 3D-CNN, and 4D-CNN) for land cover segmentation using Sentinel-2 imagery. The performance of each model was evaluated using Intersection over Union (IoU) and Dice coefficient metrics and their mean values, which are described in Section 4.3.3. The comparison was made across five CORINE Land Cover classes: Urban Areas,

Agricultural Areas, Woodland and Forest, Wetlands and Natural Areas, and Water-related Areas. Additionally, a computational performance analysis was conducted to assess the training efficiency of each architecture. All models were implemented and tested on a system with an Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and 16 GB of RAM.

9.3.1. Performance Analysis of CNN Architectures

The results show the segmentation performance in different land cover classes using CNN architectures designed to process the spectral, spatial and temporal information of Sentinel-2 imagery. The performance metrics are presented in three tables: Table 9.1 shows the Intersection over Union (IoU) scores for each class and Table 9.2 presents the corresponding Dice coefficients.

Table 9.1. IoU Scores Comparison of Different CNN Architectures

Model	Per-class IoU				
	Urban Areas	Agricultural Areas	Woodland & Forest	Wetlands & Natural Areas	Water-related Areas
1D-CNN	0.241	0.303	0.647	0.220	0.921
2D-CNN	0.748	0.722	0.875	0.734	0.968
3D-CNN	0.926	0.855	0.942	0.810	0.999
4D-CNN	0.308	0.562	0.477	0.127	0.920

Table 9.2. Dice Coefficient Comparison of Different CNN Architectures

Model	Per-class Dice Coefficient				
	Urban Areas	Agricultural Areas	Woodland & Forest	Wetlands & Natural Areas	Water-related Areas
1D-CNN	0.389	0.465	0.786	0.360	0.959
2D-CNN	0.856	0.839	0.934	0.846	0.984
3D-CNN	0.962	0.922	0.970	0.895	0.999
4D-CNN	0.471	0.720	0.646	0.225	0.958

The 1D-CNN model, which was implemented to focus on the spectral information of the satellite image, showed modest overall performance with an achieved average IoU of 0.466 and Dice of 0.592. Nevertheless, the model achieved excellent results in predicting the 'Water' class (IoU 0.921 and Dice 0.959) compared to other CLC categories due unique spectral signature of water bodies. Water has specific absorption and reflection in different parts of the spectrum, especially in the green (B03) and short-wave infrared (SWIR) spectrum, which allows clear differentiation from other categories such as vegetation or built-up areas [337]. Furthermore, the model failed to classify complex types of land cover, which emphasizes the limitation of the model to perform classification based on spectral values, excluding the possibility of detecting the spatial relationship of individual classes.

The 2D-CNN model showed significant improvement with an average IoU value of 0.810 and a Dice score of 0.892. It has consistent performance in different land cover classes, where it achieved the best results in the classification of water bodies (IoU 0.968, Dice 0.984) and woodland/forest areas (IoU 0.875, Dice 0.934). The improved performance of the model indicates the importance of spatial context in land cover classification, although the model still shows possibility for improvement in the classification of urban (IoU 0.748, Dice 0.856), agricultural (IoU 0.722, Dice 0.839) and wetlands (IoU 0.734, Dice 0.846).

The 3D-CNN model achieved the best results in land cover classification with mean IoU and Dice scores of 0.906 and 0.950, respectively. The model outperformed other architectures in all land cover classes. This architecture achieved an IoU value below 0.9 only in the agricultural (IoU 0.855, Dice 0.922) and wetlands (IoU 0.810, Dice 0.895) classes, while a Dice value below 0.9 was recorded only for wetlands. This exceptional performance of the model highlights the advantage of simultaneous processing of spectral and spatial features of satellite images for accurate land cover segmentation.

The 4D-CNN model, similar to the 1D-CNN, did not achieve satisfactory results in land cover classification, despite its more complex architecture, which includes temporal information in addition to spatial and spectral components. The increased complexity of the 4D architecture, while theoretically more comprehensive, might lead to challenges in effectively learning from the limited number of temporal instances. The model might also be more sensitive to seasonal variations that are not necessarily indicative of land cover class differences, such as temporary changes in vegetation appearance that do not reflect the underlying land cover category. The model achieved a mean IoU value of 0.479 and a Dice value of 0.604. IoU values across individual classes ranged from 0.127 to 0.920, and Dice values from 0.225 to 0.958, indicating high variability in model performance.

The training process for all CNN architectures was configured with a maximum of 100 epochs and implemented with an early stopping mechanism (patience = 10) to prevent overfitting. Table 9.3 presents the computational requirements of each CNN architecture. While 1D-CNN was the only model to complete all 100 epochs, 4D-CNN required the fewest epochs (28) before early stopping was triggered. In terms of processing time, 2D-CNN showed the longest total training time of 43.78 hours, while 4D-CNN completed training in just 0.08 hours.

Table 9.3. Computational Performance of CNN Architectures

Model	Epochs Completed	Time per Epoch	Total Training Time (hours)	Early Stopping Triggered
1D-CNN	100	8m 30s	14.17	No
2D-CNN	71	37m	43.78	Yes
3D-CNN	43	54m 35s	39.12	Yes
4D-CNN	28	10s	0.08	Yes

9.3.2. Qualitative Analysis

Figure 9.3 shows a visual comparison between the original CLC map and the predicted maps based on 1D, 2D, 3D, and 4D-CNN models of the observed area. The Sentinel-2 scene from July 16, 2023 was used for the qualitative analysis, while four scenes from 2023 (March 23, June 21, September 9, and December 28) were used for the 4D-CNN model.

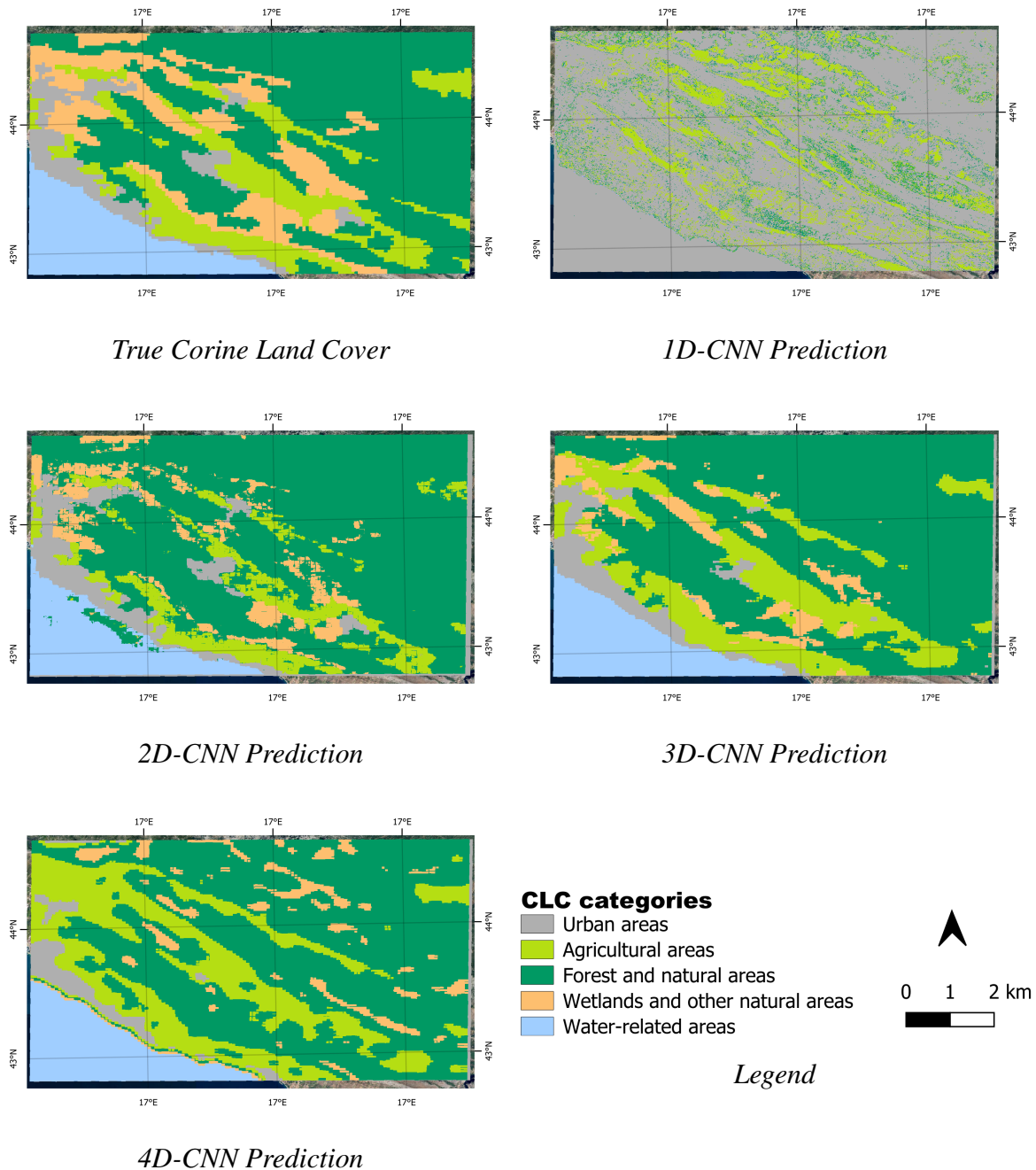


Figure 9.3. Comparison of True Land Cover Map with CNN Predictions

The map predicted by the 1D-CNN model shows significant noise and fragmentation in all classes, creating a 'salt and pepper' effect across the map. Such results confirm the limitations of the model that relies exclusively on the spectral information of the satellite

image, which is not sufficient for precise land cover segmentation.

The map predicted by the 2D-CNN model shows improved spatial coherence of classes compared to the map predicted by 1D-CNN model. However, the model shows uncertainty in boundary areas between different classes, especially where Agricultural and Wetland classes meet Urban or Forest classes. This is particularly visible in the central part of the map, where Agricultural areas are fragmented along the Forest borders.

The map predicted using the 3D-CNN model shows the closest similarity to the original CLC map. The model proved to be successful in recognizing spatial patterns and class boundaries, among which the Forest area and Water body classes stand out in particular. Also, transitions between different types of land cover are more natural and consistent with the reference data.

The map predicted using the 4D-CNN model shows the main features of the classes, but the finer details that can be seen in the original CLC map are lost, especially in areas with mixed land cover types. Despite being the most complex data set, containing spectral, spatial and temporal features, this model oversimplifies some land cover patterns. The temporal feature seems to introduce some uncertainty into class boundaries, and this is particularly visible in Agricultural and Wetland areas.

9.4. Discussion

A comprehensive comparison of the four CNN architectures used in this study for land cover segmentation is presented in Table 9.4. Experimental results showed significant variations in both quantitative and qualitative analyses.

Table 9.4. Comprehensive Performance Analysis of CNN Architectures

Model	Mean IoU/Dice	Best Class	Worst Class	Performance Consistency	Classes >0.9 IoU
1D-CNN	0.466/0.592	Water (0.921)	Wetlands (0.220)	Low	1
2D-CNN	0.810/0.892	Water (0.968)	Agricultural (0.722)	Medium	1
3D-CNN	0.906/0.950	Water (0.999)	Wetlands (0.810)	High	3
4D-CNN	0.479/0.604	Water (0.920)	Wetlands (0.127)	Low	1

The 1D-CNN model, although computationally efficient, showed limited performance with mean IoU/Dice scores of 0.466/0.592. While it performed well in Water class detection (IoU 0.921), it showed poor performance in other classes, especially Wetlands (IoU 0.220). Additionally, the predicted map exhibited a "salt and pepper" visual effect, suggesting that

spectral information alone is not sufficient for accurate land cover segmentation.

The 2D-CNN model showed significant improvement with mean IoU/Dice scores of 0.810/0.892. This suggests that the inclusion of spatial information led to better class coherence, which was also evident in the predicted map during qualitative analysis. The disadvantages of this model were the longer computational time (43.78 hours) and the lack of precise definition of class boundaries, especially in Agricultural areas (IoU 0.722).

The 3D-CNN model achieved the highest mean IoU/Dice values of 0.906/0.950, thus proving to be the most efficient architecture for land cover segmentation. The model demonstrated the advantage of combining spectral and spatial information. Despite the relatively long training time of the model (39.12 hours), the high segmentation accuracy justifies the computational complexity.

The 4D-CNN model, despite the most complex architecture, showed performance comparable to the 1D-CNN model (IoU/Dice: 0.479/0.604). Although the model achieved the shortest computation time, its low consistency in performance and poor class segmentation suggest that adding a time information may introduce unnecessary complexity without improving land cover segmentation.

The above results show that the most accurate land cover segmentation is achieved by combining spectral and spatial information using the 3D-CNN model. Adding temporal information (4D-CNN) or relying solely on spectral information (1D-CNN) did not improve model performance. It is important to note that, although the CLC map was used as reference data, it is a static representation of land cover that is updated every six years. In contrast, the proposed segmentation approach based on Sentinel-2 satellite images enables the monitoring of dynamic land cover changes with high temporal resolution, which is especially important for areas with rapid land cover changes or for monitoring seasonal changes.

9.5. Conclusion

This study evaluated the performance of different CNN architectures (1D-CNN, 2D-CNN, 3D-CNN, and 4D-CNN) for land cover segmentation using Sentinel-2 imagery. The 3D-CNN architecture outperformed other models and achieved mean IoU/Dice scores of 0.906/0.950. These results highlight the importance of effectively combining both spectral and spatial information. While the CLC map provided reliable reference data, its static nature limits monitoring of dynamic land cover changes. The proposed approach using Sentinel-2 imagery and CNN-based segmentation offers potential for more frequent and automated land cover monitoring. Future research could focus on optimizing spatial-spectral feature extraction while reducing computational costs.

10. CONCLUSION

The main focus of this doctoral thesis was to explore the capabilities of 1D, 2D, 3D, and 4D-CNN architectures in the analysis of multispectral satellite imagery, which represent the foundation of more complex architectures in deep learning. The research focused on extracting spectral, spatial, and temporal features from satellite images by applying appropriate degree of CNN architecture. To gain insight into the achievements in this field, the work first provides a detailed literature review that systematically presents the datasets used, machine learning techniques, application domains, and problem-specific metrics for different CNN architectures. The findings are synthesized in the form of an ontology that is proposed to serve as a foundation for decision-making, thus providing a structured framework for understanding the selection of an appropriate convolution degree for an Earth observation scenario.

Furthermore, the findings were evaluated through empirical research by selecting case-study applications. First, the application of 1D-CNN architecture was evaluated by predicting the Secchi disk depth parameter, considering only the spectral features of Sentinel-3 OLCI satellite data. The 1D-CNN architecture demonstrated strong performance, achieving an R^2 value of 0.89 on the test dataset, and outperformed other well-known algorithms such as C2RCC. The next phase of research focused on the implementation of 2D-CNN, which achieved 91% accuracy in road detection using spatial features from Sentinel-2 satellite imagery. The model also outperformed more complex architectures such as DeepLabV3, ResNet-50 and U-Net, demonstrating that the choice of architecture depends on the problem characteristics rather than just its complexity. Building upon previous architectures, the implemented 3D-CNN successfully predicted fire propagation potential with an accuracy of 96% on the test dataset, utilizing both spectral and spatial characteristics of Sentinel-2 satellite imagery. The model also outperformed and demonstrated high results, both quantitatively and qualitatively, when compared to 1D-CNN and 2D-CNN. Subsequently, a 4D-CNN model was implemented for burned area detection using Sentinel-2 satellite imagery time series, proving the model's effectiveness in processing spatial, spectral, and temporal features. The model was tested with three different window sizes in Split-Dalmatia County, where it achieved optimal results with a 32 x 32 pixel window, reaching 97% accuracy on the test dataset. Qualitative analysis confirmed the model's ability to successfully detect large burned areas, with some limitations in detecting smaller areas. Finally, a comparative anal-

ysis of all four convolutional neural networks was conducted on land cover segmentation to gain insight into their advantages and limitations. This analysis confirmed the hypothesis that 3D-CNN would achieve the best results (IoU/Dice scores of 0.906/0.950) due to its capability to simultaneously extract both spectral and spatial features from satellite images, which are important for correctly identifying different land cover classes.

Original scientific contributions of this dissertation are:

- Determining the impact of convolution degree on convolutional neural network results depending on the properties of problems being solved in remote sensing. The dissertation synthesizes results of a systematic literature review in the form of an ontology, providing a practical framework for determining the convolution degree that would enhance the performance of deep learning applications in remote sensing Earth observation tasks.
- New methods for predicting spectral, spatial, and temporal phenomena using convolutional neural networks of appropriate degree in remote sensing. The research was conducted through five different case studies, resulting in novel methods for predicting Earth observation parameters based on multispectral satellite imagery:
 - A 1D-CNN architecture for Secchi disk depth prediction using Sentinel-3 satellite imagery spectral features;
 - A 2D-CNN architecture for road detection using Sentinel-2 satellite imagery spatial features;
 - A 3D-CNN architecture for fire propagation potential prediction using Sentinel-2 satellite imagery spectral-spatial features;
 - A 4D-CNN architecture for burned area detection using Sentinel-2 satellite imagery spectral-spatial-temporal features;
 - A comparative analysis of 1D, 2D, 3D and 4D-CNN architectures for land cover segmentation using Sentinel-2 satellite imagery.
- A novel approach for determining the type of remote sensing problem in relation to the applicable degree of convolution. The research provides a blueprint for describing remote sensing problems by outlining the application domain, machine learning task, and spectral, spatial, spectral-spatial, and temporal dimensions. The problem description further contributes to the selection of the convolution degree based on the specific requirements of remote sensing application.

Additional useful result of this research, which do not represent scientific contribution is:

- Practical knowledge and a set of recommendations for selecting the degree of convolution based on the properties of the problem being solved by CNN.

Findings presented in this thesis hold significant value for both scientists and practitioners seeking a systematic framework for selecting appropriate approaches in remote sensing applications. The combination of systematic literature review and empirical evaluation of CNN-based deep learning applications provides scientifically valid results for informed decision-making. The case studies encompass diverse aspects of remote sensing, including machine learning tasks (regression, classification and segmentation), application domains (sea water quality monitoring, urban planning, land cover mapping, and fire management) and remote sensing tasks (parameter prediction, object detection, and change detection).

Future research will aim to develop a software framework based on the findings from this doctoral thesis, which will enable the prediction of phenomena in remote sensing using 1D, 2D, 3D and 4D convolutional neural networks with an expert system for selecting the appropriate degree of convolution.

In the context of future work, several key directions can be addressed:

- Integration of different data sources such as Sentinel-1 satellite radar data or aerial imagery, which could improve the robustness and accuracy of the models, reducing the limitations of individual data sources.
- Development of more efficient and optimized architectures by finding optimal hyperparameters or combining different degrees of convolutional layers within a single model.
- Better generalization of models across different geographical regions using transfer learning, which could particularly benefit areas with limited datasets.
- Real or near-real time systems, particularly important for applications where response time is crucial.
- Interpretability and explainability of model decisions, which would result in more reliable models, especially important in critical situations such as natural disasters.

In conclusion, this thesis provides a comprehensive and systematic investigation of convolution degree's role in CNN-based deep learning for remote sensing applications. It highlights the critical influence of convolution degree on application performance, emphasizing its importance in enhancing CNNs' capability to extract meaningful information from remote sensing data. This work advances the understanding of convolution mechanics in remote sensing contexts and their contribution to optimizing deep learning models for practical, high-impact applications.

BIBLIOGRAPHY

- [1] Space Force, Space-track.org public satellite catalog, 2024, u.S. Space Force Space Surveillance Network (SSN) satellite tracking database, accessed 18 November 2024.
- [2] F. Del Canto Viterale, Transitioning to a new space age in the 21st century: A systemic-level approach, *Systems*, 11, 5, 2023.
- [3] SpaceX, Starlink, 2024, accessed: 2024-11-20.
- [4] Planet Labs PBC, Planet - daily satellite imagery and insights, 2024, accessed: 2024-11-20.
- [5] J. McDowell, Jonathan's space pages: Starlink statistics, 2024, accessed: 2024-11-20.
- [6] W. Hazeleger, J. Aerts, P. Bauer, M. Bierkens, G. Camps-Valls, M. Dekker, F. Doblas-Reyes, V. Eyring, C. Finkenauer, A. Grundner et al., Digital twins of the earth with and for humans, *Communications Earth & Environment*, 5, 1, 463, 2024.
- [7] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen and Y. Zhu, Big data for remote sensing: Challenges and opportunities, *Proceedings of the IEEE*, 104, 11, 2207–2219, 2016.
- [8] B. Zhao, M. Liu, J. Wu, X. Liu, M. Liu and L. Wu, Parallel computing for obtaining regional scale rice growth conditions based on wofost and satellite images, *IEEE Access*, 8, 223675–223685, 2020.
- [9] L. Zhang and L. Zhang, Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities, *IEEE Geoscience and Remote Sensing Magazine*, 10, 2, 270–294, 2022.
- [10] S. Azam, S. Montaha, K. U. Fahim, A. R. H. Rafid, M. S. H. Mukta and M. Jonkman, Using feature maps to unpack the cnn 'black box' theory with two medical datasets of different modality, *Intelligent Systems with Applications*, 18, 200233, 2023.
- [11] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data (2nd ed.)*, CRC Press, 2016.
- [12] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing, Chapter 2, 4, 8, 10, 11 and 12*, pp. 31–55, 285–286, 339–356, Guilford Press, 2011.
- [13] E. Chuvieco and A. Huete, *Fundamentals of Satellite Remote Sensing, Chapter 2 and 3*, pp. 21–112, Taylor and Francis, 2009.

- [14] Casimiro, Maria and Ferreira, Luís and Leal, João and Pereira, and Monteiro, Bernardo, Ionizing Radiation for Preparation and Functionalization of Membranes and Their Biomedical and Environmental Applications, *Membranes*, 9, 163, 12 2019.
- [15] T. Lillesand, R. W. Kiefer and J. Chipman, *Remote sensing and image interpretation, Chapter 1: Concepts and Foundations of Remote Sensing*, pp. 1–59, John Wiley & Sons, 2015.
- [16] M. E. Klein, B. J. Aalderink, R. Padoan, G. De Bruin and T. A. Steemers, Quantitative hyperspectral reflectance imaging, *Sensors*, 8, 9, 5576–5618, 2008.
- [17] J. E. Estes and D. S. Simonett, Fundamentals of image interpretation, *Manual of remote sensing*, 1, 869–1076, 1975.
- [18] S. Liang and J. Wang, Advanced Remote Sensing (Second Edition) - Chapter 1, 2 and 5, *Advanced Remote Sensing (Second Edition)*, 1–57, Academic, 2020.
- [19] L. Zhu, J. Suomalainen, J. Liu, J. Hyypä, H. Kaartinen, H. Haggren et al., A review: Remote sensing sensors, *Multi-purposeful application of geospatial data*, 19–42, 2018.
- [20] L. Fan, *Automated Remote Sensing Image Interpretation with Limited Labeled Training Data*, Phd thesis, University of Waterloo, 2015, available at <https://uwaterloo.ca/items/1c801bab-4301-4c83-b1c4-8f139906d394>.
- [21] M. I. Skolnik, Introduction to radar, *Radar handbook*, 2, 21, 1962.
- [22] L. Cutrona, Synthetic aperture radar, *Radar handbook*, 2, 2333–2346, 1990.
- [23] R. Harrap and M. Lato, An overview of lidar: collection to application, *NGI publication*, 2, 1–9, 2010.
- [24] K. Tempfli, G. Huurneman, W. Bakker, L. L. Janssen, W. Feringa, A. Gieske, K. Grabmaier, C. Hecker, J. Horn, N. Kerle et al., *Principles of remote sensing: an introductory textbook*, International Institute for Geo-Information Science and Earth Observation, 2009.
- [25] M. J. Badzmierowski, D. S. McCall and G. Evanylo, Using Hyperspectral and Multi-spectral Indices to Detect Water Stress for an Urban Turfgrass System, *Agronomy*, 9, 8, 2019.
- [26] NASA Official, Passive Sensors, Online Available: <https://earthdata.nasa.gov/learn/remote-sensors/passive-sensors>, [Accessed on 13 June 2024.].
- [27] J. M. Meyer, R. F. Kokaly and E. Holley, Hyperspectral remote sensing of white mica: A review of imaging and point-based spectrometer studies for mineral resources, with spectrometer design considerations, *Remote Sensing of Environment*, 275, 113000, 2022.
- [28] C. Li and H. Xiong, A geometric and radiometric simultaneous correction model (grscm) framework for high-accuracy remotely sensed image preprocessing, *Photogrammetric Engineering & Remote Sensing*, 83, 9, 621–632, 2017.

- [29] a. P. L. c. Sinergise Solutions d.o.o., Eo browser, <https://apps.sentinel-hub.com/eo-browser/>, 2024, [Accessed on 20. June 2024.].
- [30] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm and F. Gascon, Sen2Cor for Sentinel-2, 3, 10 2017.
- [31] K. Tarrio, X. Tang, J. G. Masek, M. Claverie, J. Ju, S. Qiu, Z. Zhu and C. E. Woodcock, Comparison of cloud detection algorithms for Sentinel-2 imagery, *Science of Remote Sensing*, 2, 100010, 2020.
- [32] C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher and A. Ruescas, Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters, *Living Planet Symposium*, 740, 54, 2016.
- [33] R. T. Wilson, Py6S: A Python interface to the 6S radiative transfer model., *Comput. Geosci.*, 51, 2, 166–171, 2013.
- [34] T. Cooley, G. P. Anderson, G. W. Felde, M. L. Hoke, A. J. Ratkowski, J. H. Chetwynd, J. A. Gardner, S. M. Adler-Golden, M. W. Matthew, A. Berk et al., FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation, *IEEE international geoscience and remote sensing symposium*, 3, 1414–1418, IEEE, 2002.
- [35] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes and J. van den Bosch, MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code, *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–4, 2014.
- [36] K. Bajpai and R. Soni, Analysis of image enhancement techniques used in remote sensing satellite imagery, *International Journal of Computer Applications*, 169, 10, 1–11, 2017.
- [37] T. Taxt and A. Schistad Solberg, Data fusion in remote sensing, *Data Analysis in Astronomy, Proceedings of the Fifth Workshop. Ettore Majorana Centre for Scientific Culture, Erice, Italy*, 269–280, 1998.
- [38] C. Pohl and J. L. V. Genderen, Review article multisensor image fusion in remote sensing: Concepts, methods and applications, *International Journal of Remote Sensing*, 19, 5, 823–854, 1998.
- [39] Y. Huang, Z.-x. Chen, Y. Tao, X.-z. Huang and X.-f. Gu, Agricultural remote sensing big data: Management and applications, *Journal of Integrative Agriculture*, 17, 9, 1915–1931, 2018.
- [40] J. M. Piwowar, Getting your imagery at the right level, *Cartouche*, 41, 2001.
- [41] W. J. Larson and J. R. Wertz, Space Mission Analysis and Design: Vol. 8, *Space Technology Library (3rd ed.)*. New York: Springer, 47–72, 1999.
- [42] European Space Agency - ESA, The Sentinel missions, Online Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions, [Accessed on 20. June 2024.].

- [43] European Space Agency, Sentinel-1c fuelled ahead of liftoff, November 2024, accessed 18 November 2024.
- [44] D. Radočaj, J. Obhodaš, M. Jurišić and M. Gašparović, Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review, *Land*, 9, 11, 2020.
- [45] G. Kaplan and U. Avdan, Object-based water body extraction model using sentinel-2 satellite imagery, *European Journal of Remote Sensing*, 50, 1, 137–143, 2017.
- [46] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama and M. Ranagalage, Sentinel-2 data for land cover/use mapping: A review, *Remote Sensing*, 12, 14, 2020.
- [47] C. Donlon, B. Berruti, S. Mecklenberg, J. Nieke, H. Rebhan, U. Klein, A. Buongiorno, C. Mavrocordatos, J. Frerick, B. Seitz, P. Goryl, P. Féménias, J. Stroede and R. Sciarra, The sentinel-3 mission: Overview and status, *2012 IEEE International Geoscience and Remote Sensing Symposium*, 1711–1714, 2012.
- [48] T. D. Acharya and I. Yang, Exploring landsat 8, *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, 4, 4, 4–10, 2015.
- [49] C. Parkinson, Aqua: an Earth-Observing Satellite mission to examine water and other climate variables, *IEEE Transactions on Geoscience and Remote Sensing*, 41, 2, 173–183, 2003.
- [50] NASA Earth Science, Aqua earth-observing satellite mission, 2023, [Accessed on 18 November 2024.].
- [51] NASA Earth Observatory, Terra mission overview and status, 2023, [Accessed on 18 November 2024.].
- [52] L. Guanter, M. D. C. González-Sanpedro and J. Moreno, A method for the atmospheric correction of envisat/meris data over land targets, *International Journal of Remote Sensing*, 28, 3-4, 709–728, 2007.
- [53] W. Dierckx, S. Sterckx, I. Benhadj, S. Livens, G. Duhoux, T. V. Achteren, M. Francois, K. Mellab and G. Saint, PROBA-V mission for global vegetation monitoring: standard products and image quality, *International Journal of Remote Sensing*, 35, 7, 2589–2614, 2014.
- [54] P. Mhangara, W. Mapurisa and N. Mudau, Comparison of image fusion techniques using satellite pour l’observation de la terre (spot) 6 satellite imagery, *Applied Sciences*, 10, 5, 2020.
- [55] European Space Agency, WorldView-2, Online Available: <https://earth.esa.int/eogateway/missions/worldview-2>, [Accessed on 22 June 2024.].
- [56] European Space Agency, WorldView-3, Online Available: <https://earth.esa.int/eogateway/missions/worldview-3>, [Accessed on 22 June 2024.].
- [57] J. D. Kelleher, *Deep Learning, Chapter 1*, pp. 1–38, MIT press, 2019.

- [58] L. Zhang, L. Zhang and B. Du, Deep learning for remote sensing data: A technical tutorial on the state of the art, *IEEE Geoscience and Remote Sensing Magazine*, 4, 2, 22–40, 2016.
- [59] B. Rasti, Y. Chang, E. Dalsasso, L. Denis and P. Ghamisi, Image restoration for remote sensing: Overview and toolbox, *IEEE Geoscience and Remote Sensing Magazine*, 10, 2, 201–230, 2022.
- [60] W. Huang, L. Xiao, Z. Wei, H. Liu and S. Tang, A new pan-sharpening method with deep neural networks, *IEEE Geoscience and Remote Sensing Letters*, 12, 5, 1037–1041, 2015.
- [61] N. Zerrouki and D. Bouchaffra, Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification?, *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 864–869, IEEE, 2014.
- [62] J. Lv, Q. Shen, M. Lv, Y. Li, L. Shi and P. Zhang, Deep learning-based semantic segmentation of remote sensing images: a review, *Frontiers in Ecology and Evolution*, 11, 1201125, 2023.
- [63] Y. Gao and J. F. Mas, A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions, *Online journal of earth sciences*, 2, 1, 27–35, 2008.
- [64] S. Jiang, H. Huang, J. Yang, X. Zhang and S. Wang, Innovative research on small object detection and recognition in remote sensing images using yolov5, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 77–83, 2024.
- [65] K. Dyson, A. P. Nicolau, D. Saah and N. Clinton, *Interpreting an Image: Regression*, 149–168, Springer International Publishing, Cham, 2024.
- [66] P. Ye, G. Liu and Y. Huang, Geographic scene understanding of high-spatial-resolution remote sensing images: Methodological trends and current challenges, *Applied Sciences*, 12, 12, 2022.
- [67] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci and M. Parmar, A review of convolutional neural networks in computer vision, *Artificial Intelligence Review*, 57, 4, 99, 2024.
- [68] S. Albawi, T. A. Mohammed and S. Al-Zawi, Understanding of a convolutional neural network, *2017 International Conference on Engineering and Technology (ICET)*, 1–6, 2017.
- [69] X. Wang, H. Ren and A. Wang, Smish: A novel activation function for deep learning methods, *Electronics*, 11, 4, 2022.
- [70] H. Izadkhah, Chapter 8 - medical image processing: an insight to convolutional neural networks, H. Izadkhah, editor, *Deep Learning in Bioinformatics*, 175–213, Academic Press, 2022.
- [71] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning: Convolutional Networks - Chapter 9*, pp. 326–366, MIT Press, 2016, <http://www.deeplearningbook.org>.

- [72] K. Phil, Matlab deep learning with machine learning, neural networks and artificial intelligence, Chapter 6: Convolutional Neural Network, *Apress, New York*, 2017.
- [73] A. Ajit, K. Acharya and A. Samanta, A review of convolutional neural networks, *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 1–5, 2020.
- [74] R. Yamashita, M. Nishio, R. K. G. Do and K. Togashi, Convolutional neural networks: an overview and application in radiology, *Insights into imaging*, 9, 611–629, 2018.
- [75] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti and D. De, *Fundamental Concepts of Convolutional Neural Network*, 519–567, Springer International Publishing, Cham, 2020.
- [76] C. Shorten, T. M. Khoshgoftaar and B. Furht, Text data augmentation for deep learning, *Journal of big Data*, 8, 1, 101, 2021.
- [77] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, *Journal of big Data*, 8, 1, 1–74, 2021.
- [78] S. Bera and V. Shrivastava, Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification, *International Journal of Remote Sensing*, 41, 2664–2683, 04 2020.
- [79] W. Hu, Y. Huang, L. Wei, F. Zhang and H. Li, Deep convolutional neural networks for hyperspectral image classification, *Journal of Sensors*, 2015, 1–12, 2015.
- [80] P. Dou, H. Shen, Z. Li and X. Guan, Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system, *International Journal of Applied Earth Observation and Geoinformation*, 103, 102477, 2021.
- [81] S. Ojaghi, Y. Bouroubi, S. Foucher, M. Bergeron and C. Seynat, Deep learning-based emulation of radiative transfer models for top-of-atmosphere brdf modelling using sentinel-3 olci, *Remote Sensing*, 15, 3, 835, 2023.
- [82] S. Park and N.-W. Park, Effects of class purity of training patch on classification performance of crop classification with convolutional neural network, *Applied Sciences*, 10, 11, 3773, 2020.
- [83] M.-G. Park and N.-W. Park, Application of deep learning algorithms considering spatio-temporal features for crop classification, *40th Asian Conference on Remote Sensing, ACRS 2019: Progress of Remote Sensing Technology for Smart Future*, 2020.
- [84] X. Zhou, W. Zhou, F. Li, Z. Shao and X. Fu, Vegetation type classification based on 3d convolutional neural network model: A case study of baishuijiang national nature reserve, *Forests*, 13, 6, 2022.
- [85] S. Ji, C. Zhang, A. Xu, Y. Shi and Y. Duan, 3d convolutional neural networks for crop classification with multi-temporal remote sensing images, *Remote Sensing*, 10, 1, 2018.

- [86] M. Giannopoulos, G. Tsagkatakis and P. Tsakalides, 4d u-nets for multi-temporal remote sensing data classification, *Remote Sensing*, 14, 3, 2022.
- [87] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattemberg, M. Wicke, Y. Yu and X. Zheng, Tensorflow: Large-scale machine learning on heterogeneous systems, <https://www.tensorflow.org/>, 2015.
- [88] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems*, 32, 2019.
- [89] M. Giannopoulos, G. Tsagkatakis and P. Tsakalides, 4d convolutional neural networks for multi-spectral and multi-temporal remote sensing data classification, *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1541–1545, 2022.
- [90] A. Ivanda, L. Šerić and M. Braović, Exploring the applications of convolutional neural networks in analyzing multispectral satellite imagery: A systematic review, *Big Data Mining and Analytics*, 2024, in press.
- [91] S. Borzov and O. Potaturkin, Spectral-spatial methods for hyperspectral image classification. review, *Optoelectronics, instrumentation and data processing*, 54, 582–599, 2018.
- [92] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu and S. Yu, A survey: Deep learning for hyperspectral image classification with few labeled samples, *Neurocomputing*, 448, 179–204, 2021.
- [93] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid and A. Abbas, Modern trends in hyperspectral image analysis: A review, *IEEE Access*, 6, 14118–14129, 2018.
- [94] X. Yuan, J. Shi and L. Gu, A review of deep learning methods for semantic segmentation of remote sensing imagery, *Expert Systems with Applications*, 169, 114417, 2021.
- [95] S. Bera, V. K. Shrivastava and S. C. Satapathy, Advances in hyperspectral image classification based on convolutional neural networks: A review, *CMES-COMPUTER MODELING IN ENGINEERING & SCIENCES*, 133, 2, 219–250, 2022.
- [96] Y. Li, H. Zhang, X. Xue, Y. Jiang and Q. Shen, Deep learning for remote sensing image classification: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8, 6, e1264, 2018.
- [97] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin and B. A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177, 2019.

- [98] T. Kattenborn, J. Leitloff, F. Schiefer and S. Hinz, Review on convolutional neural networks (cnn) in vegetation remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24–49, 2021.
- [99] J. E. Ball, D. T. Anderson and C. S. Chan, Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community, *Journal of applied remote sensing*, 11, 4, 042609–042609, 2017.
- [100] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu and F. Fraundorfer, Deep learning in remote sensing: A comprehensive review and list of resources, *IEEE geoscience and remote sensing magazine*, 5, 4, 8–36, 2017.
- [101] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan et al., The prisma 2020 statement: an updated guideline for reporting systematic reviews, *International journal of surgery*, 88, 105906, 2021.
- [102] S. M. M. Nejad, D. Abbasi-Moghadam, A. Sharifi, N. Farmonov, K. Amankulova and M. Lászl, Multispectral crop yield prediction using 3d-convolutional neural networks and attention convolutional lstm approaches, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 254–266, 2023.
- [103] E. O. Box and K. Fujiwara, Vegetation types and their broad-scale distribution, *Vegetation ecology*, 455–485, 2013.
- [104] M. Weiss, F. Jacob and G. Duveiller, Remote sensing for agricultural applications: A meta-review, *Remote Sensing of Environment*, 236, 111402, 2020.
- [105] D. Roy, M. Wulder, T. Loveland, W. C.E., R. Allen, M. Anderson, D. Helder, J. Irons, D. Johnson, R. Kennedy, T. Scambos, C. Schaaf, J. Schott, Y. Sheng, E. Vermote, A. Belward, R. Bindschadler, W. Cohen, F. Gao, J. Hipple, P. Hostert, J. Huntington, C. Justice, A. Kilic, V. Kovalsky, Z. Lee, L. Lymburner, J. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. Wynne and Z. Zhu, Landsat-8: Science and product vision for terrestrial global change research, *Remote Sensing of Environment*, 145, 154–172, 2014.
- [106] A. Savtchenko, D. Ouzounov, S. Ahmad, J. Acker, G. Leptoukh, J. Koziana and D. Nickless, Terra and aqua modis products available from nasa ges daac, *Advances in Space Research*, 34, 4, 710–714, 2004, trace Constituents in the Troposphere and Lower Stratosphere.
- [107] Y. Chen, L. Tang, Z. Kan, A. Latif, X. Yang, M. Bilal and Q. Li, Cloud and cloud shadow detection based on multiscale 3d-cnn for high resolution multispectral imagery, *IEEE Access*, 8, 16505–16516, 2020.
- [108] Q. Zhang, Q. Yuan, Z. Li, F. Sun and L. Zhang, Combined deep prior with low-rank tensor svd for thick cloud removal in multitemporal images, *ISPRS Journal of Photogrammetry and Remote Sensing*, 177, 161–173, 2021.
- [109] H. Jiang and N. Lu, Multi-scale residual convolutional neural network for haze removal of remote sensing images, *Remote Sensing*, 10, 6, 945, 2018.

- [110] F. Kong, K. Hu, Y. Li, D. Li and S. Zhao, Spectral–spatial feature partitioned extraction based on cnn for multispectral image compression, *Remote Sensing*, 13, 1, 9, 2020.
- [111] Z. Zhu, Y. Tao and X. Luo, Hcnnet: A hybrid convolutional neural network for spatiotemporal image fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–16, 2022.
- [112] X. Zhao, Y. Ma, Y. Xiao, J. Liu, J. Ding, X. Ye and R. Liu, Atmospheric correction algorithm based on deep learning with spatial-spectral feature constraints for broadband optical satellites: Examples from the hy-1c coastal zone imager, *ISPRS Journal of Photogrammetry and Remote Sensing*, 205, 147–162, 2023.
- [113] A. B. Molini, D. Valsesia, G. Fracastoro and E. Magli, Deep learning for super-resolution of unregistered multi-temporal satellite images, *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5, IEEE, 2019.
- [114] F. Salvetti, V. Mazzia, A. Khaliq and M. Chiaberge, Multi-image super resolution of remotely sensed images using residual attention deep neural networks, *Remote Sensing*, 12, 14, 2207, 2020.
- [115] F. Dorr, Satellite image multi-frame super resolution using 3d wide-activation neural networks, *Remote Sensing*, 12, 22, 3812, 2020.
- [116] M. R. Ibrahim, R. Benavente, F. Lumbreras and D. Ponsa, 3drrdb: Super resolution of multiple remote sensing images using 3d residual in residual dense blocks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 323–332, 2022.
- [117] N. Li, R. Wang, H. Zhao and W. Wei, An improved feature extraction method based on context features for multi-spectral remote sensing imagery, *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 1–4, IEEE, 2019.
- [118] S. He, R. Zhou, S. Li, S. Jiang and W. Jiang, Disparity estimation of high-resolution remote sensing images with dual-scale matching network, *Remote Sensing*, 13, 24, 5050, 2021.
- [119] T. Igeta and A. Iwasaki, An unsupervised network for stereo matching of very high resolution satellite imagery, *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, 971–974, IEEE, 2022.
- [120] L. Wang, X. Wang, Z. Zhao, Y. Wu, J. Xu, H. Zhang, J. Yu, Q. Sun and Y. Bai, Multi-factor status prediction by 4d fractal cnn based on remote sensing images, *Fractals*, 30, 02, 2240101, 2022.
- [121] A. D. Gordon, *Classification*, chap. Chapter 1: Introduction, CRC Press, 1999.
- [122] J. Al-Doski, S. B. Mansorl and H. Z. M. Shafri, Image classification in remote sensing, *Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia*, 3, 10, 2013.

- [123] M. Hossin and M. N. Sulaiman, A review on evaluation metrics for data classification evaluations, *International journal of data mining & knowledge management process*, 5, 2, 1, 2015.
- [124] N. Japkowicz and M. Shah, *Evaluating learning algorithms: a classification perspective*, Cambridge University Press, 2011.
- [125] C. Ter Braak and C. Looman, Regression, *Data analysis in community and landscape ecology*, 29–77, Cambridge University Press, 1995.
- [126] A. V. Tatachar, Comparative assessment of regression models based on model evaluation metrics, *International Research Journal of Engineering and Technology (IRJET)*, 8, 09, 2395–0056, 2021.
- [127] D. Kaur and Y. Kaur, Various image segmentation techniques: a review, *International Journal of Computer Science and Mobile Computing*, 3, 5, 809–814, 2014.
- [128] V. Dey, Y. Zhang and M. Zhong, *A review on image segmentation techniques with remote sensing perspective*, 38, na Vienna, Austria, 2010.
- [129] M. Rejaur Rahman and S. Saha, Multi-resolution segmentation for object-based classification and accuracy assessment of land use/land cover classification using remotely sensed data, *Journal of the Indian Society of Remote Sensing*, 36, 189–201, 2008.
- [130] L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Li and J. Marcato, The segment anything model (sam) for remote sensing applications: From zero to one shot, *International Journal of Applied Earth Observation and Geoinformation*, 124, 103540, 2023.
- [131] S. Jeong, J. Ko and J.-M. Yeom, Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in south and north korea, *Science of The Total Environment*, 802, 149726, 2022.
- [132] F. Sabo, M. Meroni, F. Waldner and F. Rembold, Is deeper always better? evaluating deep learning models for yield forecasting with small data, *Environmental Monitoring and Assessment*, 195, 10, 1153, 2023.
- [133] Z. Xu, H. Sun, T. Zhang, H. Xu, D. Wu and J. Gao, Evaluating established deep learning methods in constructing integrated remote sensing drought index: A case study in china, *Agricultural Water Management*, 286, 108405, 2023.
- [134] S. Vulova, F. Meier, A. D. Rocha, J. Quanz, H. Nouri and B. Kleinschmit, Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence, *Science of The Total Environment*, 786, 147293, 2021.
- [135] X. Zhou, Q. Xin, Y. Dai and W. Li, A deep-learning-based experiment for benchmarking the performance of global terrestrial vegetation phenology models, *Global Ecology and Biogeography*, 30, 11, 2178–2199, 2021.
- [136] M. Fathi, R. Shah-Hosseini and A. Moghimi, 3d-resnet-bilstm model: A deep learning model for county-level soybean yield prediction with time-series sentinel-1, sentinel-2 imagery, and daymet data, *Remote Sensing*, 15, 23, 5551, 2023.

- [137] P. M. Maier, S. Keller and S. Hinz, Deep learning with WASI simulation data for estimating chlorophyll a concentration of inland water bodies, *Remote Sensing*, 13, 4, 718, 2021.
- [138] D. Fan, H. He, R. Wang, Y. Zeng, B. Fu, Y. Xiong, L. Liu, Y. Xu and E. Gao, Chlnet: A novel hybrid 1d cnn-svr algorithm for estimating ocean surface chlorophyll-a, *Frontiers in Marine Science*, 9, 934536, 2022.
- [139] S. S. Mukonza and J.-L. Chiang, Micro-climate computed machine and deep learning models for prediction of surface water temperature using satellite data in mundane water reservoir, *Water*, 14, 18, 2935, 2022.
- [140] M. Salah, H. Higa, J. Ishizaka and S. I. Salem, 1d convolutional neural network-based chlorophyll-a retrieval algorithm for sentinel-2 multispectral instrument in various trophic states., *Sensors & Materials*, 35, 2023.
- [141] A. Ivanda, L. Šerić, D. Žagar and K. Oštir, An application of 1d convolution and deep learning to remote sensing modelling of secchi depth in the northern adriatic sea, *Big earth data*, 8, 1, 82–114, 2024.
- [142] Y. Zeng, T. Liang, D. Fan and H. He, A novel algorithm for the retrieval of chlorophyll a in marine environments using deep learning, *Water*, 15, 21, 3864, 2023.
- [143] M. Salah, H. Higa, J. Ishizaka and S. I. Salem, B1d-cnn: A novel convolution neural network-based chlorophyll-a retrieval algorithm for sentinel-2 data, *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, 3950–3953, IEEE, 2023.
- [144] V. Sagan, M. Maimaitijiang, S. Bhadra, M. Maimaitiyiming, D. R. Brown, P. Sidike and F. B. Fritschi, Field-scale crop yield prediction using multi-temporal worldview-3 and planetscope satellite data and deep learning, *ISPRS journal of photogrammetry and remote sensing*, 174, 265–281, 2021.
- [145] J. Lee, J. Im, D.-H. Cha, H. Park and S. Sim, Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data, *Remote Sensing*, 12, 1, 108, 2019.
- [146] J. Zhong, J. Sun, Z. Lai and Y. Song, Nearshore bathymetry from icesat-2 lidar and sentinel-2 imagery datasets using deep learning approach, *Remote Sensing*, 14, 17, 4229, 2022.
- [147] M. Papadomanolaki, S. Christodoulidis, K. Karantzalos and M. Vakalopoulou, Unsupervised multistep deformable registration of remote sensing imagery based on deep learning, *Remote Sensing*, 13, 7, 1294, 2021.
- [148] E. Zhang, Y. Fu, J. Wang, L. Liu, K. Yu and J. Peng, Msac-net: 3d multi-scale attention convolutional network for multi-spectral imagery pansharpening, *Remote Sensing*, 14, 12, 2761, 2022.
- [149] Z. Wang, S. Fang and J. Zhang, Spatiotemporal fusion model of remote sensing images combining single-band and multi-band prediction, *Remote Sensing*, 15, 20, 4936, 2023.

- [150] A. S. Terliksiz and D. T. Altýlar, Use of deep neural networks for crop yield prediction: A case study of soybean yield in lauderdale county, alabama, usa, *2019 8th international conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4, IEEE, 2019.
- [151] M. Qiao, X. He, X. Cheng, P. Li, H. Luo, L. Zhang and Z. Tian, Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3d convolutional neural networks, *International Journal of Applied Earth Observation and Geoinformation*, 102, 102436, 2021.
- [152] N. Wang, Z. Ma, P. Huo and X. Liu, Predicting crop yield using 3d convolutional neural network with dimension reduction and metric learning, *2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, 1004–1008, IEEE, 2023.
- [153] N. Wang, Z. Ma, P. Huo, X. Liu, Z. He and K. Lu, 3d convolutional neural network with dimension reduction and metric learning for crop yield prediction based on remote sensing data, *Applied Sciences*, 13, 24, 13305, 2023.
- [154] R. Fernandez-Beltran, T. Baidar, J. Kang and F. Pla, Rice-yield prediction with multi-temporal sentinel-2 data and 3d cnn: A case study in nepal, *Remote Sensing*, 13, 7, 1391, 2021.
- [155] T. Fei, B. Huang, X. Wang, J. Zhu, Y. Chen, H. Wang and W. Zhang, A hybrid deep learning model for the bias correction of sst numerical forecast products using satellite data, *Remote Sensing*, 14, 6, 1339, 2022.
- [156] L. Wang, W. Li, X. Wang and J. Xu, Remote sensing image analysis and prediction based on improved pix2pix model for water environment protection of smart cities, *PeerJ Computer Science*, 9, e1292, 2023.
- [157] G. Chen, Q. Pei and M. Kamruzzaman, Remote sensing image quality evaluation based on deep support value learning networks, *Signal Processing: Image Communication*, 83, 115783, 2020.
- [158] L. Zhang, P. Liu, L. Wang, J. Liu, B. Song, Y. Zhang, G. He and H. Zhang, Improved 1-km-resolution hourly estimates of aerosol optical depth using conditional generative adversarial networks, *Remote Sensing*, 13, 19, 3834, 2021.
- [159] G. Bahl and F. Lafarge, Scanner neural network for on-board segmentation of satellite images, *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, 3504–3507, IEEE, 2022.
- [160] H. Ghandorh, W. Boulila, S. Masood, A. Koubaa, F. Ahmed and J. Ahmad, Semantic segmentation and edge detection—approach to road detection in very high resolution satellite images, *Remote Sensing*, 14, 3, 613, 2022.
- [161] R. Li, S. Zheng, C. Duan, L. Wang and C. Zhang, Land cover classification from remote sensing images based on multi-scale fully convolutional network, *Geo-spatial information science*, 25, 2, 278–294, 2022.

- [162] E. Saralioglu and O. Gungor, Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network, *Geocarto International*, 37, 2, 657–677, 2022.
- [163] S. Mohammadi, M. Belgiu and A. Stein, 3d fully convolutional neural networks with intersection over union loss for crop mapping from multi-temporal satellite images, *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 5834–5837, IEEE, 2021.
- [164] I. Gallo, L. Ranghetti, N. Landro, R. La Grassa and M. Boschetti, In-season and dynamic crop mapping using 3d convolution neural networks and sentinel-2 time series, *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, 335–352, 2023.
- [165] S. Mohammadi, M. Belgiu and A. Stein, Improvement in crop mapping from satellite image time series by effectively supervising deep neural networks, *ISPRS Journal of Photogrammetry and Remote Sensing*, 198, 272–283, 2023.
- [166] E. Kalinicheva, J. Sublime and M. Trocan, Unsupervised satellite image time series clustering using object-based approaches and 3d convolutional autoencoder, *Remote Sensing*, 12, 11, 1816, 2020.
- [167] K. Meshkini, F. Bovolo and L. Bruzzone, An unsupervised change detection approach for dense satellite image time series using 3d cnn, *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4336–4339, IEEE, 2021.
- [168] K. Meshkini, F. Bovolo and L. Bruzzone, A 3d cnn approach for change detection in hr satellite image time series based on a pretrained 2d cnn, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 143–150, 2022.
- [169] N. Noy and D. McGuinness, Ontology development 101: A guide to creating your first ontology, *Knowledge Systems Laboratory*, 32, 01 2001.
- [170] A. A. Salatino, T. Thanapalasingam, A. Mannocci, A. Birukou, F. Osborne and E. Motta, The Computer Science Ontology: A Comprehensive Automatically-Generated Taxonomy of Research Areas, *Data Intelligence*, 2, 3, 379–416, 07 2020.
- [171] Protégé Project, Protégé, <https://protege.stanford.edu/>, 2023, accessed: 2024-11-07.
- [172] S. Lohmann, V. Link, E. Marbach and S. Negru, Webvowl: Web-based visualization of ontologies, *Knowledge Engineering and Knowledge Management: EKAW 2014 Satellite Events, VISUAL, EKMI, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers. 19*, 154–158, Springer, 2015.
- [173] Z. Lee, S. Shang, C. Hu, K. Du, A. Weidemann, W. Hou, J. Lin and G. Lin, Secchi disk depth: A new theory and mechanistic model for underwater visibility, *Remote sensing of environment*, 169, 139–149, 2015.
- [174] D. V. Chapman, *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring, Chapter 3 - Selection of water quality variables*, 59–125, CRC Press, 1996.

- [175] E. T. Harvey, J. Walve, A. Andersson, B. Karlson and S. Kratzer, The effect of optical properties on secchi depth and implications for eutrophication management, *Frontiers in Marine Science*, 5, 496, 2019.
- [176] K. Teubner, I. Teubner, K. Pall, W. Kabas, M. Tolotti, T. Ofenböck and M. T. Dokulil, New emphasis on water transparency as socio-ecological indicator for urban water: Bridging ecosystem service supply and sustainable ecosystem health, *Frontiers in Environmental Science*, 8, 573724, 2020.
- [177] A. Morel and L. Prieur, Analysis of variations in ocean color 1, *Limnology and oceanography*, 22, 4, 709–722, 1977.
- [178] Y. He, Z. Gong, Y. Zheng and Y. Zhang, Inland reservoir water quality inversion and eutrophication evaluation using BP neural network and remote sensing imagery: a case study of Dashahe reservoir, *Water*, 13, 20, 2844, 2021.
- [179] Z. Lee and J. Tang, The two faces of “case-1” water, *Journal of Remote Sensing*, 2022.
- [180] C. D. Mobley, D. Stramski, W. Paul Bissett and E. Boss, Optical modeling of ocean waters: Is the case 1-case 2 classification still useful?, *Oceanography*, 17, SPL. ISS. 2, 60, 2004.
- [181] L. Andres, K. Boateng, C. Borja-Vega and E. Thomas, A review of in-situ and remote sensing technologies to monitor water and sanitation interventions, *Water*, 10, 6, 756, 2018.
- [182] A. Ivanda, L. Šerić, M. Bugarić and M. Braović, Mapping chlorophyll-a concentrations in the Kaštela Bay and Brač Channel using ridge regression and Sentinel-2 satellite images, *Electronics*, 10, 23, 3004, 2021.
- [183] R. Katlane, C. Dupouy, B. El Kilani, J. C. Berges et al., Estimation of chlorophyll and turbidity using sentinel 2A and EO1 data in Kneiss Archipelago Gulf of Gabes, Tunisia, *International Journal of Geosciences*, 11, 10, 708, 2020.
- [184] A. Dogliotti, K. Ruddick, B. Nechad, D. Doxaran and E. Knaeps, A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters, *Remote Sensing of Environment*, 156, 157–168, 2015.
- [185] Z. Mao, J. Chen, D. Pan, B. Tao and Q. Zhu, A regional remote sensing algorithm for total suspended matter in the East China Sea, *Remote Sensing of Environment*, 124, 819–831, 2012.
- [186] Y. Zhang, K. Shi, X. Sun, Y. Zhang, N. Li, W. Wang, Y. Zhou, W. Zhi, M. Liu, Y. Li et al., Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods, *GIScience & Remote Sensing*, 59, 1, 1367–1383, 2022.
- [187] C. Zhao, D. Yu, L. Yang, Y. Zhou, H. Gao, X. Bian and Y. Li, Remote sensing algorithms of seawater transparency: A review, *2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS)*, 744–749, 2022.

- [188] L. F. Arias-Rodriguez, U. F. Tüzün, Z. Duan, J. Huang, Y. Tuo and M. Disse, Global water quality of inland waters with harmonized landsat-8 and sentinel-2 using cloud-computed machine learning, *Remote Sensing*, 15, 5, 2023.
- [189] E. T. Slonecker, D. K. Jones and B. A. Pellerin, The new Landsat 8 potential for remote sensing of colored dissolved organic matter (CDOM), *Marine Pollution Bulletin*, 107, 2, 518–527, 2016, resetting the Bar: Establishing Baselines for Persistent Contaminants after Hurricane Sandy in the Coastal Environments of New Jersey and New York, USA.
- [190] C. A. Baughman, B. M. Jones, K. K. Bartz, D. B. Young and C. E. Zimmerman, Reconstructing Turbidity in a Glacially Influenced Lake Using the Landsat TM and ETM+ Surface Reflectance Climate Data Record Archive, Lake Clark, Alaska, *Remote Sensing*, 7, 10, 13692–13710, 2015.
- [191] E. Anassontzis, A. Ball, A. Belias, A. Fotiou, G. Grammatikakis, H. Kontogiannis, P. Koske, S. Koutsoukos, V. Lykoussis, E. Markopoulos, A. Psallidas, L. Resvanis, I. Siotis, S. Stavrakakis, G. Stavropoulos and V. Zhukov, Water transparency measurements in the deep Ionian Sea, *Astroparticle Physics*, 34, 4, 187–197, 2010.
- [192] A. Y. Morel and H. R. Gordon, Report of the working group on water color, *Boundary-Layer Meteorology*, 18, 3, 343–355, 1980.
- [193] M. H. Gholizadeh, A. M. Melesse and L. Reddi, A comprehensive review on water quality parameters estimation using remote sensing techniques, *Sensors*, 16, 8, 1298, 2016.
- [194] D. Ficek and T. Zapadka, Variability of bio-optical parameters in lake jasioń północny and lake jasioń południowy, *Limnological Review*, 10, 2, 67–76, 2010.
- [195] A. A. Padial and S. M. Thomaz, Prediction of the light attenuation coefficient through the secchi disk depth: empirical modeling in two large neotropical ecosystems, *Limnology*, 9, 143–151, 2008.
- [196] Y. Zhang, X. Liu, Y. Yin, M. Wang and B. Qin, Predicting the light attenuation coefficient through secchi disk depth and beam attenuation coefficient in a large, shallow, freshwater lake, *Hydrobiologia*, 693, 29–37, 2012.
- [197] B. Lednicka and M. Kubacka, Semi-empirical model of remote-sensing reflectance for chosen areas of the southern Baltic, *Sensors*, 22, 3, 1105, 2022.
- [198] S. Budhiman, T. W. Hobma and Z. Vekerdy, Remote sensing for mapping tsm concentration in mahakam delta: An analytical approach, *Proceedings of the 13th OMISAR Workshop on Validation and Application of Satellite Data for Marine Resources Conservation*, Bali, Indonesia, 2004, conference or Workshop Item.
- [199] C. A. Paulson and J. J. Simpson, Irradiance Measurements in the Upper Ocean, *Journal of Physical Oceanography*, 7, 6, 952 – 956, 1977.
- [200] F. Stravisi et al., Optical seawater properties in the Gulf of Trieste, *BOLLETTINO DELLA SOCIETÀ ADRIATICA DI SCIENZE*, 79, 61–75, 1999.

- [201] Z. Lee, S. Shang, K. Du and J. Wei, Resolving the long-standing puzzles about the observed Secchi depth relationships, *Limnology and Oceanography*, 63, 6, 2321–2336, 2018.
- [202] B. Umer and V. Malačič, Biexponential decrease of PAR in coastal waters (Northern Adriatic), *Geofizika*, 39, 1, 1–29, 2022.
- [203] Z. Lee, S. Shang, K. Du, G. Lin, T. Liu and L. Zoffoli, Estimating the transmittance of visible solar radiation in the upper ocean using Secchi disk observations, *Journal of Geophysical Research: Oceans*, 124, 3, 1434–1444, 2019.
- [204] V. Sagan, K. T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B. A. Greeling, S. Maalouf and C. Adams, Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing, *Earth-Science Reviews*, 205, 103187, 2020.
- [205] D. A. Maciel, C. C. F. Barbosa, E. M. L. de Moraes Novo, R. Flores Júnior and F. N. Begliomini, Water clarity in Brazilian water assessed using Sentinel-2 and machine learning methods, *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, 134–152, 2021.
- [206] M. Shen, H. Duan, Z. Cao, K. Xue, T. Qi, J. Ma, D. Liu, K. Song, C. Huang and X. Song, Sentinel-3 olci observations of water clarity in large lakes in eastern China: Implications for SDG 6.3.2 evaluation, *Remote Sensing of Environment*, 247, 111950, 2020.
- [207] H. J. Rubin, D. A. Lutz, B. G. Steele, K. L. Cottingham, K. C. Weathers, M. J. Ducey, M. Palace, K. M. Johnson and J. W. Chipman, Remote sensing of lake water clarity: Performance and transferability of both historical algorithms and machine learning, *Remote Sensing*, 13, 8, 1434, 2021.
- [208] R. J. W. Brewin, T. G. Brewin, J. Phillips, S. Rose, A. Abdulaziz, W. Wimmer, S. Sathyendranath and T. Platt, A printable device for measuring clarity and colour in lake and nearshore waters, *SENSORS*, 19, 4, FEB 2 2019.
- [209] M. Hoyer, V and D. E. Canfield, Jr., Volunteer-collected water quality data can be used for science and management, *LAKE AND RESERVOIR MANAGEMENT*, 37, 3, 235–245, AUG 26 2021.
- [210] Secchi Disk Project, <https://www.playingwithdata.com/secchi-disk-project/>, 2013, accessed: 2022-09-30.
- [211] N. Menon, G. George, R. Ranith, V. Sajin, S. Murali, A. Abdulaziz, R. J. W. Brewin and S. Sathyendranath, Citizen science tools reveal changes in estuarine water quality following demolition of buildings, *REMOTE SENSING*, 13, 9, MAY 2021.
- [212] D. Canfield, C. Brown, R. Bachmann and M. Hoyer, Volunteer lake monitoring: Testing the reliability of data collected by the Florida Lakewatch program, *LAKE AND RESERVOIR MANAGEMENT*, 18, 1, 1–9, MAR 2002.
- [213] K. M. A. Luis, J. E. Rheuban, M. T. Kavanaugh, D. M. Glover, J. Wei, Z. Lee and S. C. Doney, Capturing coastal water clarity variability with Landsat 8, *MARINE POLLUTION BULLETIN*, 145, 96–104, AUG 2019.

- [214] E. S. Deutsch, J. A. Cardille, T. Koll-Egyed and M.-J. Fortin, Landsat 8 lake water clarity empirical algorithms: Large-scale calibration and validation using government and citizen science data from across Canada, *REMOTE SENSING*, 13, 7, APR 2021.
- [215] A. Kulshreshtha and P. Shanmugam, Estimation of underwater visibility in coastal and inland waters using remote sensing data, *Environmental monitoring and assessment*, 189, 1–18, 2017.
- [216] Y. Cui, Z. Yan, J. Wang, S. Hao and Y. Liu, Deep learning-based remote sensing estimation of water transparency in shallow lakes by combining Landsat 8 and Sentinel 2 images, *Environmental Science and Pollution Research*, 29, 4401–4413, 2022.
- [217] H. Guo, X. Zhu, J. Jeanne Huang, Z. Zhang, S. Tian and Y. Chen, An enhanced deep learning approach to assessing inland lake water quality and its response to climate and anthropogenic factors, *Journal of Hydrology*, 620, 129466, 2023.
- [218] Y. He, Z. Lu, W. Wang, D. Zhang, Y. Zhang, B. Qin, K. Shi and X. Yang, Water clarity mapping of global lakes using a novel hybrid deep-learning-based recurrent model with LandsatOLI images, *Water Research*, 215, 118241, 2022.
- [219] M. Ahmed, R. Mumtaz, Z. Anwar, A. Shaukat, O. Arif and F. Shafait, A multi-step approach for optically active and inactive water quality parameter estimation using deep learning and remote sensing, *Water*, 14, 13, 2022.
- [220] J. Pyo, S. M. Hong, J. Jang, S. Park, J. Park, J. H. Noh and K. H. Cho, Drone-borne sensing of major and accessory pigments in algae using deep learning modeling, *GI-Science & Remote Sensing*, 59, 1, 310–332, 2022.
- [221] H. Shanjiang, H. Yan, T. Bangyi, Y. Jiayong and C. Weibiao, Classification of sea and land waveforms based on deep learning for airborne laser bathymetry, *Eng. Infrared and Laser Engineering*, 48, 11, 1113004–1113004, 2019.
- [222] M. Zavatarelli, F. Raicich, D. Bregant, A. Russo and A. Artegiani, Climatological biogeochemical characteristics of the Adriatic Sea, *Journal of Marine Systems*, 18, 1, 227–263, 1998.
- [223] G. H. Blake and D. Topalović, *The maritime boundaries of the Adriatic Sea*, 1–5, Ibru, 1996.
- [224] G. Rodrigues, M. Potes, A. M. Penha, M. J. Costa and M. M. Morais, The use of Sentinel-3/OLCI for monitoring the water quality and optical water types in the largest Portuguese reservoir, *Remote Sensing*, 14, 9, 2172, 2022.
- [225] EO Browser, Sinergise Ltd, <https://apps.sentinel-hub.com/eo-browser/>, accessed: 2022-09-20.
- [226] Sentinel-3 OLCI L1B, <https://docs.sentinel-hub.com/api/latest/data/sentinel-3-olci-l1b/>, Accessed: 2022-12-01.
- [227] Sentinel-3 OLCI Land User Handbook, <https://sentinel.esa.int/documents/247904/4598066/Sentinel-3-OLCI-Land-Handbook.pdf>, Accessed: 2023-04-27.

- [228] F. Steinmetz, P.-Y. Deschamps and D. Ramon, Atmospheric correction in presence of sun glint: application to meris, *Opt. Express*, 19, 10, 9783–9800, May 2011.
- [229] H. R. Gordon and M. Wang, Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawifs: a preliminary algorithm, *Appl. Opt.*, 33, 3, 443–452, Jan 1994.
- [230] Q. Vanhellemont and K. Ruddick, Atmospheric correction of sentinel-3/olci data for mapping of suspended particulate matter and chlorophyll-a concentration in belgian turbid coastal waters, *Remote Sensing of Environment*, 256, 112284, 2021.
- [231] M. Shah, M. S. Raval and S. Divakaran, A Deep Learning Perspective to Atmospheric Correction of Satellite Images, *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 346–349, 2022.
- [232] Radiometric Resolution - 21 bands in VIS/SWIR, <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-olci/resolutions/radiometric>, Accessed: 2022-12-01.
- [233] R. Fernandez-Moran, L. Gómez-Chova, L. Alonso, G. Mateo-García and D. López-Puigdollers, Towards a novel approach for Sentinel-3 synergistic OLCI/SLSTR cloud and cloud shadow detection based on stereo cloud-top height estimation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 238–253, 2021.
- [234] E. U. H. Qazi, A. Almorjan and T. Zia, A One-Dimensional Convolutional Neural Network (1D-CNN) Based Deep Learning System for Network Intrusion Detection, *Applied Sciences*, 12, 16, 7986, 2022.
- [235] M. Bandyopadhyay, Multi-stack hybrid CNN with non-monotonic activation functions for hyperspectral satellite image classification, *Neural Computing and Applications*, 33, 21, 14809–14822, 2021.
- [236] G. Van Rossum and F. L. Drake Jr, *Python tutorial*, 620, Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.
- [237] F. Chollet et al., Keras, 2015.
- [238] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., Tensorflow: A system for large-scale machine learning, *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 265–283, 2016.
- [239] P. Diez, P. Neittaanmäki, J. Periaux, T. Tuovinen and J. Pons-Prats, *Computation and Big Data for Transport: Digital Innovations in Surface and Air Transport Systems*, 54, 221–222, Springer Nature, 2020.
- [240] R. A. Lateef and A. R. Abbas, Tuning the Hyperparameters of the 1D CNN Model to Improve the Performance of Human Activity Recognition, *Engineering and Technology Journal*, 40, 04, 547–554, 2022.
- [241] S. V. Stehman and R. L. Czaplewski, Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles, *Remote Sensing of Environment*, 64, 3, 331–344, 1998.

- [242] A. Knudby, *Remote Sensing: 7. Accuracy assessment*, 72–76, 2021.
- [243] J. Yue, H. Feng, G. Yang and Z. Li, A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy, *Remote Sensing*, 10, 1, 66, 2018.
- [244] J. Li and A. D. Heap, A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, *Ecological Informatics*, 6, 3, 228–241, 2011.
- [245] U. Sara, M. Akter and M. S. Uddin, Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study, *Journal of Computer and Communications*, 7, 3, 8–18, 2019.
- [246] P. Qin, S. Simis and G. Tilstone, Radiometric validation of atmospheric correction for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC, *Remote Sensing of Environment*, 200, 263–280, 10 2017.
- [247] E. Mansouri, F. Feizi, A. Jafari Rad and M. Arian, Remote-sensing data processing with the multivariate regression analysis method for iron mineral resource potential mapping: a case study in the Sarvian area, central Iran, *Solid Earth*, 9, 2, 373–384, 2018.
- [248] A. Agarwal and M. Saxena, Assessment of pollution by physicochemical water parameters using regression analysis: a case study of Gagan river at Moradabad-India, *Advances in Applied Science Research*, 2, 2, 185–189, 2011.
- [249] A. Upreti, Machine learning application in g.i.s. and remote sensing: An overview, *Preprints*, 2022, 2022070302, 1–14, 2022.
- [250] G. James, D. Witten, T. Hastie and R. Tibshirani, *An introduction to statistical learning, Chapter 3: Linear Regression*, 112, 59–120, Springer, 2013.
- [251] G. James, D. Witten, T. Hastie and R. Tibshirani, *An introduction to statistical learning, Chapter 6: Linear Model Selection and Regularization*, 112, 217–236, Springer, 2013.
- [252] W.-Y. Loh, Classification and regression trees, *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1, 1, 14–23, 2011.
- [253] L. Breiman, Random forests, *Machine learning*, 45, 5–32, 2001.
- [254] A. Prabhat and V. Khullar, Sentiment classification on big data using naïve bayes and logistic regression, *2017 International Conference on Computer Communication and Informatics (ICCCI)*, 1–5, 2017.
- [255] A. J. Smola and B. Schölkopf, A tutorial on support vector regression, *Statistics and computing*, 14, 199–222, 2004.
- [256] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830, 2011.

- [257] QGIS Development Team, *QGIS Geographic Information System*, QGIS Association.
- [258] E. Eker-Develi, J.-F. Berthon and G. Free, Impact of environmental factors on phytoplankton composition and their marker pigments in the northern Adriatic Sea, *Oceanologia*, 64, 4, 615–630, 2022.
- [259] J. L. Mancini, Numerical estimates of coliform mortality rates under various conditions, *Journal (Water Pollution Control Federation)*, 2477–2484, 1978.
- [260] M. De Marchis, G. Freni and E. Napoli, Modelling of *E. coli* distribution in coastal areas subjected to combined sewer overflows, *Water Science and Technology*, 68, 5, 1123–1136, 09 2013.
- [261] G. Huang, R. A. Falconer and B. Lin, Integrated hydro-bacterial modelling for predicting bathing water quality, *Estuarine, Coastal and Shelf Science*, 188, 145–155, 2017.
- [262] A. Lešek and D. Žagar, Use of the particle tracking method for modelling the transport and deceasing of *Escherichia coli* in the sea, *Acta hydrotechnica*, 31, 55, 119–142, 2018.
- [263] A. Lešek, G. Novak, M. Četina and D. Žagar, Modelling transport and decease of *Escherichia coli* in the coastal sea by using the particle tracking method, 633–634, Abstract book: hosted by Spain Water and IWHR, China. Warszawa: IAHR, 2020/2021, 2020.
- [264] J. Soriano-González, E. P. Urrego, X. Sòria-Perpinyà, E. Angelats, C. Alcaraz, J. Delegido, A. Ruíz-Verdú, C. Tenjo, E. Vicente and J. Moreno, Towards the combination of c2rcc processors for improving water quality retrieval in inland and coastal areas, *Remote Sensing*, 14, 5, 1124, 2022.
- [265] D. Kyryliuk and S. Kratzer, Evaluation of Sentinel-3A OLCI products derived using the Case-2 Regional CoastColour processor over the Baltic Sea, *Sensors*, 19, 16, 3609, 2019.
- [266] S. Hafeez, M. S. Wong, H. C. Ho, M. Nazeer, J. Nichol, S. Abbas, D. Tang, K. H. Lee and L. Pun, Comparison of machine learning algorithms for retrieval of water quality indicators in case-ii waters: A case study of hong kong, *Remote Sensing*, 11, 6, 2019.
- [267] J. Chen, S. Chen, R. Fu, D. Li, H. Jiang, C. Wang, Y. Peng, K. Jia and B. J. Hicks, Remote sensing big data for water environment monitoring: Current status, challenges, and future prospects, *Earth's Future*, 10, 2, e2021EF002289, 2022.
- [268] N. Nasir, A. Kansal, O. Alshaltone, F. Barneih, A. Shanableh, M. Al-Shabi and A. Al Shammaa, Deep learning detection of types of water-bodies using optical variables and ensembling, *Intelligent Systems with Applications*, 18, 200222, 2023.
- [269] Y. Li, K. Shi, Y. Zhang, G. Zhu, Y. Zhang, Z. Wu, M. Liu, Y. Guo and N. Li, Analysis of water clarity decrease in xin'anjiang reservoir, china, from 30-year landsat tm, etm+, and oli observations, *Journal of Hydrology*, 590, 125476, 2020.

- [270] A. O. Ok, Automated extraction of buildings and roads in a graph partitioning framework, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 79–84, 2013.
- [271] Z. Chen, L. Deng, Y. Luo, D. Li, J. Marcato Junior, W. Nunes Gonçalves, A. Awal Md Nurunnabi, J. Li, C. Wang and D. Li, Road extraction in remote sensing data: A survey, *International Journal of Applied Earth Observation and Geoinformation*, 112, 102833, 2022.
- [272] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty and A. Alamri, Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review, *Remote Sensing*, 12, 9, 2020.
- [273] V. Mnih, *Machine learning for aerial image labeling*, University of Toronto (Canada), 2013.
- [274] S. Saito, T. Yamashita and Y. Aoki, Multiple object extraction from aerial imagery with convolutional neural networks, *Electronic Imaging*, 28, 1–9, 2016.
- [275] Z. Zhong, J. Li, W. Cui and H. Jiang, Fully convolutional networks for building and road extraction: Preliminary results, *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1591–1594, 2016.
- [276] Y. Wei, Z. Wang and M. Xu, Road structure refined cnn for road extraction in aerial image, *IEEE Geoscience and Remote Sensing Letters*, 14, 5, 709–713, 2017.
- [277] P. Li, Y. Zang, C. Wang, J. Li, M. Cheng, L. Luo and Y. Yu, Road network extraction via deep learning and line integral convolution, *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1599–1602, 2016.
- [278] J. Radoux, G. Chomé, D. C. Jacques, F. Waldner, N. Bellemans, N. Matton, C. Lamarche, R. D’Andrimont and P. Defourny, Sentinel-2’s potential for sub-pixel landscape feature detection, *Remote Sensing*, 8, 6, 2016.
- [279] T. Hoeseer and C. Kuenzer, Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends, *Remote Sensing*, 12, 10, 2020.
- [280] S. Oehmcke, C. Thrysoe, A. Borgstad, M. A. V. Salles, M. Brandt and F. Gieseke, Detecting hardly visible roads in low-resolution satellite time series data, *2019 IEEE international conference on big data (big data)*, 2403–2412, IEEE, 2019.
- [281] R. Abdelfattah and K. Chokmani, A semi automatic off-roads and trails extraction method from sentinel-1 data, *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3728–3731, IEEE, 2017.
- [282] C. Ayala, R. Sesma, C. Aranda and M. Galar, A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery, *Remote Sensing*, 13, 16, 2021.
- [283] P. Mooney, M. Minghini et al., A review of openstreetmap data, *Mapping and the citizen sensor*, 37–59, 2017.

- [284] Copernicus Data Space Ecosystem, Copernicus data space ecosystem, Programme of the European Union, managed by the European Space Agency (ESA), 2024, accessed: August 24, 2024.
- [285] Copernicus Data Space Ecosystem, Sentinel-2 Mission, <https://documentation.dataspace.copernicus.eu/Data/SentinelMissions/Sentinel2.html>, 2023, accessed: 2024-09-02.
- [286] E. Beauxis-Aussalet and L. Hardman, Visualization of confusion matrix for non-expert users, *IEEE Conference on Visual Analytics Science and Technology (VAST)-Poster Proceedings*, 1–2, sn, 2014.
- [287] A. Tharwat, Classification assessment methods, *Applied computing and informatics*, 17, 1, 168–192, 2021.
- [288] Z. Qian, Y. Cao, Z. Shi, L. Qiu and C. Shi, A semantic segmentation method for remote sensing images based on deeplab v3, *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, 396–400, IEEE, 2021.
- [289] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778, 2016.
- [290] M. Shafiq and Z. Gu, Deep residual learning for image recognition: A survey, *Applied Sciences*, 12, 18, 2022.
- [291] N. Y. Q. Abderrahim, S. Abderrahim and A. Rida, Road segmentation using u-net architecture, *2020 IEEE international conference of moroccan geomatics (Morgeo)*, 1–4, IEEE, 2020.
- [292] L. Šerić, A. Ivanda, M. Bugarić and D. Stipaničev, Empirical fire propagation potential from a balanced dataset, *9th International Conference on Forest Fire Research*, 162–165, 2022.
- [293] E. Commission, J. R. Centre, D. De Rigo, G. Libertà, T. Houston Durrant, T. Artés Vivancos and J. San-Miguel-Ayanz, *Forest fire danger extremes in Europe under climate change – Variability and uncertainty*, Publications Office of the European Union, 2017.
- [294] J. Costa-Saura, D. Spano, C. Sirca and V. Bacciu, Contrasting patterns and interpretations between a fire spread simulator and a machine learning model when mapping burn probabilities: A case study for mediterranean areas, *Environmental Modelling & Software*, 163, 105685, 2023.
- [295] E. Chuvieco, M. Yebra, S. Martino, K. Thonicke, M. Gómez-Giménez, J. San-Miguel, D. Oom, R. Velea, F. Mouillot, J. R. Molina, A. I. Miranda, D. Lopes, M. Salis, M. Bugaric, M. Sofiev, E. Kadantsev, I. Z. Gitas, D. Stavrakoudis, G. Eftychidis, A. Bar-Massada, A. Neidermeier, V. Pampanoni, M. L. Pettinari, F. Arrogante-Funes, C. Ochoa, B. Moreira and D. Viegas, Towards an integrated approach to wildfire risk assessment: When, where, what and how may the landscapes burn, *Fire*, 6, 5, 2023.

- [296] M. González, S. Gómez-González, A. Lara, R. Garreaud and I. Díaz-Hormazábal, The 2010-2015 megadrought and its influence on the fire regime in central and south-central chile, *Ecosphere*, 9, e02300, 08 2018.
- [297] D. Stipaničev, M. Bugarić, D. Krstinić, L. Šerić, T. Jakovčević, M. Braović and M. Štula, New generation of automatic ground based wildfire surveillance systems, *Advances in forest fire reasearch*, 2014.
- [298] H. R. Pourghasemi, A. Gayen, R. Lasaponara and J. P. Tiefenbacher, Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling, *Environmental research*, 184, 109321, 2020.
- [299] A. B. Massada, A. D. Syphard, S. I. Stewart and V. C. Radeloff, Wildfire ignition-distribution modelling: a comparative study in the huron–manistee national forest, michigan, usa, *International journal of wildland fire*, 22, 2, 174–183, 2012.
- [300] L. A. Dimuccio, R. Ferreira, L. Cunha and A. C. de Almeida, Regional forest-fire susceptibility analysis in central portugal using a probabilistic ratings procedure and artificial neural network weights assignment, *International Journal of Wildland Fire*, 20, 6, 776–791, 2011.
- [301] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia and J. M. Pereira, Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest, *Forest Ecology and Management*, 275, 117–129, 2012.
- [302] H. Hong, P. Tsangaratos, I. Ilija, J. Liu, A.-X. Zhu and C. Xu, Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. the case of dayu county, china, *Science of the total environment*, 630, 1044–1056, 2018.
- [303] N. N. Thach, D. B.-T. Ngo, P. Xuan-Canh, N. Hong-Thi, B. H. Thi, H. Nhat-Duc and T. B. Dieu, Spatial pattern assessment of tropical forest fire danger at thuan chau area (vietnam) using gis-based advanced machine learning algorithms: A comparative study, *Ecological informatics*, 46, 74–85, 2018.
- [304] S. J. Kim, C.-H. Lim, G. S. Kim, J. Lee, T. Geiger, O. Rahmati, Y. Son and W.-K. Lee, Multi-temporal analysis of forest fire probability using socio-economic and environmental variables, *Remote Sensing*, 11, 1, 86, 2019.
- [305] G. Zhang, M. Wang and K. Liu, Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china, *International Journal of Disaster Risk Science*, 10, 3, 386–403, 2019.
- [306] G. Zhang, M. Wang and K. Liu, Deep neural networks for global wildfire susceptibility modelling, *Ecological Indicators*, 127, 107735, 2021.
- [307] H. V. Le, D. A. Hoang, C. T. Tran, P. Q. Nguyen, V. H. T. Tran, N. D. Hoang, M. Amiri, T. P. T. Ngo, H. V. Nhu, T. V. Hoang and D. Tien Bui, A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas, *Ecological Informatics*, 63, 101300, 2021.

- [308] A. Bjånes, R. De La Fuente and P. Mena, A deep learning ensemble model for wildfire susceptibility mapping, *Ecological Informatics*, 65, 101397, 2021.
- [309] European Forest Fire Information System, Effis - european forest fire information system, 2024, accessed: 2024-08-22.
- [310] D. M. Szpakowski and J. L. R. Jensen, A review of the applications of remote sensing in fire ecology, *Remote Sensing*, 11, 22, 2019.
- [311] B. Wotton, D. Martell and K. Logan, Climate change and people-caused forest fire occurrence in ontario, *Climatic change*, 60, 3, 275–295, 2003.
- [312] M. Conedera, D. Torriani, C. Neff, C. Ricotta, S. Bajocco and G. B. Pezzatti, Using monte carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region, *Forest Ecology and Management*, 261, 12, 2179–2187, 2011.
- [313] C. Quintano, A. Fernández-Manso and O. Fernández-Manso, Combination of landsat and sentinel-2 msi data for initial assessing of burn severity, *International journal of applied earth observation and geoinformation*, 64, 221–225, 2018.
- [314] E. Barboza Castillo, E. Y. Turpo Cayo, C. M. de Almeida, R. Salas López, N. B. Rojas Briceño, J. O. Silva López, M. Á. Barrena Gurbillón, M. Oliva and R. Espinoza-Villar, Monitoring wildfires in the northeastern peruvian amazon using landsat-8 and sentinel-2 imagery in the gee platform, *ISPRS International Journal of Geo-Information*, 9, 10, 564, 2020.
- [315] T. A. Lima, R. Beuchle, A. Langner, R. C. Grecchi, V. C. Griess and F. Achard, Comparing sentinel-2 msi and landsat 8 oli imagery for monitoring selective logging in the brazilian amazon, *Remote sensing*, 11, 8, 961, 2019.
- [316] O. Satir, S. Berberoglu and C. Donmez, Mapping regional forest fire probability using artificial neural network model in a mediterranean forest ecosystem, *Geomatics, Natural Hazards and Risk*, 7, 5, 1645–1658, 2016.
- [317] M. Mohajane, R. Costache, F. Karimi, Q. B. Pham, A. Essahlaoui, H. Nguyen, G. Lan- eve and F. Oudija, Application of remote sensing and machine learning algorithms for forest fire mapping in a mediterranean area, *Ecological Indicators*, 129, 107869, 2021.
- [318] G. P. Petropoulos, C. Kontoes and I. Keramitsoglou, Burnt area delineation from a uni-temporal perspective based on landsat tm imagery classification using support vector machines, *International Journal of Applied Earth Observation and Geoinformation*, 13, 1, 70–80, 2011.
- [319] S. T. Seydi, M. Akhoondzadeh, M. Amani and S. Mahdavi, Wildfire damage assessment over australia using sentinel-2 imagery and modis land cover product within the google earth engine cloud platform, *Remote Sensing*, 13, 2, 2021.
- [320] D. P. Roy, H. Huang, L. Boschetti, L. Giglio, L. Yan, H. H. Zhang and Z. Li, Landsat-8 and sentinel-2 burned area mapping-a combined sensor multi-temporal change detection approach, *Remote Sensing of Environment*, 231, 111254, 2019.

- [321] M. Syifa, M. Panahi and C.-W. Lee, Mapping of post-wildfire burned area using a hybrid algorithm and satellite data: The case of the camp fire wildfire in california, usa, *Remote Sensing*, 12, 4, 623, 2020.
- [322] F. Ngadze, K. S. Mpakairi, B. Kavhu, H. Ndaimani and M. S. Maremba, Exploring the utility of sentinel-2 msi and landsat 8 oli in burned area mapping for a heterogenous savannah landscape, *PLoS One*, 15, 5, e0232962, 2020.
- [323] Q. Li, J. Tian and Q. Tian, Deep learning application for crop classification via multi-temporal remote sensing images, *Agriculture*, 13, 4, 906, 2023.
- [324] O. Sefrin, F. M. Riese and S. Keller, Deep learning for land cover change detection, *Remote Sensing*, 13, 1, 78, 2020.
- [325] U. Khusni, H. I. Dewangkoro and A. M. Arymurthy, Urban area change detection with combining cnn and rnn from sentinel-2 multispectral remote sensing data, *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*, 171–175, IEEE, 2020.
- [326] M. Rußwurm and M. Körner, Multi-temporal land cover classification with sequential recurrent encoders, *ISPRS International Journal of Geo-Information*, 7, 4, 129, 2018.
- [327] N. Di Mauro, A. Vergari, T. M. A. Basile, F. G. Ventola, F. Esposito et al., End-to-end learning of deep spatio-temporal representations for satellite image time series classification., *DC@ PKDD/ECML*, 2017.
- [328] C. Pelletier, G. I. Webb and F. Petitjean, Temporal convolutional neural network for the classification of satellite image time series, *Remote Sensing*, 11, 5, 523, 2019.
- [329] M. A. Belenguer-Plomer, M. A. Tanase, E. Chuvieco and F. Bovolo, Cnn-based burned area mapping using radar and optical data, *Remote Sensing of Environment*, 260, 112468, 2021.
- [330] Y. Ban, P. Zhang, A. Nascetti, A. R. Bevington and M. A. Wulder, Near real-time wildfire progression monitoring with sentinel-1 sar time series and deep learning, *Scientific reports*, 10, 1, 1322, 2020.
- [331] M. M. Pinto, R. Libonati, R. M. Trigo, I. F. Trigo and C. C. DaCamara, A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images, *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 260–274, 2020.
- [332] S. T. Seydi, M. Hasanlou and J. Chanussot, Burnt-net: Wildfire burned area mapping with single post-fire sentinel-2 data and deep learning morphological neural network, *Ecological Indicators*, 140, 108999, 2022.
- [333] D. Lee, S. Son, J. Bae, S. Park, J. Seo, D. Seo, Y. Lee and J. Kim, Single-temporal sentinel-2 for analyzing burned area detection methods: A study of 14 cases in republic of korea considering land cover, *Remote Sensing*, 16, 5, 2024.
- [334] M. Šebela, Burned Area Detection Script for Sentinel-2, https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/burned_area_ms/, 1 2024, last modified: January 23, 2024. Accessed: 19 October 2024.

- [335] E. Union, Copernicus land monitoring service information, 2018, accessed: 2024-10-29.
- [336] Sentinel Hub, CORINE Land Cover README, 2024, accessed: 2024-11-06.
- [337] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li and X. Li, Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band, *Remote Sensing*, 8, 4, 2016.

Curriculum Vitae

Antonia Ivanda

Antonia Ivanda (née Senta) was born on July 20, 1995, in Split, Croatia. After completing the Mathematics and Science High School in Split, in 2014 she enrolled in the undergraduate program in Computing at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB) in Split. She obtained her Bachelor's degree in Computing in 2017 and her Master's degree in Computing in 2019. In the same year, she enrolled in the postgraduate (PhD) program in Electrical Engineering and Information Technology in the field of Computing at FESB, where she also began her professional career as a Teaching Assistant at the Department of Modeling and Intelligent Systems. Since then, she has been actively involved in scientific research and teaching activities, conducting laboratory exercises in the following courses: Digital Image Forensics Analysis, Parallel Programming, Artificial Intelligence, Introduction to Distributed Information Systems, and Introduction to Programming.

Her research focuses on the application of artificial intelligence methods in remote sensing data analysis. She has authored and co-authored 23 publications between 2021 and 2024, including 8 journal papers and 15 conference proceedings papers. During her doctoral studies, Antonia participated in the *10th Advanced training course on land remote sensing: EO and AI for forestry* held in Ljubljana, Slovenia in 2021. She was also involved in teaching activities within the CEEPUS Summer School in the field of *Internet of Things and Teleinformatics*. Additionally, she actively participated in various international conferences, notably the *9th International Conference on Forest Fire Research* held in Coimbra, Portugal, and the *Joint Urban Remote Sensing Event* held in Heraklion, Crete, Greece.

Životopis

Antonia Ivanda

Antonia Ivanda (rođ. Senta) rođena je 20. srpnja 1995. u Splitu. Nakon završetka Prirodoslovno-matematičke gimnazije u Splitu, 2014. godine upisuje prijediplomski studij Računarstva na Fakultetu elektrotehnike, strojarstva i brodogradnje (FESB) u Splitu. Zvanje prvostupnice inženjerke računarstva stekla je 2017. godine, a magistre inženjerke računarstva 2019. godine. Iste godine upisuje poslijediplomski studij Elektrotehnike i informacijske tehnologije u polju računarstva na FESB-u, gdje ujedno započinje i svoj profesionalni put kao asistentica pri Katedri za modeliranje i inteligentne sustave. Od tada aktivno sudjeluje u znanstvenom i nastavnom procesu izvođenjem vježbi iz predmeta: Forenzička analiza digitalne slike, Paralelno programiranje, Umjetna inteligencija, Uvod u distribuirane informacijske sustave i Uvod u programiranje.

Njezino područje istraživanja usmjereno je na primjenu metoda umjetne inteligencije u analizi podataka dobivenih daljinskim istraživanjem. Autorica je i koautorica ukupno 23 publikacije u razdoblju od 2021. do 2024. godine, od čega je 8 radova objavljeno kao prilog u časopisima, a 15 kao prilog sa skupa (u zborniku). Tijekom doktorskog studija Antonia je sudjelovala na naprednom tečaju *10th Advanced training course on land remote sensing: EO and AI for forestry* održanom 2021. godine u Ljubljani, Sloveniji. Također je bila uključena u nastavne aktivnosti u okviru CEEPUS ljetne škole u području *Internet of Things and Teleinformatics*. Osim toga, aktivno je sudjelovala na raznim međunarodnim konferencijama, od čega bi posebno istaknula *9th International Conference on Forest Fire Research* održanoj u Coimbri, Portugal i *Joint Urban Remote Sensing Event* održanoj u Heraklionu na Kreti, Grčka.