

SATELLITE-DERIVED BATHYMETRY USING MACHINE LEARNING METHODS

CHIFUNIRO NGALANDE

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF EARTH AND SPACE SCIENCE
GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE

YORK UNIVERSITY
TORONTO, ONTARIO

APRIL 2025

© CHIFUNIRO NGALANDE, 2025

Abstract

This study evaluates the performance of machine and deep learning (ML/DL) models for satellite-derived bathymetry (SDB) compared to traditional empirical and physics-based methods in high-latitude freshwater and saltwater environments. The models were trained on data from several Canadian sites, and their transferability was tested on unseen sites. DL models, U-Net, SegNet and DeepLabv3 +, achieved about two times higher F1 scores than traditional methods. From the class-wise scores, ML/DL models can reliably predict up to a depth of 8m in freshwater, due to greater water transparency, and up to 3m in saltwater, due to sediments that prevent light penetration. Traditional methods struggled in deeper and turbid waters, but performed similarly to the machine learning Random Forest model in both environments. Overall, the ML/DL models have difficulties generalizing to unseen data or data from geographic locations not included in the training process, particularly in saltwater due to poor light penetration. The depth results met the Zones of Confidence (ZOC) categories, CATZOC C and CATZOC D, of the International Hydrographic Organization (IHO) depth accuracy standards. Overall, this study highlights the advantages of machine learning over traditional methods while identifying challenges in model generalization and data diversity.

Acknowledgments

I would like to express my sincere gratitude to the following individuals who have supported me over the past two years as I completed this thesis.

Firstly, I thank my supervisor, Prof. Costas Armenakis, for his guidance, encouragement, and genuine faith in me. Through his expertise and mentorship, I had the opportunity to grow both academically and professionally. His support extended beyond academics to my personal well-being, and I appreciate his patience, especially when the project faced numerous challenges.

I also thank my supervisory committee member, Prof. Baoxin Hu, for her valuable insights and contributions to my research. I am grateful to my examination committee members, Prof. Gunho Sohn and Prof. Amaneh Jadidi Mardkheh, for their thoughtful assessment and constructive suggestions that enriched this work.

I am grateful to my colleagues in the Earth and Space Science department, Zahra Arjmandi, Evangelos Bousias, and Neda Kolahi, for their support and guidance over the past two years. Special thanks to the Graduate Program Assistant, Marcia Gaynor, for her consistent help and genuine care in resolving every issue I faced.

I am deeply grateful to my family for their love and encouragement. I thank my dad, Dr. Ellias Ngalande, my mom, Mrs Alice Ngalande, Fatsani, Thokozani, Thandizo, Lindani, Lonjezo, and Kella, for their persistent prayers and endless support. I am blessed with a loving family.

I am eternally grateful to my extended family in Canada, the Open Arms Seventh Day Baptist church, for their prayers, support, and for treating me as family. I thank them for being a true blessing in my life. Their love and hospitality made the past two years truly memorable.

Finally, I give all glory and thanks to my Father in Heaven who made all this possible through His great power and wisdom. I am grateful that His love sustained me and that *He brought me out into a broad place; He delivered me because He delighted in me*. To God be the glory!

Contents

	Page
Abstract	ii
Acknowledgments	iii
Contents	iv
List of Figures	viii
List of Tables	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Research Objectives	5
1.3 Contributions	6
1.4 Thesis Structure	7
2 Fundamentals	8
2.1 Spectral reflectance properties of water	8
2.2 The radiative transfer in water	10
2.3 Physical properties of water	12
2.3.1 Inherent Optical Properties	13
2.3.2 Apparent Optical Properties	15
2.3.3 Boundary Conditions	15

2.4	Physics-based Modelling	16
2.5	Radiative Transfer Model	17
2.5.1	Assumptions	17
2.5.2	RTM Equation	18
2.6	Empirical Modelling	18
2.6.1	Log-Linear Transform	19
2.6.2	Band Ratio Transform	19
2.7	Machine Learning	20
2.7.1	Random Forest Classifier	22
2.8	Deep Learning	23
2.8.1	Encoder-Decoder CNN architecture	24
2.8.2	ResNet-50	27
2.8.3	U-Net	30
2.8.4	SegNet	32
2.8.5	DeepLabv3+	34
3	Related Works	38
3.1	Empirical algorithm studies	38
3.1.1	Summary	44
3.2	Physics-based studies	44
3.2.1	Summary	48
3.3	Machine Learning models: Random Forest	49
3.3.1	Random Forest without Optimization	49
3.3.2	Hyperparameter Tuning	52
3.3.3	Selection of Predictors	54
3.3.4	Combined Optimization	56
3.3.5	Summary	58
3.4	Deep Learning models: DCNNs	59

3.4.1	Semantic Segmentation	59
3.4.2	Multi-channel Training	61
3.4.3	Model Transferability	64
3.4.4	Summary	65
4	Methodology	66
4.1	Study Areas	68
4.2	Datasets	72
4.3	Data preprocessing	73
4.3.1	Satellite Image preprocessing	73
4.3.2	Bathymetric data preprocessing	74
4.4	Image data annotation	74
4.4.1	Image data labelling for RF model	75
4.4.2	Image data labelling for DCNNs	76
4.5	Model Training	77
4.5.1	Random Forest Training	78
4.5.2	DCNNs Training	80
4.6	Model Transferability	83
4.7	Comparison with deterministic models	84
5	Experiments and Results	85
5.1	Random Forest classification results	85
5.1.1	Saltwater Random Forest model	86
5.1.2	Freshwater Random Forest model	91
5.2	DCNNs semantic segmentation results	95
5.2.1	U-Net Results	95
5.2.2	SegNet Results	104
5.2.3	DeepLabv3+ Results	114

5.3	Comparison with Empirical and Physics-based models	124
5.4	Performance Comparison	127
5.4.1	Training and Validation	127
5.4.2	Testing	128
5.5	Impact of fresh and saltwater conditions on SDB modelling	130
6	Conclusions and Future Work	132
6.1	Future Work	135
	References	137
	Appendices	144
	Appendix A: International Hydrographic Organization standards	144
	Appendix B: Random Forest Hyperparameters	146
B.1	<i>Mtry</i> Parameter	146
B.2	<i>Ntree</i> Parameter	147
B.3	<i>Kn</i> Parameter	148
	Appendix C: Satellite Image Data	149

List of Figures

1.1	Graphic representation of SDB	4
2.1	Light penetration in open ocean and coastal waters	9
2.2	Bidirectional reflectance	10
2.3	Optical properties of water	13
2.4	Radiative Transfer Model	16
2.5	Classification by the Random Forest model	21
2.6	A simple convolutional neural network	24
2.7	A basic encoder-decoder architecture	25
2.8	ResNet-50 architecture	29
2.9	U-Net architecture	30
2.10	Overlap-tile strategy of U-Net	31
2.11	SegNet architecture	32
2.12	SegNet max-pooling indices	33
2.13	Feature extraction from a 1D input. (a) Sparse feature extraction with ordinary convolution. (b) Dense feature extraction with atrous convolution with sampling rate $r = 2$	34
2.14	Atrous Spatial Pyramid Pooling	35
2.15	DeepLabv3+ architecture	36
4.1	Methodology	67
4.2	Freshwater test site in Lake Ontario	70

4.3	Combination (a) and pure saltwater (b) test sites	71
4.4	Labelled mask of Point Clark, Lake Huron for training	76
4.5	Confusion Matrix	80
5.1	Saltwater RF model confusion matrix	87
5.2	Saltwater RF model feature importance	88
5.3	Random Forest prediction of Graham Island	89
5.4	Random Forest prediction of the Rimouski Port	90
5.5	Freshwater RF model confusion matrix	92
5.6	Freshwater RF model feature importance	93
5.7	Random Forest prediction of Athol	94
5.8	U-Net saltwater loss curves	95
5.9	U-Net saltwater validation predictions	97
5.10	Graham Island U-Net predictions	98
5.11	Rimouski U-Net predictions	100
5.12	U-Net freshwater loss curves	101
5.13	U-Net freshwater validation predictions	102
5.14	Athol U-Net predictions	103
5.15	SegNet saltwater loss curves	105
5.16	SegNet saltwater validation predictions	106
5.17	Graham Island SegNet predictions	107
5.18	Rimouski SegNet predictions	109
5.19	SegNet freshwater loss curves predictions	110
5.20	SegNet freshwater validation predictions	112
5.21	Athol SegNet predictions	113
5.22	DeepLabv3+ saltwater loss curves	114
5.23	DeepLabv3+ saltwater validation predictions	116
5.24	Graham Island DeepLabv3+ predictions	117

5.25	Rimouski DeepLabv3+ predictions	119
5.26	DeepLabv3+ freshwater loss curves predictions	120
5.27	DeepLabv3+ freshwater validation predictions	121
5.28	Athol SegNet predictions	122
5.29	Validation of SDB using deterministic models in Graham Island	125
5.30	Validation of SDB using deterministic models in Rimouski	125
5.31	Validation of SDB using deterministic models in Athol	126
5.32	Machine learning model validation performance in fresh and saltwater	128
5.33	SDB model test performance comparison using F1-score	129

List of Tables

3.1	Different band combinations as predictors	54
3.2	Modified predictors	57
3.3	Optimizer specifications	60
3.4	Sentinel-2 Band combinations	61
4.1	Study Areas	68
4.2	Study Areas for training and testing model transferability	71
4.3	Sentinel 2-A multispectral bands	73
4.4	Depth classes	75
4.5	Distribution of class features	77
4.6	Features included in the Random Forest as predictors	78
4.7	Class distribution of points in fresh and salt water tabular datasets	79
4.8	Optimal Random Forest hyperparameters for salt and freshwater	79
4.9	Optimal hyperparameters for DCNNs in fresh and salt water	81
5.1	Saltwater RF classification report	86
5.2	Classification report of Graham Island	89
5.3	Classification report of the Rimouski Port	90
5.4	Freshwater RF classification report	91
5.5	Classification report of Athol	94
5.6	Best training and validation epoch scores of U-Net saltwater model	96
5.7	U-Net saltwater class-wise validation scores	97

5.8	Graham Island U-Net prediction scores	99
5.9	Rimouski U-Net prediction scores	100
5.10	Best training and validation epoch scores of U-Net freshwater model	101
5.11	U-Net freshwater class-wise validation scores	102
5.12	Athol U-Net prediction scores	103
5.13	Best training and validation epoch scores of SegNet saltwater	105
5.14	SegNet saltwater class-wise validation scores	106
5.15	Graham Island SegNet prediction scores	108
5.16	Rimouski SegNet prediction scores	110
5.17	Best training and validation epoch scores of SegNet freshwater	111
5.18	SegNet freshwater class-wise validation scores	112
5.19	Athol SegNet prediction scores	113
5.20	Best training and validation epoch scores of DeepLabv3+ saltwater	115
5.21	DeepLabv3+ saltwater class-wise validation scores	115
5.22	Graham Island DeepLabv3+ prediction scores	118
5.23	Rimouski DeepLabv3+ prediction scores	119
5.24	Best training and validation epoch scores of DeepLabv3+ freshwater	120
5.25	DeepLabv3+ freshwater class-wise validation scores	122
5.26	Athol DeepLabv3+ prediction scores	123
5.27	Empirical Model F1 scores	124
5.28	Physics-based model F1 scores	124
A.1	IHO ZOC Categories	144
C.1	Saltwater Study Areas	149
C.2	Freshwater Study Areas	149

List of Abbreviations

AI	Artificial Intelligence
AOP	Apparent Optical Property
ANN	Artificial Neural Network
C2RCC	Case 2 Regional Coast Colour
CD	Chart Datum
CE	Cross Entropy
CHS	Canadian Hydrographic Service
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
ESA	European Space Agency
FL	Focal Loss
FN	False Negative
FNN	Feedforward Neural Network
FP	False Positive
IOP	Inherent Optical Property

MAE	Mean Absolute Error
ML	Machine Learning
MSL	Mean Sea Level
NIR	Near Infrared
NN	Neural Network
RF	Random Forest
RMSE	Root Mean Square Error
R_{rs}	Remote-Sensing Reflectance
RTM	Radiative Transfer Model
RTE	Radiative Transfer Equation
SDB	Satellite-Derived Bathymetry
SNAP	Sentinel Applications Platform
SGD	Stochastic Gradient Descent
TOA	Top of Atmosphere
TN	True Negative
TP	True Positive
UTM	Universal Transverse Mercator
WASI	Water Colour Simulator
WGS	World Geodetic System

Chapter 1

Introduction

1.1 Background and Motivation

Coastal zones are unique and significant geographical regions. Though they occupy a relatively small portion of the Earth's land surface, they are dynamic and ecologically rich systems that include a wide range of marine and coastal-terrestrial habitats, from intertidal zones to deep water ecosystems. According to Reimann et al. (2023), as of 2020, nearly 2.15 billion people live in the near-coastal zone, defined as land within 100km of the coast and an elevation of 100m, and there is much economic activity in this region, such as commercial fishing, tourism, pipeline construction, and marine transportation. Marine transportation accounts for 90% of global trade and Selamoglu (2021) states that more than 1.8 million principal marine vessels were registered in 2016, including bulk cargo carriers, oil tankers, container ships, passenger ships and ferries. With diverse activity occurring in coastal waters and the transportation of sediments along the coast, the bathymetry of these environments changes frequently (Yunus, et al., 2019). For this reason, accurate and high spatial resolution bathymetric surveys of coastal shallow waters are crucial to ensure safe marine navigation and the appropriate management of coastal ecosystems (Hamylton, et al., 2015).

Traditionally, water depth has been measured using laborious and expensive methods, such as

sounding technology like the multibeam echosounder (MBES) onboard marine vessels (Lyzenga, 1985). MBES produces high-resolution measurements and bathymetric maps that accurately represent the seafloor and is the preferred method for carrying out bathymetric surveys. However, the traditional method of using sounding techniques is limited in spatial and temporal coverage as marine vessels cannot manoeuvre in extremely shallow waters and access remote and protected locations (Yunus, et al., 2019). In addition, the operation of such equipment is costly and time-consuming to employ frequently. To overcome the constraints of using MBES, Light Detecting and Ranging (LiDAR) technology onboard an aircraft, which uses active remote sensing techniques, has been used to measure the water depth of extremely shallow and remote locations. Airborne LiDAR Bathymetry (ALB) uses a sensor to emit pulses of light that can penetrate the water column and reflect off the seafloor. The return signal is recorded, and the water depth is retrieved. Surveying coastal areas using LiDAR is expensive as the aircraft can only cover a small area at a time (Gholamalifard, et al., 2013) due to the maximum altitude at which the laser scanner can operate. Running airborne campaigns also requires a license (Traganos, et al., 2018). With the traditional methods' spatial, temporal, and logistical constraints, there are many data gaps in the bathymetric measurements of coastal zones. Only a small fraction of the coastal zones of the Earth have been surveyed (Yunus, et al., 2019), and traditional bathymetric surveying methods cannot produce the information required fast enough. Therefore, there is a need for extensive and real-time measurements of water depth to fill in the gaps in the existing bathymetric maps of Canadian coastal zones to ensure safe marine navigation and appropriate coastal management.

Consequently, there has been a shift from these expensive and time-consuming methods to affordable techniques to measure water depth in remote locations and fill in the gaps of existing bathymetric maps. Passive remote sensing techniques using multispectral scanners onboard satellite platforms offer a quick, flexible, and affordable method for bathymetric surveying called satellite-derived bathymetry (Caballero & Stumpf, 2019). Satellite-derived bathymetry (SDB) modelling is a relatively novel technique for the bathymetric surveying of shallow coastal zones. It involves pas-

sive remote sensing techniques to estimate water depth from satellite imagery. In Satellite-derived Bathymetry, high-resolution satellite imagery is used to estimate water depth. The bathymetric derivation is based on capturing the water leaving reflectance by the satellite sensor and then transforming it to water depth; that is, the water depth is determined as a function of the multispectral reflectance. The captured reflectance not only depends on the reflectance from the bottom surface but also on the absorption and scattering properties of any suspended material in the water column and the atmospheric effects on the water reflectance values of the images. Passive systems are appealing as high-resolution satellite imagery is readily available and has global coverage (Lyzenga, 1985), making it ideal for shallow water mapping (Gholamalifard, et al., 2013). Although atmospheric influences and water clarity inhibit passive remote sensing techniques, the short revisit time of satellites over the exact geographical location provides many images and data to select from for optimal water depth prediction (Caballero & Stumpf, 2019).

Therefore, research on satellite-derived bathymetry (SDB), depicted in Figure 1.1, has increased, showing promising results in clear, tropical and subtropical waters. The relationship between the intensity of light reflected and emitted from the water's surface and in-situ water depth data is modelled using empirical, analytical, or semi-analytical approaches, which are deterministic models. Deterministic models are mathematical models that have fixed and estimated parameters. Most studies on SDB use two empirical algorithms, Stumpf's Band Ratio transform algorithm and Lyzenga's Log-Linear transform algorithm, to calibrate the relationship between the water-leaving reflectance and water depth (Stumpf, et al., 2003). However, empirical algorithms assume that the physical properties of water are uniform across an entire study area (Ilori & Knudby, 2020) and hence need to be re-calibrated for a different location. Therefore, semi-analytical models, such as the Radiative Transfer Model (RTM), a physics-based model that considers the physical properties of water and simulates the behaviour of light in the water column, make better predictions of depth (Mobley, 2001). However, the physics-based model reduces the complex relationship of reflectance and water depth to a mathematical approximation. Because of this, accurate estimates of the phys-

ical properties of water and precise measurements of radiance, which is observed by the satellite sensor, are required. Studies by Yunus et al. (2019) and Stumpf et al. (2003) have shown that SDB using empirical algorithms can predict water depths up to 10m in coastal zones and 30m in inland lakes and rivers and prove that it is the most feasible method to map extensive and inaccessible areas and satisfy the demand for real-time measurements of water depth. Nevertheless, empirical and physics-based models of predicting water depth are not as accurate as the traditional methods due to the assumptions and mathematical approximations that are made (Caballero & Stumpf, 2019).

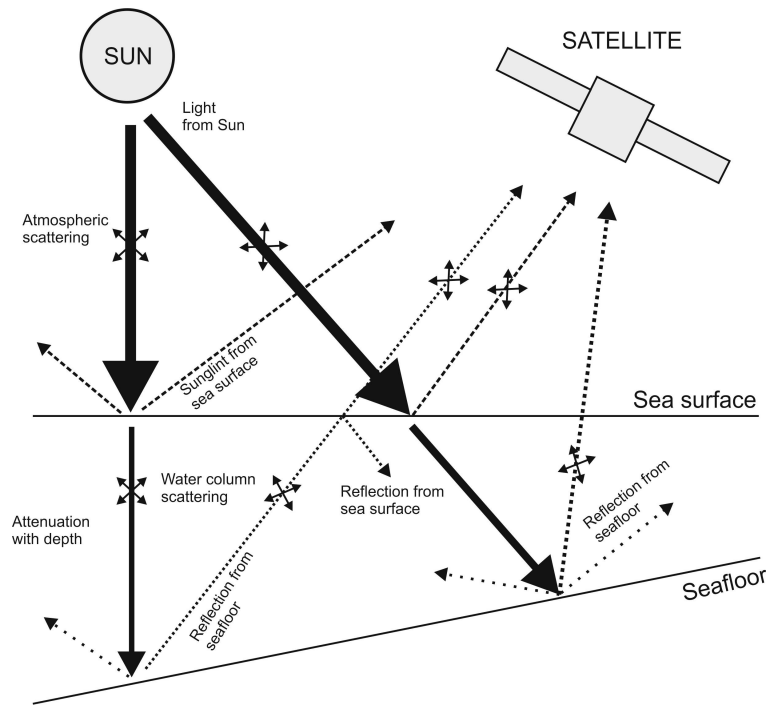


Figure 1.1: Graphic representation of SDB

Source: Westley (2021). Satellite-derived bathymetry for maritime archaeology: Testing its effectiveness at two ancient harbours in the Eastern Mediterranean. *Journal of Archaeological Science: Reports*. 38, 103030.

For that reason, data-driven approaches, such as machine and deep learning approaches, are being explored, as they can learn the relationship between spectral data in satellite imagery and water depth to make accurate and automated water depth predictions. This is achieved using computer vision techniques and regarding the prediction task as a classification problem, as illustrated in recent studies by Wilson et al. (2020). According to our literature research, there is ongoing research in SDB using data-driven models, especially in tropical and subtropical regions. However, there is limited work in the high-latitude areas beyond a latitude of 35° in different bathymetric environ-

ments. Although the water depth in Canadian coastal zones has been predicted using deterministic models in studies by Ilori (2020) and Yunus et al. (2019), there is limited research using machine learning models in SDB. Therefore, the motivation of this project is to investigate the performance of several machine learning models in predicting water depth in Canadian high-latitude salt and freshwater conditions, and compare their accuracy to the current empirical and physics-based models. Saltwater is characterized by high concentrations of salt and dissolved ions and sediments, which increase its viscosity and hence its resistance to internal flow. Talley et al. (2011) state that due to the mass of dissolved salts, saltwater is denser than freshwater (Talley, et al., 2011). In that regard, freshwater floats above saltwater in estuaries as it is less dense. Freshwater has high concentrations of organic matter, such as phytoplankton, which absorb and scatter light. However, the lower concentration of salt and sediments makes it less dense and reduces its internal resistance to flow, allowing more light to penetrate the water column (Talley, et al., 2011). By investigating the performance of machine learning models in saltwater and freshwater, the impacts of salinity on SDB can be assessed.

1.2 Research Objectives

This investigation aims to examine the use of data-driven approaches like machine and deep learning models to predict water depth in shallow coastal zones not exceeding 10m of depth from satellite imagery and compare their accuracy against the current empirical and semi-analytical SDB modelling techniques. This comparison will identify the most transferable model that requires little computational costs and produces accurate water depth predictions in different conditions. This will enable frequent measurements of coastal regions, which can aid many applications. The objectives of the investigation are to:

1. Assess the performance of ML and DL models and investigate the applicability of a generalized approach to build a transferable model.
2. Identify the spectral data from satellite imagery useful for depth prediction using the Random

Forest feature selection measure.

3. Identify the range of depths within which all SDB approaches are reliable.
4. Compare the accuracy of data-driven models to the current empirical and physics-based models.
5. Investigate and determine the impact of different bathymetric conditions in high-latitude regions on the retrieval of depth.
6. Determine if data-driven approaches meet the International Hydrographic Organization (IHO) accuracy standards.

The International Hydrographic Organization (IHO) standards, in Appendix A, provide a guideline for the accuracy expected from this investigation. As the IHO has assigned SDB in CATZOC C and CATZOC D, this means that SDB cannot replace the traditional bathymetric survey methods, but is sufficient to fill in the gaps in existing bathymetric maps and predict depth in highly remote areas that are not frequently navigated.

1.3 Contributions

Much SDB research has been conducted using deterministic approaches, which are adequate in estimating water depth to a specific depth range with an accuracy acceptable by IHO standards. However, based on our literature review, there is limited research on using data-driven models for satellite-derived bathymetry in high-latitude regions. Therefore, through this study, we aim to contribute to SDB in the following ways.

1. Contributing to the growing research in SDB modelling using machine learning methods, especially in high-latitude waters.
2. Comparing the performance of various SDB models in different salinity conditions and in high-latitude waters. This comparison will highlight the bathymetric conditions in which de-

terministic or data-driven models excel or underperform to guide the choice of an appropriate model.

3. Investigating the applicability of knowledge transfer ML/DL models for accurate predictions for different water conditions.

1.4 Thesis Structure

This thesis is organized into six chapters. The first is the introduction of the background, problem, and motivation of this study and the description of the objectives of this research work. The fundamental concepts of the radiative transfer of light in water and the deep neural network architectures that will be used in the study are provided in Chapter 2. Chapter 3 includes a review of related research in SDB using various modelling techniques. Chapter 4 presents the methods used in the study, while Chapter 5 presents the experiments, results, and the comparison to deterministic models. Finally, Chapter 6 includes the conclusions drawn from the study, plus areas that should be considered in future works.

Chapter 2

Fundamentals

This chapter presents some fundamental concepts and approaches related to SDB modelling.

2.1 Spectral reflectance properties of water

Optical remote sensing has provided means to derive valuable information about the Earth's surface through measured reflectance. The directional spectral reflectance from a surface depends on the angular distribution of the incident radiant energy and its interactions with the surface (Verhoef, 1998). These interactions result from the surface's properties (Mobley, 2022) and hence are a matter of interest to study and understand.

Modelling the reflectance from a surface enables the determination of properties that contribute to the directional reflectance. Verhoef (1998) states this can be done through forward physical models, which mathematically represent the properties of the surface that contribute to the directional spectral reflectance. The Radiative Transfer Model (RTM) is a forward physical model that simulates the propagation of light energy through a medium or surface (Mobley, 2001). The water surface absorbs almost all radiation from the sun and reflects a small portion into the atmosphere to be measured by remote sensing technology (Mobley, 2001). It is unique in the way it interacts with solar radiation, and these interactions are described mathematically by the RTM.

According to Webb (2021), water absorbs incident light from the sun, which causes the amount of light that penetrates through the water column to decline with depth. In addition, suspended organic and inorganic matter in water decreases the penetrative power of light. The rate of decline for the blue, green, and red wavelengths is different in open ocean and coastal waters, as the former has lower concentrations of suspended matter. At the same time, the latter is characterised by sediments from runoff and microscopic organisms such as phytoplankton. Light of extremely short and long wavelengths, blue and red, is absorbed and scattered faster by water molecules and organic matter than light of middle wavelengths (Webb, 2021), which explains the blue-green colour observed in shallow coastal waters. Figure (2.1) shows light penetration in an open ocean and in coastal waters.

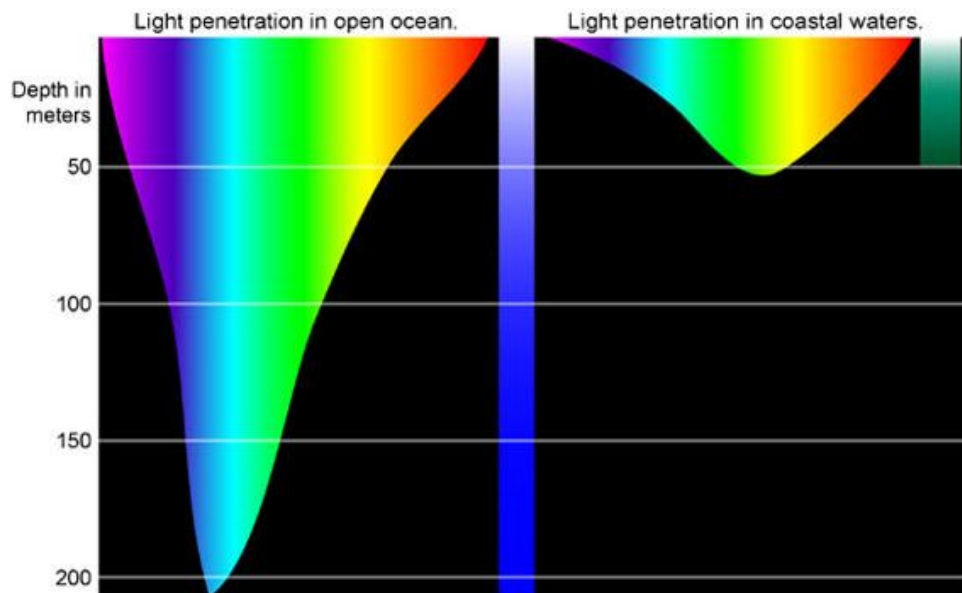


Figure 2.1: Light penetration in open ocean and coastal waters
Source: Webb, P. (2021). *Introduction to Oceanography*, chapter Light. Roger Williams University.

The rate of decline in light intensity is greater in coastal waters than in the open ocean, with the green wavelength interval being the most penetrative, reaching a depth of 50m in optimal conditions. In comparison, blue light can reach a depth of 200m in the open ocean (Webb, 2021). This decline and absorption of light intensity in coastal waters means that very little light reaches the bottom of the seafloor to be reflected and detected by satellite sensors. Therefore, for purposes of this thesis, the depth range was restricted to 10m as the study by Duan et al. (2022) revealed that

beyond 15m, prediction errors are greater.

2.2 The radiative transfer in water

Water absorbs electromagnetic radiation (EMR) with wavelengths below 250nm and above 700nm (Mobley, 2001) and reflects EMR in the visible region of the spectrum, from 500nm to 600nm (Sayer, 2007), as this energy can penetrate through the water column and is scattered by the constituents in the water. With this knowledge, water can be distinguished from other surface types through optical remote sensing techniques and radiation measurements, as it has a unique spectral signature.

The spectral reflectance from the water surface is directional as it depends upon the zenith and azimuth angles of the sun and the sensor (Sayer, 2007). The direction in which the radiance is reflected from the surface is a function of the direction of the incident radiance (Mobley, 2022).

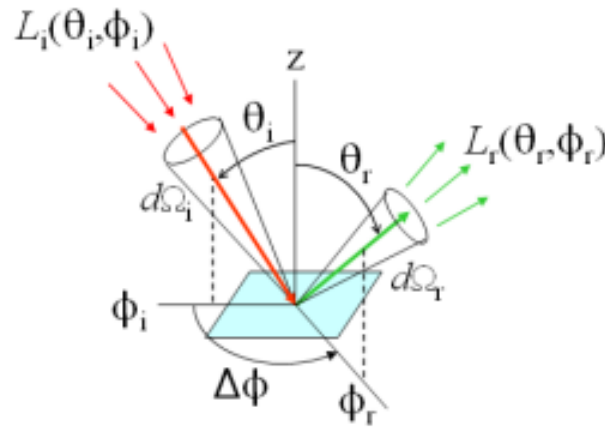


Figure 2.2: Bidirectional reflectance

Source: Mobley, C. D. (2001). Radiative transfer in the ocean. *Encyclopedia of ocean sciences*, 4:2321–2330.

Figure (2.2) shows the geometry of the directional reflectance from water, with the incident radiant energy, $L_i(\theta_i, \phi_i)$, with the solid angle Ω_i centred in the direction of (θ_i, ϕ_i) , where θ_i is the solar zenith angle, and ϕ_i is the solar azimuth angle and the reflected radiance $L_r(\theta_r, \phi_r)$ along the solid angle Ω_r centred in the direction of (θ_r, ϕ_r) , where θ_r is the viewing zenith angle and ϕ_r is

the viewing azimuth angle of the sensor (Mobley, 2022). However, the measured radiance from the water surface is not only a result of surface scattering, as shown in Figure (2.2), but of volume scattering and bottom reflectance as well (Mobley, 2022). As energy penetrates the water column, the water elements backscatter photons in the light beam (Sayer, 2007). This process, known as volume scattering, allows the radiance from the water subsurface to be measured (Hedley, 2008). Therefore, the radiance from the water, shown in Figure (2.4) from Mobley (2022), is the sum of the radiance reflected from the surface ($L_r(\theta, \phi, \lambda)$), the upwelling radiation from the water column ($L_w(\theta, \phi, \lambda)$), and the radiance that is reflected from the bottom of the water column ($L_u(z, \theta, \phi, \lambda)$) (Lee & Liou, 2007). Figure (2.4) in section 2.4 also illustrates that the incident radiant energy at the water surface is a combination of the direct radiation from the sun, and the diffuse downwelling radiation $E_d(\lambda)$ and that $L_d(z, \theta, \phi, \lambda)$ is the downwelling radiance that is transmitted through the water column (Mobley, 2022).

Mobley (2001) employs the Radiative Transfer Model, which is a forward physical model that mathematically represents the interactions between solar radiation and water. This is the most common model used by researchers to simulate the optical interactions in water. Hedley (2008) and Cheng and Liang (2011) explain that the RTM portrays the exchange of radiant power between the radiation beam and a surface, through the physical processes of absorption, scattering, and emission (Cheng & Liang, 2011). The output of the radiative transfer equation (RTE) is the radiance distribution, $L(z, \theta, \phi, \lambda)$, which is a function of depth (z), solar zenith (θ) and azimuth angle (ϕ), and the wavelength of the propagating light beam (λ). It explains the directional and spectral structure of the light beam and is calculated using the optical properties and boundary conditions of water, which is then used to calculate the radiance at the water surface (Mobley, 2001).

To understand the directional reflectance properties of water, Mobley (2001) studies the physical properties of water and the interactions between radiant energy and water as they relate to how incident radiance from any direction is reflected into any direction (Mobley, 2022). These interactions

are defined as the optical properties of water.

2.3 Physical properties of water

The fraction of incident radiant energy reflected from the water surface, which is then detected by an optical sensor, depends upon the depth and composition of the water column.

The radiant power propagating through the water decreases exponentially with depth due to the attenuation by water constituents (Mobley, 2001). These two physical properties allow Mobley (2001) to classify water bodies into two categories. Case 1 water bodies refer to clear or eutrophic waters with a high concentration of phytoplankton, which are microscopic organisms that contain chlorophyll and absorb light photons (Mobley, 2001). Case 2 water bodies are usually shallow, coastal waters or estuaries that have a lower concentration of phytoplankton, but a high concentration of colour dissolved organic matter (CDOMs) and inorganic particles, which scatter and absorb photons (Mobley, 2001). Measuring reflectance from this water body is complex as blue light, which is the most penetrative in water, is absorbed by CDOMs (Mobley, 2001). Furthermore, organic and inorganic matter vary spatially and temporally (Mobley, 2001), as opposed to phytoplankton, which cause the optical properties of Case 2 water bodies to change frequently.

By classifying water into two categories by physical properties, Mobley (2001) shows that different water bodies shall have different optical properties, hence varying directional spectral reflectance properties. Depth and composition of the water column affect the directional spectral reflectance of water. This reflectance detected by the sensor includes the surface reflected radiance and the upwelling radiance through the water surface (Lee & Liou, 2007), and is influenced by the optical properties shown in Figure (2.3). To understand the RTM and the inputs that are required, Mobley (2001) differentiates the optical properties of water as inherent optical properties (IOPs) and apparent optical properties (AOPs).

2.3.1 Inherent Optical Properties

Figure (2.3) shows the components and inputs of RTM and how the incident radiant flux (Φ_i) is attenuated through the volume of water (ΔV) with thickness (Δr). As the incident radiant power propagates through the medium, a fraction of energy is absorbed (Φ_a) by the water constituents, another fraction is scattered ($\Phi_s(\Psi)$) at the scattering angle (Ψ) into the solid angle ($\Delta\Omega$), and the remainder (Φ_t) is transmitted through the water column (Mobley, 2001).

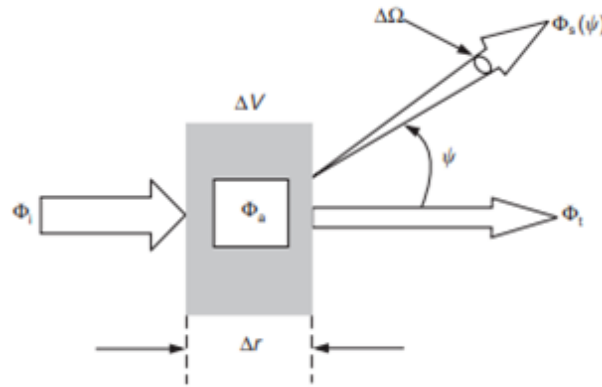


Figure 2.3: Optical properties of water

Source: Mobley, C. D. (2001). Radiative transfer in the ocean. *Encyclopedia of ocean sciences*, 4:2321–2330.

IOPs are the parameters of water that determine how light is attenuated through the water column (Mobley, 2001). These include the absorption coefficient, scattering coefficient, single-scattering albedo, phase function, and the volume scattering function (VSF). IOPs cannot be measured but are predicted from the measurable physical properties of water, such as the chlorophyll concentration, inorganic particle concentrations, depth, and the wavelength of light being propagated (Mobley, 2001). However, with the spatial and temporal variability of particulate matter concentrations (Mobley, 2001), IOPs are subject to change, which results in a different solution for the RTM with each change in IOPs.

The absorption coefficient of water, $a(\lambda)$, is the sum of all the absorption coefficients of each water element, which include phytoplankton, CDOMs, and inorganic particles (Mobley, 2001). Mobley (2001) states that predicting the absorption coefficient of water is complex, as all water constituents

absorb radiant energy in varying amounts. It is defined to be the limit of the fraction of incident radiant flux that is absorbed by the water column, as the thickness decreases (Mobley, 2001):

$$a(\lambda) \equiv \lim_{\Delta r \rightarrow 0} \frac{1}{\Phi_i(\lambda)} \frac{\Phi_a(\lambda)}{\Delta r} \quad [m^{-1}] \quad (2.1)$$

Water constituents scatter the photons in the propagating light beam at an angle. Mobley (2001) defines the spectral scattering coefficient, $b(\lambda)$, as the total scattered radiant flux per unit incident irradiance and unit volume of water:

$$b(\lambda) = \int_{4\pi} \beta(\Psi, \lambda) d\Psi = 2\pi \int_0^\pi \beta(\Psi, \lambda) \sin \Psi d\Psi \quad (2.2)$$

where $\beta(\Psi, \lambda)$, is the volume scattering function (VSF), which is the scattered intensity, $I_s(\Psi, \lambda)$, per unit incident irradiance, $E_i(\lambda)$, per unit volume of the water column and is given by (Mobley, 2001):

$$\beta(\Psi, \lambda) = \lim_{\Delta V \rightarrow 0} \frac{I_s(\Psi, \lambda)}{E_i(\lambda) \Delta V} \quad (2.3)$$

The attenuation coefficient, $c(\lambda)$, is the sum of the absorption and scattering coefficients. In contrast, the phase function, $\tilde{\beta}(\Psi, \lambda)$, is the ratio of the VSF to the spectral scattering coefficient (Verhoef, 1998). Finally, the single-scattering albedo, $\omega_o(\lambda)$, is a dimensionless ratio of the scattering coefficient to the attenuation coefficient (Mobley, 2001):

$$\omega_o = \frac{b(\lambda)}{c(\lambda)} \quad (2.4)$$

Cheng and Liang (2011) include more optical properties in their investigation involving the RTM, such as the asymmetry factor, which is the cosine of the phase function, and the refractive index of the water column. The RTM relates IOPs to the radiance distribution in water which in turn is used to obtain the directional reflectance at the surface (Mobley, 2001).

2.3.2 Apparent Optical Properties

AOPs are not inputs required in the RTM but are dependent upon both IOPs and the directional structure of the propagating light beam (Mobley, 2001). They are reliable descriptors of ocean colour as they are ratios of measurable variables, and hence do not change frequently as IOPs (Mobley, 2001). These include the irradiance ratio, $R(z, \lambda)$, which is the ratio of upwelling irradiance to the surface to the downwelling irradiance from the surface (Mobley, 2001). The remote-sensing reflectance, $R_{rs}(\theta, \phi, \lambda)$, which is the ratio of the water-leaving radiance to the downwelling irradiance and is defined in Equation (2.5) (Mobley, 2001).

$$R_{rs}(\theta, \phi, \lambda) \equiv \frac{L_w(\theta, \phi, \lambda)}{E_d(\lambda)} \quad [sr^{-1}] \quad (2.5)$$

Finally, the diffuse attenuation coefficients (K) for the radiance, downwelling and upwelling irradiance, and the photosynthetic available radiation, which describe the change in radiance with depth (Mobley, 2001).

2.3.3 Boundary Conditions

Boundary conditions are just as important as IOPs as they are inputs in the RTM and affect the spectral reflectance from the water surface. They constitute as the physical environment of the water and include surface waves, incident sky radiance, and the bottom character of the water body (Mobley, 2001) (Sayer, 2007). Surface waves are a significant boundary condition as they produce whitecaps on the surface which result in more incident radiant energy being reflected (Hedley, 2008) (Sayer, 2007). The radiance from the water surface is also a product of the radiance that is reflected from the bottom of the water column, which may either be physical or just the deepest depth (Mobley, 2022). The contribution of these boundary conditions to the radiance was not stated explicitly. Still, the study by Sayer (2007) provides the relationship between the reflectance at the water surface and the boundary conditions:

$$\rho = W\rho_{wc} + (1 - W)\rho_{gl} + (1 - \rho_{wc})\rho_{ul} \quad (2.6)$$

where ρ is the total reflectance at the surface, W is the proportion of whitecaps on the surface, ρ_{wc} is the reflectance due to whitecaps, ρ_{gl} is the reflectance due to sun glint, and ρ_{ul} is the reflectance due to underlight, which accounts for bottom reflectance (Sayer, 2007).

2.4 Physics-based Modelling

Physics-based models have also been used to predict depth by simulating the behaviour of light within the water column using semi-analytical methods. The most common physics-based model is the Radiative Transfer Model (RTM), which is further explained in section 2.5 and is shown in Figure (2.4).

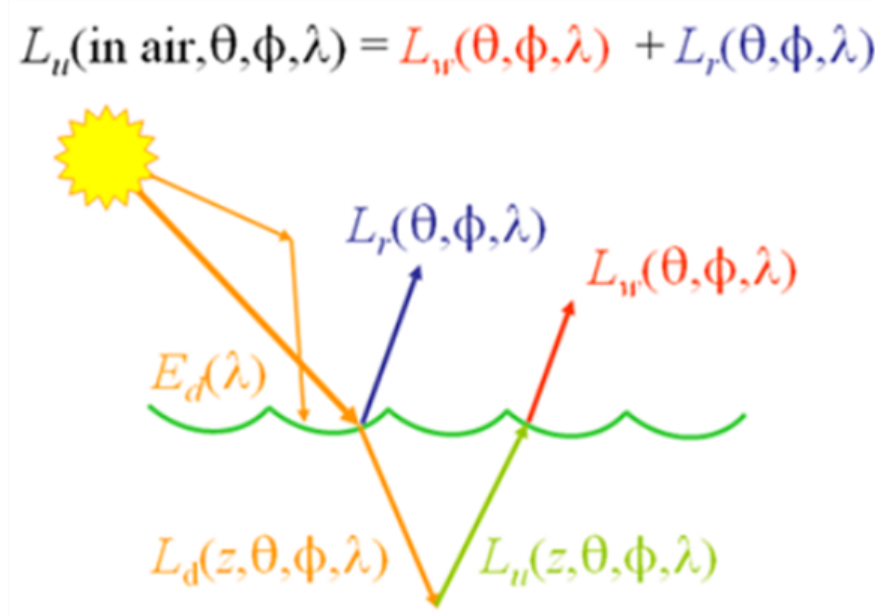


Figure 2.4: Radiative Transfer Model

Source: Mobley, C. (2022). The oceanic optics book. International Ocean Colour Coordinating Group (IOCCG)

From Figure (2.4), the water-leaving radiance, $L_w(\theta, \phi, \lambda)$, and the reflected radiance, $L_r(\theta, \phi, \lambda)$, make up the total reflectance detected by the satellite sensor and is known as the remote-sensing reflectance ($R_{rs,observed}$) (Mobley, 2022), the total radiance received at the water surface is $E_d(\lambda)$. In

contrast, $L_d(z, \theta, \phi, \lambda)$ and $L_u(z, \theta, \phi, \lambda)$ are the downwelling and upwelling radiance, respectively, and contain the depth information, z (Mobley, 2022). The remote sensing reflectance is derived from the radiance that the satellite sensor detects. It contains information on the water depth, seafloor features, and the inherent optical properties of water (IOPs) (Ilori & Knudby, 2019). A synthetic set of reflectance spectra, $R_{rs,modelled}$, is created through forward modelling using combinations of a wide range of parameterized IOPs and depths from diverse water conditions and environments. The $R_{rs,modelled}$ is stored in Look-Up Tables (LUTs). Olayinka and Knudby (2019) affirms that water depth can be estimated from satellite imagery by inverse modelling, which involves finding the $R_{rs,modelled}$ that best matches $R_{rs,observed}$ using matching techniques, such as least squares, and then inverting it. Therefore, the depth of the water column is derived by studying the optical properties and reflectance spectra of the water column (Ilori & Knudby, 2020).

2.5 Radiative Transfer Model

2.5.1 Assumptions

Before the RTE is used, Mobley (2001) states that numerous factors and conditions need to be considered to simulate the physical processes in the water column properly. As the IOPs are inputs of the RTE, they are assumed to be constant within a layer of water. Therefore, the absorption and scattering coefficients, single-scattering albedo, and the phase function only vary with depth (Mobley, 2001) (Cheng & Liang, 2011). Due to this homogeneity of optical properties in a single layer, it is also assumed that the water column is plane-parallel (Mobley, 2001) (Cheng & Liang, 2011) (Lee & Liou, 2007). Despite the variability of IOPs, the study by (Mobley, 2001) considers time-independent IOPs and boundary conditions, as the change occurs over a prolonged period. Mobley (2001) further notes that inelastic scattering of light photons can be accommodated by the RTE but was omitted from the investigation due to the difficulty of mathematically representing it.

Further assumptions to the RTM have been made in the investigation by Lee and Liou (2007),

such as clear sky conditions at the time of acquisition and that the bottom of the water body is a Lambertian surface. While in the study by Dekker et al. (2011), the dependence on the wavelength of light in predicting optical properties is ignored.

2.5.2 RTM Equation

After estimating the optical properties and boundary conditions, the RTE can be used to evaluate the radiance distribution in water (Cheng & Liang, 2011). The RTE is an integro-differential equation given by (Mobley, 2001):

$$\cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} = -c(z, \lambda)L(z, \theta, \phi, \lambda) + \int_{4\pi} L(z, \theta', \phi', \lambda') \times \beta(z, \theta', \phi' \rightarrow \theta, \phi; \lambda) d\Omega' + S(z, \theta, \phi, \lambda) \quad (2.7)$$

where (θ', ϕ') is the scattering direction, Ω' is the solid angle centered on the scattering direction, and $S(z, \theta, \phi, \lambda)$ is the energy from chlorophyll fluorescence or inelastic scattering (Mobley, 2001). From Equation (2.7), it is evident that the attenuation coefficient of water and the VSF contribute largely to the radiance distribution (Hedley, 2008), further emphasizing how critical it is to predict these parameters accurately. Due to the complex and time-consuming evaluation of the equation, Mobley (2001) calculates the radiance distribution using the Hydrolight software. The radiance distribution obtains the spectral reflectance using the corresponding definition (Mobley, 2001).

2.6 Empirical Modelling

Empirical approaches assume that incoming solar radiation is attenuated exponentially with increasing water depth (Gholamalifard, et al., 2013). To simplify the relationship between water depth and light detected by the sensor (radiance), the radiance values are logarithmically transformed to derive a linear function. Therefore, empirical modelling involves examining the linear relationship between in-situ depth data and radiance; hence, various models or transform algo-

rithms have been developed to predict water depth. These algorithms are strengthened by using combinations of wavelengths with different attenuation rates to predict depth (Hamylton, et al., 2015). The two main algorithms are the Band Ratio Transform algorithm, developed by Richard Stumpf, and the log-linear transform algorithm by David Lyzenga.

2.6.1 Log-Linear Transform

The first and most popular empirical method, which fully uses the varying attenuation rates of different light wavelengths, is the Log-Linear transform algorithm developed by David Lyzenga. This algorithm involves using multiple linear regression to model the relationship between depth and the water-leaving radiance stored in multispectral bands (Lyzenga, 1985) to predict depth, as shown in Equation (2.8):

$$z = a_0 + \sum_{i=1}^N a_i \ln(V'_i - V_{i,\infty}) \quad (2.8)$$

where a_0 and a_i are the constants obtained from the regression analysis, V'_i is the sun-glint corrected band, and $V_{i,\infty}$ is the deep-sea signal for the band i . $V_{i,\infty}$ is subtracted from the V'_i as it has virtually zero radiance. Gholamalifard et al. (2013) states that any amount of radiance from the deep-sea signal is due to scattering, which must be removed. The Log-Linear transform simplifies the relationship between radiance and water depth. It fully explains the behaviour of light in the water column by summing the logarithmically transformed radiance values in each band.

2.6.2 Band Ratio Transform

Another robust method developed to overcome the shortcomings of the Log-Linear model, as it does not consider variation in bottom character, is the Band Ratio transform algorithm by Richard Stumpf. This uses the ratio of two multispectral bands to compensate for variable bottom characters and cancel out their effects to make more accurate depth estimations. The relationship between depth and the band ratio is modelled using simple linear regression, as shown in Equation (2.9):

$$z = m_1 \frac{\ln(nV'_i)}{\ln(nV'_j)} - m_0 \quad (2.9)$$

where m_0 and m_1 are the coefficients obtained from the regression analysis, V'_i and V'_j are the sun-glint-corrected radiance from bands i and j , and n is an arbitrary integer large enough to ensure that the logarithms are positive (Caballero & Stumpf, 2019).

2.7 Machine Learning

As the deterministic models use assumptions to derive a mathematical relationship between water depth and reflectance, the intrinsic patterns in the reflectance data are neglected. Therefore, the performance of supervised machine learning (ML) models must be tested for SDB, as they study the patterns in remotely sensed data and could achieve higher accuracies than empirical and physics-based models.

ML models can simulate the non-linear relationship between the multispectral reflectance and the target classes of water depth through complex nonlinear regression and classification (Duan, et al., 2022) and hence require large amounts of training data. By learning the relationship or the attributes of the target classes of water depth, the model identifies these learned attributes in new unclassified reflectance data to predict depth (Belgiu & Dragut, 2016). These attributes can be considered as the reflectance intensity of each target class at different wavelength bands. However, as a significant proportion of the reflected signal is absorbed in the water column before reaching the sensor, the number of bands that can be used as predictors is limited (Manessa, et al., 2020). This challenges SDB by ML models as high-dimensional data is required (Manessa, et al., 2020). However, ML models like Random Forest can be optimized to adapt to limited data and produce accurate predictions.

The Random Forest is a non-parametric supervised learning model that does not make assump-

tions about the relationship between the independent and target classes (Belgiu & Dragut, 2016). It is also an ensemble learning-based algorithm that uses numerous decision trees to make a single prediction for classification or regression tasks, as shown in Figure (2.5). It has become popular for SDB as it adapts to the complex distribution of multispectral reflectance data and learns the attributes of each target class (Belgiu & Dragut, 2016). It does not enforce a linear relationship by using logarithmically transformed reflectance data but instead uses just the atmospheric corrected multispectral reflectance as input data (Duan, et al., 2022).

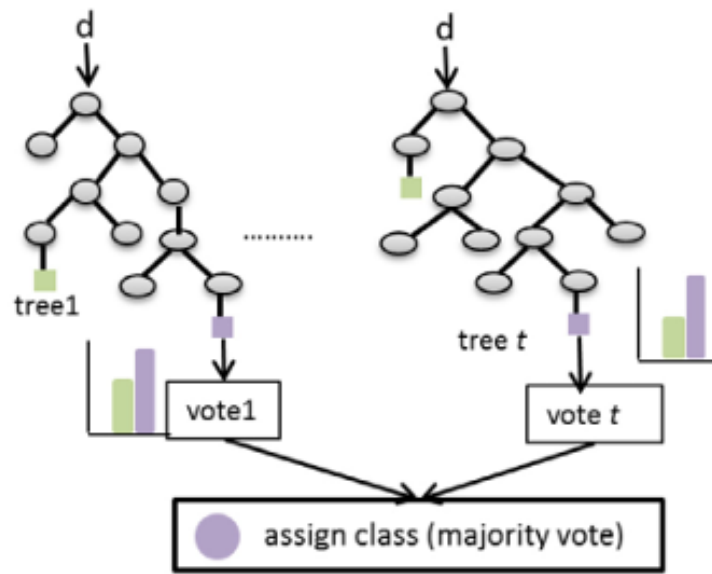


Figure 2.5: Classification by the Random Forest model

Source: Belgiu, M. and Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*. 114:24–31.

The forest is grown through bootstrap aggregating or bagging, where each decision tree is built and trained independently using a random subset of the data with replacement to reduce overfitting and the classification variance (Belgiu & Dragut, 2016). The predictions from each tree are then aggregated to obtain the final result. The model can achieve high accuracies due to the randomness incorporated during the construction and training of the model. This accuracy relies on the choice of model hyperparameters, which include the number of trees, tree depth, number of predictors, sample size, and number of leaf nodes (Manessa, et al., 2020). Research has been performed on how the RF model can be optimized for SDB-related problems through hyperparameter tuning,

selection of the best combination of predictors, or both (Manessa, et al., 2016) (Çelik, et al., 2023).

2.7.1 Random Forest Classifier

Non-parametric learning is preferred for image classification tasks over parametric learning. An algorithm is free from parameter limits and learns from the data to create a function that can map the labelled inputs to the output. The algorithm does not make any assumptions about the data distribution and is hence suitable for classifying multispectral satellite imagery, which scarcely has normal distributions (Belgiu & Dragut, 2016). Therefore, the Random Forest was used in this investigation as it is a non-parametric supervised classifier that does not make any assumptions about the form of the mapping function (Belgiu & Dragut, 2016).

The RF model is an ensemble learning approach composed of a collection of Classification and Regression Trees (CART) trained independently using the bagging approach. Belgiu and Dragut (2016) explain that bagging or bootstrap aggregating is a technique to reduce the variance in predictions from the decision trees. Each classifier or decision tree is trained on a randomly selected subset of the data, referred to as the in-bag samples. This subset, about two-thirds of the entire dataset, is replaced to ensure that data points can be selected multiple times for training (Belgiu & Dragut, 2016). The remaining third is referred to as the out-of-bag samples and is used to validate the performance of the RF model during cross-validation. The estimation error is known as the out-of-bag (OOB) error and evaluates the accuracy of the classification.

In addition, the RF is flexible in its ability to perform classification and regression tasks on different data types. This flexibility lies in the liberty to set the parameter values, which can impact the model's performance. The main parameters that significantly influence the performance of the RF (Appendix B) are the number of features to split a node ($Mtry$), the number of decision trees ($Ntree$), and the tree depth (Kn) (Scornet, 2017). Therefore, each parameter will be explored in detail to determine the optimal range of values and justify its use in our investigation.

2.8 Deep Learning

Predicting water depth from satellite imagery is challenging as the image datasets are heterogeneous, caused by atmospheric conditions, waves, sediments, brightness variations, and cloud shadows. In addition, classifying images solely based on spectral signatures is limited, as publicly available multispectral images have a large spectral resolution. As the previous section has revealed, ML models consider an individual pixel at a time, missing much contextual information, which results in extreme prediction errors (Maggiori, et al., 2016). ML models are also sensitive to noise as they attempt to model each data input and fail to generalize to unseen data. Deep learning models can generalize better to unseen data through regularization techniques. Therefore, deep learning (DL) models to predict depth will be reviewed.

With the growth of publicly available and large datasets, particularly multispectral satellite imagery, deep learning has become a popular tool in many computer vision applications, such as image segmentation and classification. As the estimation of water depth from images involves the analysis and classification of the image, computer vision techniques have been adopted to advance the study of SDB (Wilson, et al., 2020). Therefore, SDB is designed as a pixel-wise classification task to exploit the use of DL techniques in mapping pixels to class labels (Wilson, et al., 2020). Deep learning overcomes the influence of noise by using a logical brain-like structure to process and analyze satellite imagery and learn the patterns in the data to make accurate predictions. DL models also consider the spatial correlation of pixels by assuming adjacent pixels have a similar relationship to depth (Lumban-Gaol, et al., 2021). Therefore, DL models are powerful tools for SDB, which can address the non-linear relationship between the pixel reflectance value and water depth and the spatial correlation between pixels.

DL models include neural networks, supervised and unsupervised feature learning algorithms, and probabilistic models representing the patterns and structures in large datasets. Voulodimos et al. (2018) defines a neural network as a layered structure of algorithms with hidden layers that learn

a mapping function between the input and output layers. Convolution neural networks (CNNs) are a neural network inspired by the biological visual system structure to study and extract an image's visual and contextual features (Voulodimos, et al., 2018).

2.8.1 Encoder-Decoder CNN architecture

Image classification remains a key problem in SDB as assigning a label to every pixel in a heterogeneous image is complex (Maggiori, et al., 2016). As mentioned, the Random Forest algorithm does not consider the spatial relationships of pixels in the image. CNNs are used in our investigation to improve the water depth estimation.

The CNN is a form of an artificial neural network (ANN) that deals predominantly with image data and performs tasks like image classification, object detection, and pattern recognition (O'Shea, 2015). CNN can be tailor-made with different image processing features, which gives rise to many CNN architectures that can manage the complexity of image data. However, O'Shea (2015) warns that with larger images, the number of weights and trainable parameters of the CNN increases, and so does the risk of overfitting. Therefore, the aim of SDB modelling using deep learning models is to reduce the complexity of CNNs while still preserving the model's predictive power.

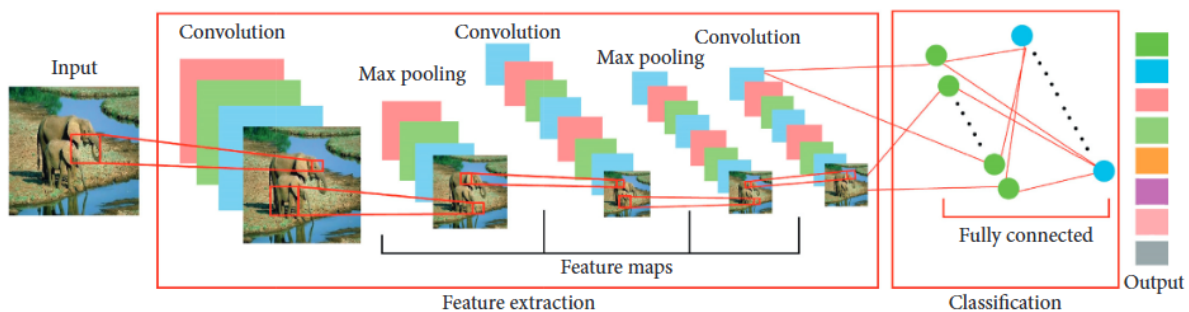


Figure 2.6: A simple convolutional neural network

Source: Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., and Dar, S. H. (2021). Satellite and scene image classification based on transfer learning and fine tuning of resnet50. *Mathematical Problems in Engineering*, 2021(1):5843816.

A CNN architecture consists of three hidden layers, the convolutional layer, the pooling layer, and the fully connected layer, as shown in Figure (2.6). CNNs are tailor-made for image classification by arranging the neurons in the hidden layer in three dimensions to represent the image's height, width and channel depth (O'Shea, 2015). The layer transforms the 3D input volume to a smaller 3D output volume passed to the next layer to produce a $(1, 1, n)$ output volume, where n represents the target variable (O'Shea, 2015). As the input volume passes through the system, its dimensionality is progressively reduced, allowing each subsequent layer to learn the complex patterns.

However, Maggiori et al. (2016) notes that the basic CNN architecture can only extract features from the input imagery by progressively compressing the input as it moves through the network to produce an output that is a lower-dimensional classification of the input. This process of transforming the input data to a simpler representation of the input is performed by an encoder (Badrinarayanan, et al., 2017). A decoder produces multi-dimensional dense predictions that match the original input size. Therefore, an encoder-decoder CNN architecture, shown in Figure (2.7), is used in our investigation.

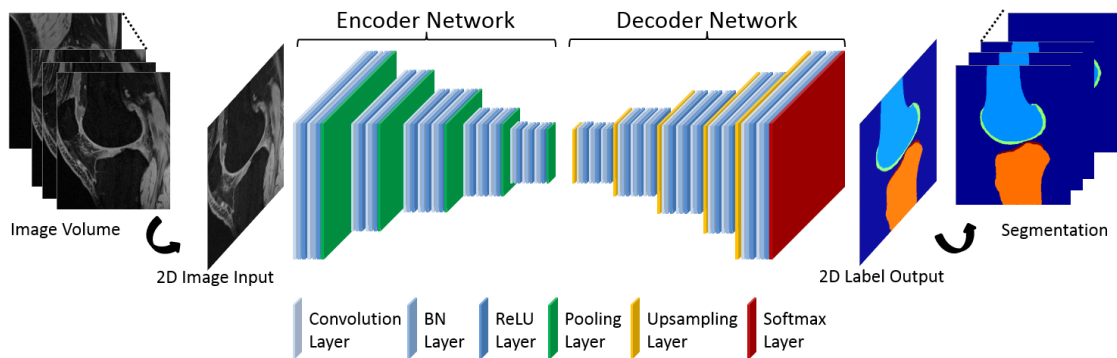


Figure 2.7: A basic encoder-decoder architecture

Source: Liu et al. (2017). Deep convolutional auto-encoder and 3d deformable approach for tissue segmentation in magnetic resonance imaging. *International Society for Magnetic Resonance in Medicine*, 25.

In the encoder network, the input layer is responsible for vectorizing the input image, which will be mapped to a feature space (Liu, et al., 2017). The next layer is the convolutional layer. This layer performs operations on the input data using multiple small $k \times k \times depth$ filters, also known

as kernels. The kernel, a matrix of learnable weights, slides over the images and performs a dot product between the kernel weights and the image patch covered by the kernel, known as the receptive field (Maggiori, et al., 2016). The kernel covers the entire image by moving in strides that the user can specify. This produces a feature map, a matrix of positive and negative output values, of the same size as the image that represents specific features or sections of the image (Mascarenhas & Agarwal, 2021). This operation allows the model to identify kernels that activate when they encounter certain features in the input image (O'Shea, 2015). However, this operation only allows the model to learn linear relationships. Therefore, an activation function is applied to the output of the convolutional layer to introduce nonlinearity for the model to learn complex patterns (Maggiori, et al., 2016). A popular activation function used in many CNNs and the study by Ai et al. (2020) is the Rectified Linear Unit (ReLU) defined in Equation (2.10):

$$y = \max(0, x) \quad (2.10)$$

where x is the input pixel value. Mascarenhas and Agarwal (2021) explain that the function receives the values of the feature map and returns the same value as the input if it is positive and returns zero if the input is negative.

Next is the pooling layer, which reduces the dimensions of the output feature maps. This form of downsampling involves aggregating the information in many neurons into fewer neurons using a kernel to reduce the number of model parameters while retaining the relevant information (Maggiori, et al., 2016). The pooling layer usually follows each set of convolutional layers to reduce the size of feature maps. There are two types of pooling layers: average pooling and max pooling. The average pooling function calculates the average input value in a selected region, while the max pooling function returns the maximum input value in a region (O'Shea, 2015). In the max pooling layer, a 2x2 kernel is applied to the feature map, reducing its size by 25% (O'Shea, 2015). The flattened layer follows the pooling layer, which transforms the multidimensional input vectors into 1D context vectors to form the latent feature space that can be passed to the decoder network.

The encoder and decoder are usually symmetric and may contain the same number of layers. The decoder network does the opposite of the encoder network by upsampling the context vectors into the original input resolution (Maggiori, et al., 2016). The upsample layer uses indices from the last pooling layer in the encoder network to perform a nonlinear transformation of the context vectors to a feature map (Badrinarayanan, et al., 2017). This operation, called upsampling, involves reconstructing the details of the image that can be fed to subsequent convolutional layers. The convolutional layers in the decoder network apply a dot product operation on the feature map using a transposed kernel from the corresponding encoder network to reconstruct the original input (Badrinarayanan, et al., 2017). As the decoder predicts the output scores from the inputs, nonlinearity is introduced to the decoder network using the ReLU activation function to allow the model to make accurate predictions from complex data.

The fully connected (FC) layer, which consists of a Softmax activation function that converts a vector, $\vec{X}$, of K raw output scores into a probability distribution of K possible class predictions that add up to 1 (Mascarenhas & Agarwal, 2021). This function is defined in Equation (2.11):

$$\sigma(\vec{X})_i = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad (2.11)$$

where x_i and x_j are the predicted probabilities of the input vector belonging to class i and j respectively, K is the number of classes, and σ is the Softmax. Badrinarayanan et al. (2017) explain that the Softmax layer returns a classified image, where each class identified had the maximum probability at the given pixel.

2.8.2 ResNet-50

DNNs require enormous datasets to train and can take days to complete. However, in most classification problems, the dataset is too small to train the entire model. Therefore, part of the pre-trained

CNN model or the entire model is used and attached to the new segmentation networks to perform these classification tasks. This process is known as transfer learning, where a pre-trained model is adapted for a new problem (Mascarenhas & Agarwal, 2021). The weights of the pre-trained model are frozen to ensure they are not updated during the training, but only the weights of the new network are updated.

In the Computer Vision field, many networks are trained on large datasets. The early layers aim to learn edge detection, the middle layers learn to identify shapes, while the later layers, which are specific to the task at hand, learn to identify specific features. Shabbir et al. (2021) explains that in transfer learning, the last few layers, usually the FC, Softmax, and classification layers, are retrained on the new dataset to reduce the computational time. Popular pre-trained models specialized in image processing include VGG, Xception, and ResNet. These models have been trained on the ImageNet dataset, which comprises over one million images and is the backbone of many segmentation networks (Shabbir, et al., 2021).

Our investigation uses the ResNet-50 model as the backbone to the U-Net, SegNet, and DeepLabv3+ models that will be trained on RGB satellite imagery. ResNet-50 is a deep residual network that learns residual information and consists of an encoder with an FC layer composed of interconnected neurons. It is considered a deep residual network with 48 convolutional layers (Mascarenhas & Agarwal, 2021). It differs from the standard CNN architecture by having a max pooling layer and an average pooling layer to make up 50 layers in total, as shown in Figure (2.8). As with deep NNs, the more complex they become, the less they can generalize making them prone to generating large prediction errors. To reduce prediction errors, the ResNet-50 network includes a residual block. In the residual block, an identity activation function, which returns the same value as its input, is used as an identity connection, or skip connection to pass the gradient from earlier to later layers and overcome degradation and vanishing gradients (Shabbir, et al., 2021).

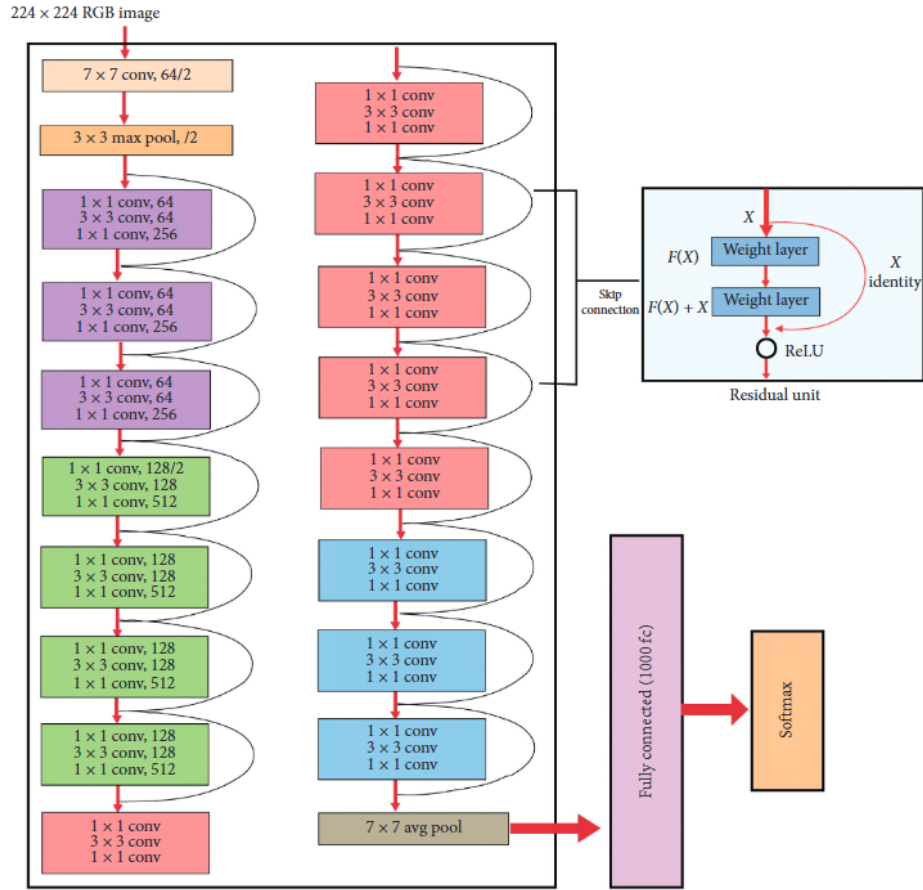


Figure 2.8: ResNet-50 architecture

Source: Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., and Dar, S. H. (2021). Satellite and scene image classification based on transfer learning and fine tuning of resnet50. *Mathematical Problems in Engineering*, 2021(1):5843816.

The first convolutional layer uses 64 7×7 kernels to perform a dot product operation across the entire input image, followed by a 3×3 max pooling layer with a stride of 2. The input size is reduced by half, while the channel depth doubles. This progressively occurs as the input passes through the network. All through the model, skip connections from the residual block ensure that the residual input information from the initial layers is added to the latter layers (Shabbir, et al., 2021). After the first convolutional layer, there are identical convolutional blocks that consist of a 1×1 kernel that restores the dimensions of the input, while the 3×3 kernel is a bottleneck layer. A 7×7 average pooling layer follows, an FC layer of 1000 neurons, and a Softmax layer computes the classification scores (Shabbir, et al., 2021).

The ResNet-50 model is used as a backbone in many segmentation models, where the final stages of the model are adjusted by adding the three layers mentioned previously.

2.8.3 U-Net

U-Net is an image segmentation network created by modifying the Fully Convolutional Network to operate with a small training set (Ronneberger, et al., 2015). Consecutive upsampling layers were appended to the encoder network of the Fully Convolutional Network. Ronneberger et al. (2015) states that the U-Net model is unique with skip connections to recover the fine-grained details from the encoder feature maps to add to the upsampled feature maps. As the encoder-decoder architecture is symmetric for typical segmentation models, the upsampling section of the U-Net model is slightly different from the downsampling section due to the increased number of feature channels from the skip connections, which creates a U-shaped architecture, as shown in Figure (2.9).

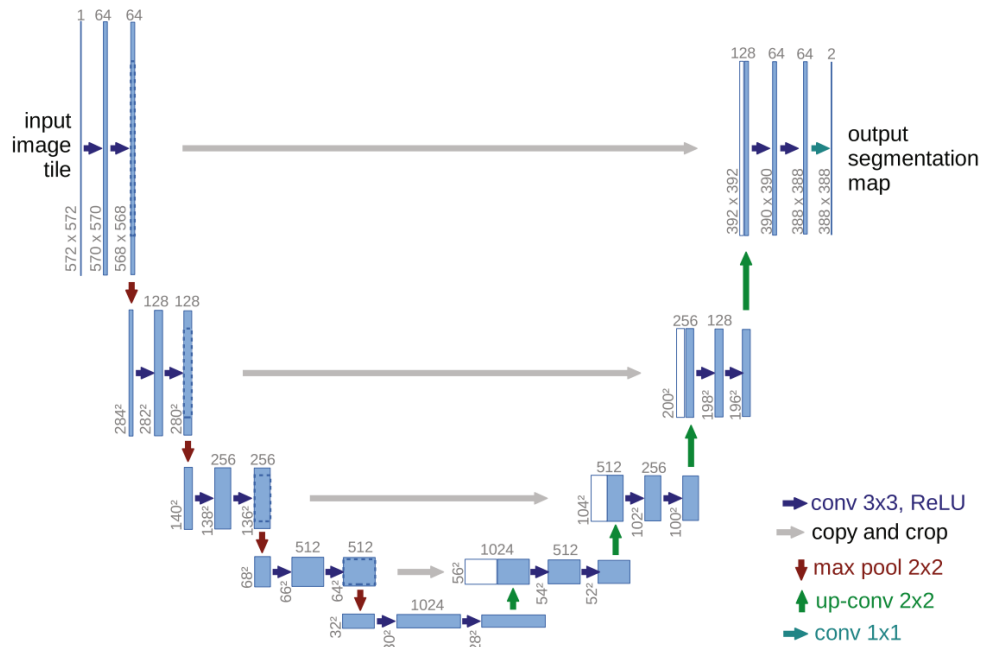


Figure 2.9: U-Net architecture

Source: Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

The left side of the network, known as the contracting path, is identical to the standard CNN encoder architecture, where a convolutional block is made up of two 3x3 convolutional kernels each followed by a ReLU activation layer (Ronneberger, et al., 2015). The contracting path contains five convolutional blocks with a 2x2 max pooling layer, with a stride of 2, following each block to downsample the feature map. The number of feature channels doubles after each downsampling operation. The right side is the expansive path and contains four convolutional blocks that upsample the encoded feature map. Before the encoded feature map is passed through the 3x3 convolutional layers, it is upsampled using a 2x2 convolutional kernel that reduces the feature channels by half (Ronneberger, et al., 2015). During the upsampling operation, as contextual information from the corresponding contracting path feature maps is added to the output, the two 3x3 convolutional kernels are used to train the model to consolidate the two pairs of information. The ReLU activation function is applied to the output to introduce nonlinearity into the model.

As a result of the skip connections, the U-Net model does not need an FC layer to calculate the output class scores. After the last upsample operation, the output feature map is consolidated to the original input image for contextual information as shown in Figure (2.9) (Ronneberger, et al., 2015). Combining the predicted feature map with its corresponding contracting path feature map is known as the overlap-tile strategy, shown in Figure (2.10).

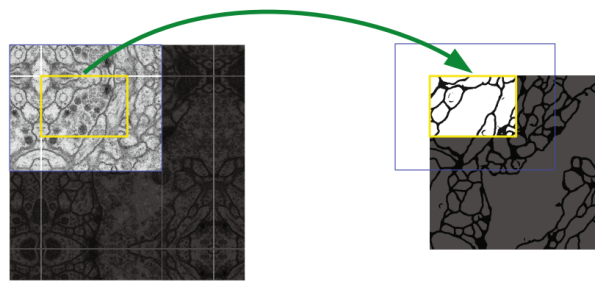


Figure 2.10: Overlap-tile strategy of U-Net

Source: Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18, pages 234–241. Springer.

The predicted feature map then passes through the last two convolutional layers and then the final

layer, which is a 1x1 convolutional kernel that maps each 64-component feature vector to the specific number of classes to finally generate the output segmentation map (Ronneberger, et al., 2015).

The efficiency of the U-Net model is due to its skip connections, which allow the model to apply the overlap-tile strategy to use contextual spatial information at each step in the learning and prediction process. It is specially equipped to predict the image boundaries as Ronneberger et al. (2015) mirror the input image in the last convolutional step to extrapolate the contextual information for the border pixels.

2.8.4 SegNet

SegNet is a deep CNN with an encoder-decoder architecture for semantic pixel-wise image segmentation. The encoder network was inspired by the VGG16 network, a 16-layer CNN, with 13 convolutional layers and 3 FC layers, used for object recognition and image classification (Badrinarayanan, et al., 2017). In the encoder, the FC layers were removed to preserve the encoded high-resolution feature maps and to reduce the number of trainable parameters. The model also uses pre-trained weights from the VGG16 network to ensure that the appropriate weights for image classification are initialized. The decoder is built by creating a decoder layer for each corresponding encoder layer. The decoder output is then passed to the pixel-wise classification layer as shown in Figure (2.11).

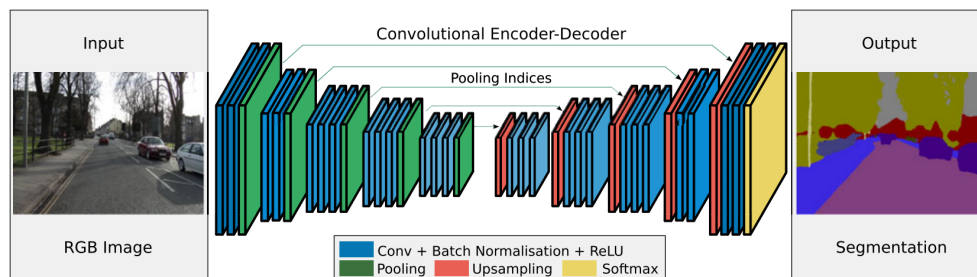
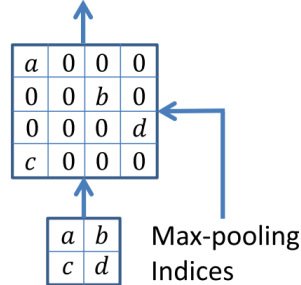


Figure 2.11: SegNet architecture

Source: Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.

In the encoder, a convolutional layer with a 7x7 kernel produces a set of normalized feature maps. This technique is called batch normalization, which involves calculating the mean and variance of the set of feature maps and using these to normalize all the feature maps (Badrinarayanan, et al., 2017). Once the layer outputs are normalized, the ReLU activation function is applied to allow the model to learn complex relationships. The feature maps are then downsampled using a 2x2 max pooling layer with a stride of 2, followed by a subsampling of a factor of 2 (Badrinarayanan, et al., 2017). However, multiple subsampling operations cause a loss of detail in feature maps, especially boundary details. Figure (2.12) shows that to mitigate this, the SegNet model stores the max-pooling indices, which Badrinarayanan et al. (2017) explains are the matrix coordinates of the maximum feature value in each pooling window.

Convolution with trainable decoder filters



Max-pooling
Indices

Figure 2.12: SegNet max-pooling indices

Source: Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.

Badrinarayanan et al. (2017) notes that the SegNet model is more efficient in its use of memory than the U-Net model, as it uses the stored max-pooling indices from the corresponding encoded feature maps for contextual information for the upsampling operations in the decoder. As the up-sampling layers use the max-pooling indices to generate sparse feature maps, no additional learning occurs in the decoder, essentially decreasing the number of trainable parameters. The sparse feature maps are fed to the convolutional layers to produce dense feature maps that are then normalized (Badrinarayanan, et al., 2017). The ReLU activation function is not applied to the outputs of the convolutional layers in the decoder as no learning occurs. The last convolutional layer in the de-

coder produces a high-dimensional feature map. The Softmax classifier receives the final feature map and classifies each pixel to return a K-channel image of class probabilities that add up to 1, where K represents the number of classes (Badrinarayanan, et al., 2017).

2.8.5 DeepLabv3+

The encoder-decoder architecture is successful in detecting object boundaries using max-pooling indices in the SegNet model and skip connections in the U-Net model, which can retrieve the spatial contextual information from the input image. However, these networks require a reduction in the dimensionality of the input data to learn fine-grained patterns and features. This is achieved by using subsequent convolutional and pooling layers to generate a feature map encoded with semantic information. Chen et al. (2018) states that though the decoder can upsample the encoded feature maps, there is still a loss of vital spatial information and resolution at specific scales necessary for semantic pixel-wise segmentation. The operations in the encoder convolutional layers, where $k \times k$ kernels with a stride of integer n , generate sparse feature maps as not all the contextual information in the input image is captured, as depicted in Figure (2.13a).

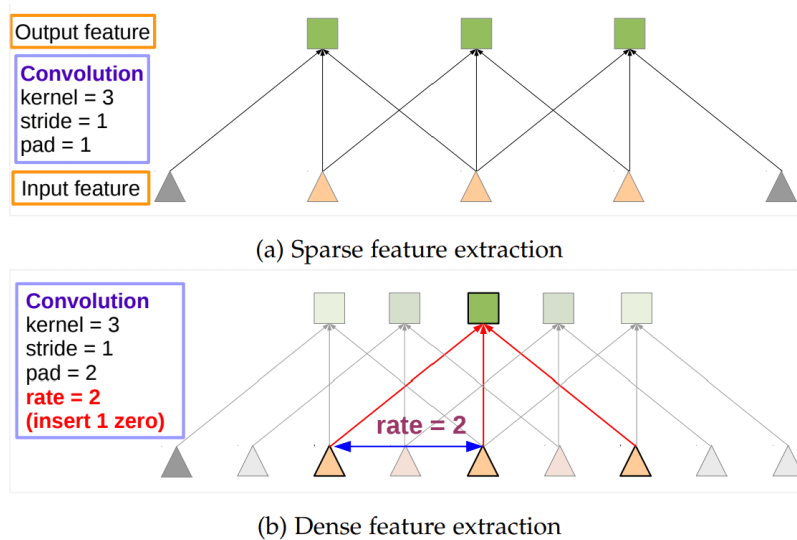


Figure 2.13: Feature extraction from a 1D input. (a) Sparse feature extraction with ordinary convolution. (b) Dense feature extraction with atrous convolution with sampling rate $r = 2$

Source: Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

For this reason, an atrous spatial pyramid pooling module is incorporated into the encoder-decoder architecture to create the DeepLabv3+ network (Chen, et al., 2018). The atrous spatial pyramid pooling (ASPP) is a module used in segmentation networks to extract dense features from an input feature layer (Chen, et al., 2017). From Figure (2.13b), an atrous convolution operation involves using the exact specifications of a standard convolutional kernel but with a sampling rate of r . The sampling rate, also known as the dilation rate, controls the distance between filter elements in a kernel and essentially the field-of-view of the kernel, while still maintaining the same number of elements (Chen, et al., 2017). An atrous convolutional layer allows the network to extract more contextual information from the input and generate denser feature maps. In this case, a sampling rate of 2 means the atrous kernel leaves one-pixel space or introduces zeros between elements. As the kernel moves by a stride of 1, it captures all the pixel positions of the image.

The ASPP module also allows the network to learn features at multiple scales. As intrinsic features may not be distinguishable at certain scales and resolutions, Chen et al. (2017) state that the ASPP uses multiple parallel atrous convolutional layers with different sampling rates and merges these multiscale feature maps to generate a more accurate feature map. This approach is illustrated in Figure (2.14)

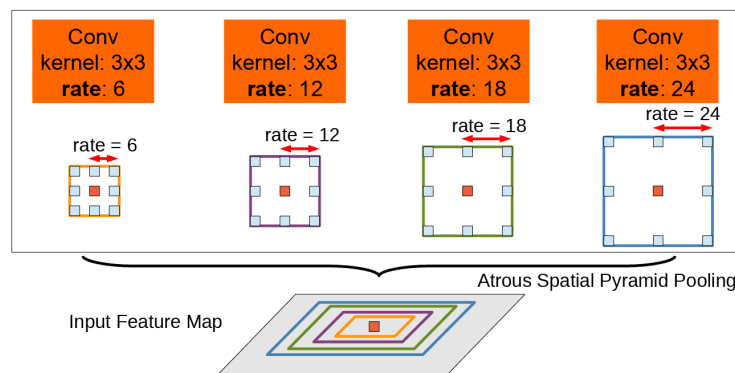


Figure 2.14: Atrous Spatial Pyramid Pooling

Source: Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

The multiscale feature maps are merged by taking the maximum response across all scales at each

image position to ensure that spatial information at various scales is utilized in the learning and prediction process. This improves the efficiency of the DeepLabv3+ network, as standard encoder-decoder architectures cannot produce dense feature maps that are more than 4 times smaller than the original input image due to limited GPU memory. Therefore, multiple parallel atrous convolutional layers provide a good alternative to generating denser feature maps.

The architecture of the DeepLabv3+ network, illustrated in Figure (2.15), allows the network to combine the efficiency and object border detection capabilities of the encoder-decoder architecture with the multiscale techniques of the ASPP module. The network uses its predecessor, the DeepLabv3 network, as its encoder module and attaches a simple decoder to it. The DeepLabv3 network has an atrous convolutional block consisting of a 1x1 convolutional layer, three 3x3 atrous convolutional layers with different sampling rates, and a pooling layer. The output from the first convolutional layer is fed through three parallel atrous layers, producing various multiscale feature maps. After the image pooling player, these dense feature maps are merged into a dense feature map through a 1x1 convolutional layer (Chen, et al., 2018).

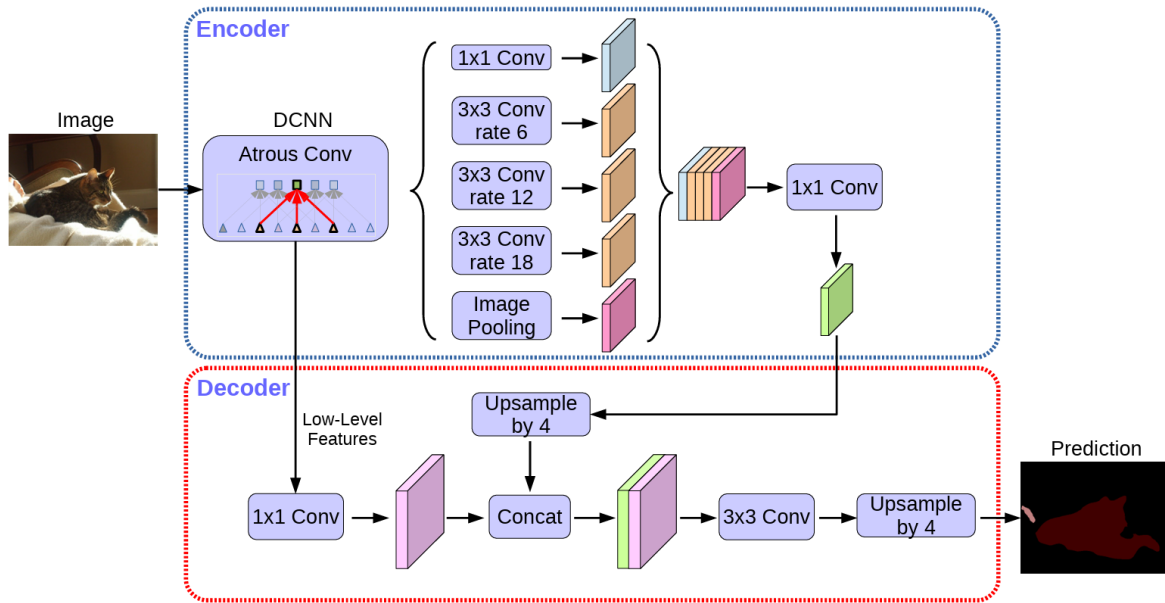


Figure 2.15: DeepLabv3+ architecture

Source: Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818

The encoded feature map is then passed to the decoder. It is upsampled by a factor of 4 and integrated with low-resolution features from the ResNet backbone that have passed through a 1x1 convolutional layer to reduce the number of channels (Chen, et al., 2018). The output is then refined through another 3x3 convolutional layer and upsampled by a factor of 4 for the final time to produce the network's prediction.

Chapter 3

Related Works

This research was motivated by the growing interest in implementing data-driven models in SDB tasks to assess their performance against the current empirical and physics-based models in terms of scalability and accuracy. In this chapter, we review related works in SDB that use different modelling approaches to orient the scope of the investigation and identify key areas that require more research. Firstly, studies that use empirical algorithms are presented, followed by physics-based studies to review the general constraints and limitations of these methods. Finally, current works that apply artificial intelligence models are reviewed.

3.1 Empirical algorithm studies

The investigation of North Cat Cay and Bimini in the Bahamas involved the Log-Linear model, where LiDAR and passive multispectral scanner (MSS) data were combined to capitalize on the high accuracy of the LiDAR system and the ability of the MSS system to capture passive signals from inaccessible areas, and therefore determine the model's coefficients (Lyzenga, 1985). In-situ data from the LiDAR scanner was co-registered to the passive signals at corresponding sample locations to extract the statistical relationship between water depth and signal intensity, as shown in Equation (2.8). According to an earlier study by Lyzenga (1978), modelling the behaviour of light in water using a pair of wavelength bands through multiple linear regression compensates for

the variable physical properties of the water column. Therefore, to find the optimum band pair for depth prediction, Lyzenga (1985) first logarithmically transformed the radiance values to obtain the pseudo-depths, as shown in Equation (3.1):

$$x_i = \ln(V'_i - V_{i,\infty}) \quad (3.1)$$

where x_i is the pseudo depth for band i . A regression analysis is performed between a sample of depths z and the pseudo depths of each possible band pair. This would result in a set of equations, as shown in Equation (3.2), with the corresponding correlation coefficient (R). The pair of bands with the highest R was used in the subsequent depth prediction of the entire study area.

$$z = a_0 + a_i x_i + a_j x_j \quad (3.2)$$

The analysis of the North Cat Cay revealed that the pair of wavelengths with the highest R , of 0.957, was $0.50 - 0.54\mu\text{m}$, which corresponds to green light, and $0.58 - 0.64\mu\text{m}$, being red light, while Bimini's optimum pair was $0.48 - 0.520\mu\text{m}$ (blue light), and $0.55 - 0.60\mu\text{m}$ (green light) had an R -value of 0.917. The calculated RMSE between the in-situ LiDAR data and the predicted depth was 0.728m for North Cat Cay and 0.928m for Bimini.

Other studies have modified the Log-Linear model due to the variation in the degree of attenuation of different wavelengths. As light passing through a medium is absorbed and scattered, shorter wavelengths are the least attenuated compared to longer wavelengths; hence, blue light penetrates deeper in the water column. For this reason, Gholamalifard et al. (2013) modified the Log-Linear model by using a single band of radiance to form the Single Band algorithm (SBA), as shown in Equation (3.3), for their investigation of the Caspian Sea in South-East Iran,

$$z = \frac{-1}{2k} \ln(V'_i - V_{i,\infty}) + \frac{1}{2k} \ln V_o \quad (3.3)$$

where, k is the water attenuation coefficient and V_o is the sensitivity factor. The attenuation co-

efficient, a function of the wavelength, sea bottom character, and the water's physical properties, is assumed to be constant for each spectral band (Gholamalifard, et al., 2013). Once the radiance was logarithmically transformed, simple linear regression was applied between the sample depth data and the pseudo-depths or transformed spectral bands. In this case, the blue band and red band were used to make separate linear equations that would be used for depth prediction. The correlation coefficients between water depth and the blue and red bands were found to be -0.2 and -0.65, respectively, showing an inverse relationship (Gholamalifard, et al., 2013). The coefficient of determination (R^2) was 4.31% and 43.47% for the blue and red bands, respectively, indicating that the variance in the depths is explained more by the transformed red band than by the transformed blue band (Gholamalifard, et al., 2013).

The results from Gholamalifard et al. (2013) revealed that the SBA using the transformed red band produced reasonable predictions, with an R -value of -0.66 as compared to the blue band with an R -value of -0.21. The poor performance of the blue band algorithm was due to the high organic matter concentration in the coastal waters of the Caspian Sea, which absorbs blue light more rapidly than red light (Gholamalifard, et al., 2013).

Gholamalifard et al. (2013) proved that the SBA is a viable alternative to the Log-Linear model when there is limited in-situ data for depth calibration. The choice of spectral band relies on the characteristics of the coastal study area. Blue light can penetrate deeper in clearer water without dissolved organic matter or phytoplankton, so the blue SBA is appropriate for depth estimation in clear and calm water. At the same time, the red SBA is better suited for estimating the depth of turbid waters as it is the least absorbed by organic matter (Gholamalifard, et al., 2013). However, the red SBA is only reliable within the 0 – 5m range, as beyond this depth, the rate of absorption of red light by water molecules increases with depth.

It is important to note that the Log-Linear model and its modifications do not account for sea

bottom types and the variation in bottom reflectance but are only suitable approaches for predicting the depth of shallow coastal zones with homogeneous water quality (Lyzenga, 1985). As the model rightly compensates for the varying attenuation coefficients of distinct wavelengths, failing to consider the difference in the bottom reflectance across wavelengths and bottom types does not allow the model to predict depth in any body of water accurately. Therefore, this leads us to examine the applicability and accuracy of Stumpf's Band Ratio Transform algorithm in the following section.

Caballero and Stumpf (2019) applied the Band Ratio model to predict the water depth of Dry Tortugas, Key West, and West Palm Beach in South Florida, USA. Sentinel 2-A and 2-B satellite imagery, at 10m resolution, were used as they contain 13 spectral bands that provide significant insight into the behaviour of light in water (Caballero & Stumpf, 2019). ALB provided the in-situ depth data for the study and was referenced to the Mean Lower Low water level (Caballero & Stumpf, 2019). The ratio transform algorithm was used to compensate for the exponential decrease of light with depth and the reflectance from different bottom types (Caballero & Stumpf, 2019), which is effective for coastal water bodies. Two bands were first logarithmically transformed to obtain a linear relationship with the in-situ depth data. Caballero and Stumpf (2019) calculated the ratio of the log-transformed bands as shown in Equation (3.4):

$$pSDB = \frac{\ln(nV'_i)}{\ln(nV'_j)} \quad (3.4)$$

where $pSDB$ is the pseudo-depth.

Two ratios were used in the investigation, the ratio of the blue band to the red band (SDB_{Red}) and the blue band to the green band (SDB_{Green}) and the correlation between these ratios and the in-situ data was examined. Caballero and Stumpf (2019) were motivated to use these ratios due to the penetrative power of blue light in clear water, and the faster absorption rates of green and red light in highly productive and turbid waters, which would provide more information in shallower water. A simple linear regression analysis was performed between the ratios or pseudo-depths and in-situ depth to derive the coefficients m_0 and m_1 . The coefficients obtained were then inserted into Equation (2.9)

to form a set of SDB_{Red} and SDB_{Green} linear equations to predict water depth. The predicted depths were compared to ALB data and chart soundings of the coastal waters in South Florida to determine the precision and accuracy of the empirical ratio transform algorithm. The SDB_{Green} ratio algorithm could accurately predict depths up to 18m in West Palm Beach with a Median Absolute Error (MedAE) of $<0.5\text{m}$. Due to the heterogeneous turbidity of the coastal zones of Key West, SDB_{Red} was preferred for depth prediction as SDB_{Green} poorly predicted depth in waters of less than 3m (Caballero & Stumpf, 2019). SDB_{Red} could predict up to 6m with a MedAE of 0.42m (Caballero & Stumpf, 2019), which revealed that the SDB_{Green} algorithm can accurately represent deep waters while the SDB_{Red} performs better in shallow waters. The results from the investigation prove that using empirical algorithms to estimate water depth from Sentinel-2 imagery provides valuable bathymetric information, as channels and straits can be identified (Caballero & Stumpf, 2019). A spatial component should be considered in SDB as the SDB_{Green} and SDB_{Red} algorithms vary in their performance in shallow and deep waters. In addition to this spatial variation, different bottom types affect the reflectance of various wavelengths.

Therefore, to better estimate bathymetry, Hamylton et al. (2015) and Traganos et al. (2018) optimized the Band Ratio algorithm through geographical error analysis. The investigation by Hamylton et al. (2015) focused on Lizard Island, a continental island fringing reef, and Sykes Reef, a planar platform reef, in the Great Barrier Reef. In contrast, Traganos et al. (2018) selected the coastal area of Apokoronas in West Crete, the southeast part of the Thermaikos Gulf, and the northwest Aegean Sea as their study area. As the study on the coastal waters of South Florida revealed that specific band ratio algorithms tend to under-predict and over-predict in some regions, these studies examined the spatial distribution of the algorithm's residuals, which are the differences between the predicted and in-situ depth data. By using exploratory spatial data analysis, the spatial patterns of the residuals were identified, which would give valuable information on the algorithm's performance (Hamylton, et al., 2015).

Hamylton et al. (2015) selected to use the blue (478nm) and green (546nm) wavebands of the WorldView-2 imagery in the Band Ratio algorithm. These reflectance bands were logarithmically transformed, and the ratio of the log-green band to the log-blue band was used in the subsequent regression analysis. The simple linear regression analysis between the pseudo-depths and the MBES collected data showed that the R^2 values were 0.76 for Lizard Island and 0.83 for Skyes Reef (Hamylton, et al., 2015). From the empirical equations derived from the regression analysis, the water depth of both study areas was predicted. The algorithm's accuracy was assessed using the correlation coefficient between the predicted depths and a set of in-situ validation points not used in the regression analysis. The R^2 value for Lizard Island was 0.74 and Skyes Reef was 0.82 (Hamylton, et al., 2015). Traganos et al. (2018) used the ratio of the log-blue band to the log-green band from Sentinel-2 imagery and found that the Band Ratio algorithm achieved an RMSE of 3.49m in the 0-25m range with an $R^2 = 0.59$ and an RMSE of 1.78m for the 0-12m range with a $R^2 = 0.5$.

The model's residuals at specific geographical points were then transformed into a continuous surface raster using Kriging interpolation to visualize their spatial arrangement. There was no distinguishable pattern in the residuals of Lizard Island, but Skyes Reef showed that there were under-predicted clusters in the shallow water region. In addition, the west of the reef was consistently over-predicted. At the same time, the east was under-predicted (Hamylton, et al., 2015) meaning that positive and negative residuals are inclined to aggregate in space. This finding by Hamylton et al. (2015) indicates that there is a distinct spatial distribution in the empirical algorithm's errors which follows Tobler's first law of geography that features close to each other are more related than features that are further away from each other. Hamylton et al. (2015) also found that regions with sandy bottoms were predicted accurately, while deep regions were constantly under-predicted, as light does not reflect from the bottom due to absorption. The spatial error maps of the coastal sites in Greece, which have varying sea bottoms from bright sand to dark algae and seagrass, showed geographical and depth-related patterns of over- and under-estimation over distinct sea bottoms (Traganos, et al., 2018). The empirical algorithm was more likely to under-predict

in shallow waters over sandy seabeds and the algae-covered deeper waters of the Apokoronas area. Over-predicted clusters were found in shallow parts of the Apokoronas area that are algae habitats and in the western and easternmost parts of the Thermaikos Gulf with seagrass bottoms (Traganos, et al., 2018). Therefore, it can be concluded that the accuracy of empirical algorithms depends on the geographic location and physical properties of the water column.

3.1.1 Summary

These studies have shown that SDB using empirical approaches is only accurate in shallow coastal zones as the number of photons absorbed by the water column increases with increasing depth, resulting in a few photons reaching the bottom to be reflected (Traganos, et al., 2018). Therefore, the water reflectance signal detected by the satellite sensor does not contain any information about the sea bottom or water depth, which causes the empirical algorithms to under-predict. Furthermore, from the spatial distribution of the algorithms' errors, it is evident that the water reflectance signal is affected not only by depth but by the bottom character and the scattering and absorption effects of the water column (Traganos, et al., 2018). Bright, sandy, and rocky bottoms can strongly reflect light, which underestimates depth in these waters. In contrast, dark sea bottoms, like algae- and seagrass-covered bottoms do not reflect much light and hence lead to the over-estimation of these waters (Traganos, et al., 2018). As empirical algorithms assume that the water's bottom character and optical properties are uniform across the study area, the true relationship between water depth and the reflectance from the water surface is not modelled (Ilori & Knudby, 2020). Therefore, the physical properties of water need to be considered when estimating water depth.

3.2 Physics-based studies

The estimation of the depth in the Florida Keys waters in the USA by Ilori and Knudby (2020) involved using Landsat-8 imagery, LiDAR in-situ depth data, and a database of $R_{rs,modelled}$ spectra with the associated water depth, seafloor reflectance, environmental conditions, and inherent

optical properties (Ilori & Knudby, 2020). Multitemporal images from May 2013 to May 2019, that were captured during optimal and sub-optimal conditions, were used in an ensemble approach to enhance the results of SDB by finding the per-pixel median depth value of all images (Ilori & Knudby, 2020). The $R_{rs,observed}$ from the first four bands of the Landsat-8 imagery was first atmospherically corrected using ACOLITE, which is an aquatic-based atmospheric correction algorithm that removes the effects of sun-glint, scattering, and noise in satellite imagery (Ilori & Knudby, 2020). $R_{rs,modelled}$ was synthesized using a collection of sand and algae-covered seafloor reflectance, IOPs, and depths, ranging from 0.1 to 25m. The $R_{rs,modelled}$ were then discretized to form a uniform representative spectral space (Ilori & Knudby, 2020). Thereafter, the $R_{rs,modelled}$ were inverted using a semi-analytical approach, the binary space partitioning approach, which recursively divides the spectral space into two nodes to find the optimum $R_{rs,modelled}$ value that best matches the $R_{rs,observed}$ value (Ilori & Knudby, 2020). After depth has been retrieved, it is corrected for tides and the final results are validated by comparing against the LiDAR in-situ data.

Ilori and Knudby (2020) found that the ensemble approach could achieve an RMSE value of 1.34m for the 0.1 – 25m depth range and an RMSE of 1.21 for the 0.1 – 12m depth range (Ilori & Knudby, 2020). These results are comparable to those from the best single image. However, the ensemble approach eclipses the single image approach as by using the median depth, the sensor-environment noise ($NE\Delta R$) in satellite images that was not removed through the atmospheric correction is eliminated (Ilori & Knudby, 2020). This approach also has some computational benefits, since by selecting multiple images with optimal and sub-optimal conditions, the time spent on visually inspecting and filtering through images is reduced (Ilori & Knudby, 2020). The results further show that the accuracy of SDB depreciates with increasing depth, as previously discovered with the empirical investigations. Beyond 15m, Ilori and Knudby (2020) observed that the percentage of the radiance detected that was reflected from the seafloor is infinitesimal.

Other studies have capitalized on the availability of multi-temporal images of the same area, such

as the study of Boundary Bay, Hudson Bay, and Baffin Bay in Canada (Ilori, 2020). The Landsat-8 satellite has a revisit time of 16 days, which offers a wide selection of images captured during different environmental conditions for SDB applications. Ilori (2020) used 54 image scenes, which were atmospherically corrected using ACOLITE to remove the effects of atmospheric attenuation and sea surface specular reflection. The atmospheric correction retrieved the $R_{rs,observed}$ used in the subsequent physics-based approach. In addition to in-situ bathymetric data, water-quality maps containing surface condition information and water column optical properties, and benthic habitat maps showing different seafloor types in the study area were included in the investigation (Ilori, 2020). With representative clear water, solar zenith and azimuth values, these were parameterized and used to forward model the $R_{rs,modelled}$ spectra, which were then stored in a LUT (Ilori, 2020). However, as the previous study by Ilori and Knudby (2020) revealed that atmospheric correction does not remove $NE\Delta R$, Ilori (2020) added sensor-environment noise to the forward-modelled $R_{rs,modelled}$ to ensure that the simulated spectra represent what is observed. 20 random $NE\Delta R$ values were appended to each spectral band to create noise-bothered spectra (Ilori, 2020). The spectra were then discretized by recursively dividing the spectral space into equal parts through binary space partitioning (Ilori, 2020).

After finding the spectra that best match $R_{rs,observed}$ and inverting the spectra, Ilori (2020) found that the physics-based model can retrieve depths up to 4.5m in clear waters and 3m in turbid water. These depths were predicted with less than 1m error at a 95% confidence level (Ilori, 2020). By adding noise to the simulated spectra, it was evident that noise significantly affected RTM inversion. There was a greater overestimation of water depth in parts of Boundary Bay, a turbid water body, which points to the impact of noise in SDB predictions. The prediction results were more precise in Hudson and Baffin Bay, which have clearer water, as spectral signatures at different depths are distinguishable (Ilori, 2020).

An earlier study of Florida Bay by Lee et al. (1999) explored ways in which the RTM could be

optimized using a hyperspectral scanner, which is a sensor that measures the spectral reflectance from the water surface and separates it into fine wavelength channels (Lee, et al., 1999). For this optimization, it was noted that even with homogeneous IOPs, the path radiance detected from deep water is not the same as that from shallow water (Lee, et al., 1999). Secondly, due to the dynamic nature of coastal regions, the optical properties of the water column will change, especially between satellite image acquisition times (Lee, et al., 1999). For this reason. The semi-analytical RTM was applied to both the computer-simulated and field-measured spectra. To create a computer-simulated set of spectra ($R_{rs,Hydrolight}$), different values of wind speed, solar zenith angle, depth, chlorophyll-a concentrations, seafloor reflectance, and other IOPs were inputted into Hydrolight. This software program models the RTM (Lee, et al., 1999). The field measurements included water depth, chlorophyll-a, and the hyperspectral T_{rs} and S_{rs} , defined below.

$$T_{rs} = \frac{L_u}{E_d} \quad , \quad S_{rs} = \frac{L_{sky}}{E_d} \quad (3.5)$$

where L_{sky} is the radiance scattered by the atmosphere, S_{rs} is the surface-reflected remote sensing reflectance, and T_{rs} is the total remote sensing reflectance, which includes R_{rs} , S_{rs} , and sun-glint effects (Lee, et al., 1999). The $R_{rs,Hydrolight}$ and $R_{rs,observed}$ were compared against the forward-modelled $R_{rs,modelled}$, and after finding the best match, the difference between the two spectra was reduced by altering the model's absorption coefficient (a), backscattering coefficient b_b , depth (H), and bottom albedo (ρ) values using a repetitive predictor-corrector process (Lee, et al., 1999). Once the difference was minimized, the $R_{rs,modelled}$ was inverted to obtain the water depth value and the IOPs. These parameters (Q_i^{der}) were compared to the measured parameters (Q_i^{mea}) and the retrieval error was measured for each configuration of the model settings, as shown in Equation (3.6) (Lee, et al., 1999).

$$\delta = \exp[mean|\ln(\frac{Q_i^{der}}{Q_i^{mea}})|] \quad (3.6)$$

For the computer-simulated $R_{rs,Hydrolight}$, the retrieval error for water depth ranging from 2 to 20m

was 5.3%, with a few overestimations for deeper waters (Lee, et al., 1999). Whereas, for the field-measured $R_{rs,observed}$, the error was 10.9% for depths ranging from 0.8 to 25m, with an R^2 value of 0.96 (Lee, et al., 1999). In more turbid waters with a complex seafloor, the depth difference was larger due to the high backscattering coefficient of the water (Lee, et al., 1999). A conclusion was made that the semi-analytical RTM is more suited for coastal waters with a typical phase function and fails to model more complex waters, such as Florida Bay. Spectral parameters, such as the spectral slope, S , and spectral shape Y , are not accurately modelled by the RTM as the parameters cannot be estimated from the measured remote sensing reflectance (Lee, et al., 1999). Without accurate estimates of the spectral slope and shape, the IOPs cannot be modelled effectively to simulate $R_{rs,modelled}$ that closely matches $R_{rs,observed}$ (Lee, et al., 1999). However, the retrieval error indicates that the physics-based model can predict water depth in shallow waters from remote sensing measurements.

3.2.1 Summary

The studies have shown that the RTM model can estimate depths in shallow regions below 5m (Ilori, 2020). The RTM assumes that the portion of $R_{rs,observed}$ from the above-surface reflectance is error-free. Therefore, the physics-based approach depends on the accuracy of the spectral reflectance measured (Ilori & Knudby, 2020) and behaves poorly when noisy reflectance data from satellite imagery is inverted to retrieve depth. Atmospherically correcting the raw signal and extracting the remote sensing reflectance is a challenge as approximately 10% of the reflectance detected contains information from the water column and seafloor (Ilori & Knudby, 2020). In addition to noise, the atmospheric correction does not account for waves, wind speed, and the solar zenith angle. Therefore, SDB is still susceptible to such environmental effects (Lee, et al., 1999). Thus, as a workaround, sensor-environment noise is added to the forward-modelled $R_{rs,modelled}$ to create a synthetic dataset that would more likely represent $R_{rs,observed}$, as done by (Ilori, 2020). In addition, this approach allows the effect of noise in SDB applications to be evaluated and quantified. Another approach to reduce the impact of noise is to use multi-temporal satellite images in

an ensemble approach. As Ilori and Knudby (2020) calculated the per-pixel median depth value across multiple images, the $NE\Delta R$ was eliminated, and the results were comparable to those from the single best image.

The empirical and physics-based models can predict water depth from satellite imagery with acceptable accuracy, according to the International Hydrographic Organization (2023). However, many assumptions are made, such as homogeneous water properties across the study area, water properties remaining constant over time, and the linear relationship between water depth and linearized reflectance. However, noise makes this relationship no longer linear (Manessa, et al., 2020). In addition, the dynamic nature of coastal regions presents a challenge for deterministic model approaches to predict water depth, which must constantly be recalibrated for different study areas. Therefore, due to these challenges, data-driven models need to be investigated to assess whether SDB estimation can be improved by analyzing and interpreting historical reflectance data to build a functional algorithm that can classify satellite imagery and predict water depth. These data-driven models are built using machine learning and deep learning algorithms, which are detailed in the following sections.

3.3 Machine Learning models: Random Forest

In this section, studies that employ machine learning models, such as the Random Forest, are reviewed. Each optimization approach of the Random Forest will be assessed by reviewing related investigations.

3.3.1 Random Forest without Optimization

The water depth of the coasts off Ganquan Island and Oahu Island in the Pacific Ocean was predicted from Landsat-8 and Sentinel-2 imagery using ML models and bathymetric in-situ LiDAR data (Duan, et al., 2022). Duan et al. (2022) compared the performance of the Random Forest

(RF) to other ML models, such as the Support Vector Machine (SVM) and Multilayer Perceptron (MLP). The default parameters of the RF were used with 100 decision trees and the RGB bands of the image were used as the only predictors (Duan, et al., 2022). The satellite imagery was radiometrically corrected to convert the images from digital number (DN) values to top-of-atmosphere (TOA) radiance. Thereafter, the images were atmospherically corrected to retrieve the surface reflectance and were further processed to remove sun-glint and separate land from the water (Duan, et al., 2022). As ML models are built using historical data and do not consider the radiative transfer model, they are sensitive to noise (Duan, et al., 2022). Therefore, a mean filter with a 3x3 window size was applied to the images to reduce the effect of noise in the training process.

All the models were reliable up to depths of 15m; beyond this, depths were significantly overestimated. It was found that there was a positive correlation between depth and the prediction error (Duan, et al., 2022), implying that the reflectance of deeper water cannot be accurately modelled. This is due to the increased absorption and scattering of light in the water column which creates noise in the signal detected by the satellite sensor resulting in very little bottom reflectance information being captured (Manessa, et al., 2016). The LiDAR test dataset was used to validate the predictions from the ML models and the RMSE, bias, and R^2 values were calculated for each island. For Ganquan Island, all models obtained an R^2 value greater than 0.85, indicating a high correlation between the predicted and in-situ depth data. The MLP performed best with an R^2 -value of 0.95, a bias of 0.123m, and an RMSE of 0.985m due to its ability to simulate the nonlinearity of light in water and adjust to variations in the reflectance caused by the water column and seafloor heterogeneity (Duan, et al., 2022). The performance of the RF was significantly lower with an R^2 -value of 0.908, a bias of 0.104m, and an RMSE of 1.337m and Duan et al. (2022) stated that this drop in accuracy is caused by the model's large prediction error in depth ranges of 0-5m and 15-20m. This was supported by the visual inspection of the model's accuracy by extracting the profiles, revealing that the predictions diverge from the reference profiles where there is an undulating seafloor (Duan, et al., 2022). However, the ML models, particularly the RF model, performed better on

Oahu Island. All models obtained R^2 values higher than 0.92, revealing that the estimated depths are consistent with the reference depths. The RF model produced extremely accurate predictions in this study area with an RMSE of 0.792m, a bias of 0.015m, and an R^2 of 0.981, and the extracted profiles closely matched the reference profiles. The performance improvement was likely due to increased training data from the study area (Duan, et al., 2022). Oahu Island also appeared to have smooth terrain, as the accuracy of RF was constant across the 25m depth range. Like Ganquan Island, the ML models were prone to underpredicting the 20-25m depths. Duan et al. (2022) concluded from the study that SDB can be reliable to a depth of 15m, as beyond this depth, the errors increase significantly.

Overall, the ML models were more accurate in predicting depth than the empirical methods and the inversion model as the complex rate of the attenuation of light in water can be approximated through their nonlinear regression capabilities (Duan, et al., 2022). During the study, the accuracy of the bathymetric predictions from Landsat-8 and Sentinel-2 imagery was also compared and Duan et al. (2022) found that predictions from the latter, with a 10m resolution for the multispectral bands, were considerably more accurate than those from Landsat-8, with a 30m resolution. This motivates the use of Sentinel-2 imagery for SDB tasks as the higher resolution and shorter revisit time of 10 days offer more satellite data to be collected and used for training. The smaller pixel size will also capture more details of the water surface's fluctuations and seafloor (Duan, et al., 2022). Duan et al. (2022) The investigation of Ganquan and Oahu Island also revealed that it is challenging for ML models to make accurate predictions over the heterogeneous landscape of coastal zones. In contrast, empirical and physics-based models are more robust (Duan, et al., 2022). This is likely due to the lack of optimization of the model for specific problems. Without optimization, the RF fails to model the complexities specific to SDB.

3.3.2 Hyperparameter Tuning

As the investigation of Ganquan and Oahu Island did not involve any optimization approach, the effect of hyperparameter tuning on the performance of RF will be explored from the study of Tavsan Island in the Sea of Marmara, Turkey, by Çelik et al. (2023).

Estimating the water depth of Tavsan Island proved challenging as the island's coast is a wave-dense environment (Çelik, et al., 2023). Waves add more complexity to SDB due to the interferences and variations of the water surface, which also cause diffuse reflection back to the satellite sensor (Çelik, et al., 2023). However, the RF model has been found to manage this nonlinearity as the model can compensate for missing values, and outliers can be uniquely calibrated for each bathymetric environment through hyperparameter tuning. To ensure highly accurate depth predictions, WorldView-2 level LV2A imagery, with a resolution of 1.84m and 8 multispectral bands, was used (Çelik, et al., 2023). The images were corrected for sun-glint, and the land was separated from the water. To minimize the effects of the waves in the image, the median filter was applied with a 5x5 window to create a smoother image with less interference. These images and in-situ depth data, collected by single-beam SONAR, were then used to train the RF regression model. Using level LV2A images that have been atmospherically corrected, the pixel values contain normalized surface reflectance values that are referenced to ground-truth reflectance values, ensuring the validity of the satellite image (Çelik, et al., 2023).

The RF model was built using three predictors, the Blue, Yellow, and NIR-1 bands as they are weakly correlated to each other but strongly correlated to water depth (Çelik, et al., 2023). 50% of the dataset was used to train and validate the RF model, while the remaining 50% was used for testing. The model's performance was then manually fine-tuned by adjusting the hyperparameters, one of the most important being the number of trees (n estimators). After some runs, the number of trees was set to 500 (Çelik, et al., 2023). Other studies, such as the land use classification of the City of Trabzon in Turkey, adjusted more hyperparameters of the RF, such as n estimators and the

m number of variables used to split each node of a decision tree (Akar & Gungor, 2012). Firstly, a decision tree from each N sample was built by randomly selecting m predictors, and the best split among these predictors was used (Akar & Gungor, 2012). In the automatic tuning of the RF, m is equal to the square root of the total number of predictors, M. To manually choose the best split, the GINI index was used, which measures the class homogeneity or “purity” of a node (Akar & Gungor, 2012). The lower the index, the more homogeneous the node is. A split was considered good if the GINI index of the child node was lower than that of the parent node and was terminated once the index was equal to zero (Akar & Gungor, 2012). This ensures that the m number of variables selected has a low correlation with each other but has a high predictive capacity (Akar & Gungor, 2012). This, in a sense, was achieved by (Çelik, et al., 2023) as the Blue, Yellow, and NIR-1 bands were selected due to their low intercorrelation but high correlation to depth.

Çelik et al. (2023) used the R², RMSE, and Mean Absolute Error (MAE) to evaluate the accuracy of the RF model. The water depth of Tavsan Island was predicted over a 15m range and Çelik et al. (2023) found that the accuracy of the RF model varied significantly over the range of depth but achieved an R² value of 0.85. There was an increase in accuracy with an increase in depth from 0 to 6m as the RMSE decreased from 1.91m at 0m to 0.93m at 6m, while the MAE decreased from 1.30m to 0.65m for the same depth range (Çelik, et al., 2023). However, the accuracy decreased with increasing depth beyond 6m, as the RSME increased from 0.93m to 2.41m and the MAE increased from 0.65m to 1.86m from the range of 6m to 15m (Çelik, et al., 2023).

Hyperparameter tuning is an appropriate approach to optimize the Random Forest for SDB as the hyperparameters can be fine-tuned to fit the complex relationships within the dataset and improve the generalization of the model (Çelik, et al., 2023)

3.3.3 Selection of Predictors

The capability of the RF model to simulate nonlinear patterns allowed Manessa et al. (2016) to predict the bathymetry of the Gili Mantra Islands and Panggang Island in Indonesia from Worldview-2 imagery. The Worldview-2 satellite captures high spatial resolution imagery at sub-2m resolution in 8 multispectral bands, six of which are in the visible range and two in the near-infrared range (Manessa, et al., 2016). Ground truth data was collected using a single-beam echosounder and the Differential Global Positioning System and was corrected for tides. The RF model was optimized using 63 different band combinations, as shown in Table 3.1.

Table 3.1: Different band combinations as predictors

Number of bands	Number of combinations
1	6
2	15
3	20
4	15
5	6
6	1

Source: Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T., and Higuchi, T. (2016). Satellite-derived bathymetry using random forest algorithm and worldview-2 imagery. [*Geoplanning: Journal of Geomatics and Planning*, 3(2):117–126..

Firstly, the pixel values of the satellite images were converted from DN to TOA radiance. The images were further corrected for atmospheric noise using Lyzenga’s correction model. The reflectance data was then logarithmically transformed before being used as inputs in the RF model (Manessa, et al., 2016). The model was used to predict depth through nonlinear regression, where individual decision trees take on numerical values rather than class labels (Manessa, et al., 2016). The performance of each combination of predictors was tested using a cross-validation method and was compared to determine the optimum combination. 10% of the ground truth data was randomly selected at each iteration to train the RF model and the remaining 90% was used as test data. After 100 iterations had been run for each predictor combination, the accuracy was assessed using RMSE and R^2 (Manessa, et al., 2016).

Due to the dissimilar water properties of the Gili Islands and Panggang Island, the combination of predictors that performed best was different. Manessa et al. (2016) found that the performance of the RF model did not depend on the use of all visible wavelength bands but on the use of four or fewer visible bands. For the Gili Islands, the best combination was with four visible bands (Coral, Green, Yellow, and Red), achieving an RMSE of 0.746m and R^2 of 0.874, while for Panggang Island, three visible bands (Green, Yellow, and Red) were sufficient for high accuracy, with an RMSE of 1.747m and R^2 of 0.858 (Manessa, et al., 2016). The accuracy of the RF model was not enhanced by including all visible bands as some bands do not contain useful information for depth prediction but contain more noise, such as the Red-Edge band in deeper water (Manessa, et al., 2016). Better performance was associated with the green, yellow, and red bands due to the rate at which these wavelengths are absorbed in water. Despite the strength of the coastal and blue wavelength bands to penetrate deeper into the water column, their inclusion hardly improved the accuracy of the RF model as they contained noise.

The predictions from the RF algorithm were validated against the results from Lyzenga's empirical multiple linear regression (MLR) algorithm. Both algorithms produced precise predictions in the shallow waters of the Gili Islands with a depth of less than 7m. In deeper waters, the RF model consistently under-estimated depth, most likely due to the increase in noise from increased absorption and scattering, while the MLR model over and under-estimated depth (Manessa, et al., 2016). There was a slight difference in performance between the RF and MLR models, with the RMSE of the RF model being 0.14m lower and the R^2 being 0.16 higher than that of the MLR model (Manessa, et al., 2016). The difference in performance between the RF and MLR models was more pronounced for Panggang Island. The predictions from RF were accurate across the entire depth range, with a few overestimations in the shallow water, while the MLR algorithm produced poor estimations. The RF also had an RMSE 1.27m lower and a R^2 value 0.47 higher than that of the MLR model (Manessa, et al., 2016).

The stark difference in performance between the Gili Islands and Panggang Island could only be ascribed to the nonlinear behaviour of light in the water column, which is difficult to model in two different water bodies. Though the reflectance data was logarithmically transformed to create a linear relationship between water depth and reflectance, the presence of noise and other inherent properties generate nonlinear conditions (Manessa, et al., 2016). Particularly in deep waters, the signal received by the satellite sensor contains more information about the water column and less about the bottom reflectance due to the absorption and scattering by the water column. This noise limits the range of depth that can be predicted by the RF model. Dark objects in shallow water and bright objects in deeper also contribute to poor estimations as the reflectance values are similar. Manessa et al. (2016) concluded that the results could be improved by tuning some hyperparameters, mainly the sampling size, node size, and the number of randomly sampled predictors for each tree. Therefore, the Random Forest algorithm cannot be optimized solely through hyperparameter tuning or the selection of best predictors but with a combination of both.

3.3.4 Combined Optimization

In another study of the Gili Islands in Indonesia, the bathymetry was estimated using SPOT-7 multispectral imagery and employed a combined optimization approach to produce accurate predictions. SPOT is an earth observation satellite that captures images at a high spatial resolution of 6m (Manessa, et al., 2020). The four bands in the visible part of the spectrum (Blue, Green, Red, and Near-Infrared) were extracted and used as predictors. To create high-dimensional training data for the RF, Manessa et al. (2020) increased the number of predictors by creating modified spectral data from the existing multispectral bands. A total of 15 predictors ($3X_i + 3\hat{X} + 3\hat{Y} + 3\hat{Z} + 1X_{ratio} + 2Y_{ij}$), where X_i is the pseudo-depth (x_i) defined in Equation (3.1), were used and are defined in Table 3.2.

Different permutations of the predictors were tested to determine which predictors are necessary for water depth estimation. This was done through a cross-validation process where 70% of the data was used to train the model and 30% of the data was set aside for testing (Manessa, et al.,

Table 3.2: Modified predictors

Modified variable	Equation
Band ratio	$X_{ratio} = \frac{X_{blue}}{X_{green}}$
Modified linearized reflectance	$\hat{X}_i = \ln(V'_i - a_0 - a_1 V'_{NIR})$ $\hat{Y}_i = \exp(\hat{X}_i)^{-1}$ $\hat{Z}_i = V'_{NIR} \exp(\hat{X}_i)^{-1}$
Depth-invariance index	$Y_{ij} = X_i - \frac{k_i}{k_j} X_j$ where $\frac{k_i}{k_j}$ = attenuation coefficient derived from bands i and j

Source: Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T., and Higuchi, T. (2016). Satellite-derived bathymetry using random forest algorithm and worldview-2 imagery. *Geopanning: Journal of Geomatics and Planning*, 3(2):117–126.

2020). The performance of the model, with its combination of predictors and automatic tuning, was evaluated using R^2 and RMSE. Manessa et al. (2020) found that by increasing the number of predictors from 3 to 15, the accuracy improved notably, from an R^2 of 0.95 to 0.97 and RMSE of 1.45m to 1.1m. This preliminary test revealed that building an RF model using a few predictors produces imprecise results and indicates that more predictors in the model can identify intrinsic patterns in the data. The effect of hyperparameter tuning on the model's performance was also tested separately. To optimize the model, the node size, sampling size, and $Mtry$ were manually tuned using tuneRanger to find the optimal values (Manessa, et al., 2020). The hyperparameter optimization improved the performance of the model with just three predictors from an R^2 of 0.95 to 0.97 and a RMSE of 1.45m to 1.03m (Manessa, et al., 2020).

Optimizing the RF model solely through hyperparameter tuning has a much greater effect on the accuracy than increasing the number of predictors. However, by using the combined approach of hyperparameter tuning and increasing the number of predictors, the model produced highly accurate water depth predictions. By manually tuning the model and using 15 predictors, the accuracy of the model increased and an R^2 of 0.98 and RMSE of 0.78m was achieved (Manessa, et al., 2020). Therefore, it can be concluded that ML models, such as the Random Forest, can be optimized for

extracting water depth from multispectral satellite imagery.

3.3.5 Summary

The Random Forest machine learning model has become a powerful tool for SDB as it is an ensemble learning technique which uses numerous decision trees to make a prediction. In addition, the model can be uniquely calibrated or optimized for specific bathymetric environments and conditions and data availability levels.

The study of Tavsan Island by Çelik et al. (2023) and of the Gili Mantra Islands and Panggang Island by Manessa et al. (2016) proved that the use of the RF and multispectral imagery provides the liberty to select the best combination of predictors or bands for modelling water depth. However, the selection of predictors must be well advised as not all multispectral bands of a satellite image are essential for modelling water depth and the best predictors may be different for distinct study areas. Both studies also showed the necessity of studying the intercorrelation between bands as certain depth ranges may have similar spectral signatures which are difficult to differentiate between and classify (Akar & Gungor, 2012). The Green and Red bands were excluded from the Tavsan Island study due to their strong correlation to the blue band and so were the NIR-2 and red-edge bands for their strong correlation to the NIR-1 band (Çelik, et al., 2023). However, with hyperparameter tuning, the RF can differentiate between target classes with similar spectral signatures.

To optimize the model hyperparameter tuning is applied to find the best set of model parameters to build the RF (Akar & Gungor, 2012). The parameters to be optimized typically include the number of estimators, the maximum depth of the tree, minimum samples for a node split, and the minimum number of leaves.

3.4 Deep Learning models: DCNNs

Deep Convolutional Neural Networks (DCNNs) have been used for SDB to perform pixel-wise segmentation and classification using labelled training data (Wilson, et al., 2020). To ensure CNNs can recognize and classify image features, large, diverse and representative labelled datasets are required for training. Al Najar et al. (2023) mentions that labelled satellite imagery is difficult to obtain but can be supplemented with masks created from in-situ data. Multi-temporal satellite images and in-situ bathymetric data from a range of bathymetric and environmental conditions are collected to increase the size and diversity of the dataset for training the models to make accurate predictions in different types of water conditions (Al Najar, et al., 2023). Many variations of the CNN are created based on the configurations of the hidden layers and algorithms, creating unique CNN architectures. Different architectures are specific to certain tasks, and it is necessary to evaluate the performance of several architectures in estimating water depth from satellite imagery (Lumban-Gaol et al., 2021). These architectures can also be customized for specific problems and datasets by regularization.

3.4.1 Semantic Segmentation

An investigation of the Devils Lake Area in the USA involved deep learning techniques to estimate depth from Sentinel-2A imagery using a fully convolutional neural network with an encoder-decoder architecture (Wilson, et al., 2020). This type of CNN architecture is also known as SegNet, designed for semantic segmentation tasks.

The Sentinel-2A imagery was labelled using depth measurements from the Humminbird GPS SONAR device (Wilson, et al., 2020). The depth measurements were transformed into a raster image which was then registered to the satellite images to be subdivided into 10x10 image tiles. The satellite imagery was tiled into 38 patches, where 27 were used for training, 3 for validation, and 8 for testing (Wilson, et al., 2020). While training the model, the input tiles were further subdivided

into smaller tiles, using a sliding window, to guarantee that the minute details in the satellite image were identified and mapped to the corresponding depth data. The training dataset was randomly grouped in batches of size 16 for each epoch and the model was trained for 50 epochs and was evaluated after each epoch. Wilson et al. (2020) used the cross-entropy loss function, which measures the difference between the target values and the predicted values, to monitor the performance of the model. The loss was then propagated backwards through the model to update the weights for the next forward pass (Wilson, et al., 2020). The weights were updated using the stochastic gradient descent as an optimizer, whose parameter values are summarised in Table 3.3. To ensure the model was not overfitting and could generalize on new data, the model was regularized using an L-2 norm weight decay, where a penalty is added to the loss function to reduce the model's large weights (Wilson, et al., 2020).

Table 3.3: Optimizer specifications

Stochastic gradient descent optimizer	
Momentum	0.9
Learning Rate	0.01
Weight Decay	0.001

Source: Wilson, B., Kurian, N. C., Singh, A., and Sethi, A. (2020). Satellite-derived bathymetry using deep convolutional neural network. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2280–2283. IEEE.

The accuracy of the classification was measured using the average F-1 score, which counts the number of actual negatives in each depth class and the ROC-AUC score which measures how successfully the model can differentiate between classes (Wilson, et al., 2020). The model achieved an F-1 score of 0.734 and a ROC-AUC score of 0.82 on the test data (Wilson, et al., 2020). The SegNet model could capture the inherent patterns in satellite imagery due to its architecture, which requires fewer trainable parameters and can manage inputs of variable sizes (Wilson, et al., 2020). However, the study used the three RGB bands from the Level-2A imagery and did not utilize all the multiple spectral bands that are available. This proves that deep learning models can be effective in predicting water depth from three-channel satellite imagery and can be improved by using multiple

channels in the classification process (Wilson, et al., 2020).

3.4.2 Multi-channel Training

Lumban-Gaol et al. (2021) demonstrated the use of all the available multispectral bands of Level-2A imagery in their study of Puerto Rico Island. Sentinel-2 Level-2A images, which store surface reflectance, were used as the inputs to the CNN. The images were labelled using MBES and LiDAR bathymetric data which were corrected for tides. The bathymetric data also included measurements above mean sea level (MSL) for the sake of training the CNN to distinguish between water and land and identify the coastline. Various combinations of nine spectral bands, shown in Table 3.4, were used to evaluate the significance of each band.

Table 3.4: Sentinel-2 Band combinations

Channels	Number of Channels	Bands
RGB	3	Red, Green, Blue
RGBN	4	Red, Green, Blue, NIR
RGBNSS	6	Red, Green, Blue, NIR, SWIR1, SWIR2
All bands	9	Red, Green, Blue, NIR, SWIR1, SWIR2, Red-Edge 1, Red-Edge 2, Red-Edge 2

Source: Wilson, B., Kurian, N. C., Singh, A., and Sethi, A. (2020). Satellite-derived bathymetry using deep convolutional neural network. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2280–2283. IEEE.

The best image scene was collected for the coast the Ponce area, the Southwest area, and the coast of San Juan. The Ponce and Southwest areas are reasonably turbid while the San Juan area is extremely turbid (Lumban-Gaol, et al., 2021). The images were subdivided into smaller tiles and 10,820 image tiles (80%) of these were selected for training and validation, with 2,705 image tiles (20%) set aside for testing (Lumban-Gaol, et al., 2021). However, different CNN architectures, all using the RGB combination of bands, were compared first to determine the optimal architecture to predict water depth. Each architecture was trained on the Ponce area of the island and the prediction

error was evaluated using the RMSE and R^2 . A variety of window sizes were implemented during the construction of the CNNs, and batch normalization was included to prevent the model from memorizing the patterns in the data and overfitting (Lumban-Gaol, et al., 2021). The Rectified Linear Unit (ReLU) function was used as the activation layer function to introduce non-linearity into the network, which would assist the CNN in modelling the complex relationship between water surface reflectance and water depth, and calculate the output from the input values (Lumban-Gaol, et al., 2021).

After several trials, Lumban-Gaol et al. (2021) found that the RMSE from all the architectures ranged from 1.48m to 1.94m for the RMSE and 0.89 and 0.94 for the R^2 coefficient, which indicates that the choice in architecture is significant to ensure accurate predictions. In this preliminary test, it was found that a CNN with one convolutional layer did not return accurate results, and therefore, the final CNN would require at least three convolutional layers as a deeper network displayed improvements in the predictions. A larger window size also improved the network's accuracy by approximately 10cm (Lumban-Gaol, et al., 2021). The architecture that was found to best predict water depth was the CNN with three convolutional layers and a pooling layer, which was referred to as CNN4, and a 9x9 window size (Lumban-Gaol, et al., 2021).

The CNN4 was then fine-tuned using different combinations of the learning rate, batch size and dropout rate, and the validation accuracy was assessed using the MAE and RMSE after each training epoch (Lumban-Gaol, et al., 2021). A learning rate of 0.0001, batch size of 521, and dropout rate of 0.3 were found to produce the best performance (Lumban-Gaol, et al., 2021). Different band combinations were then explored to assess any further improvement in accuracy from using bands other than the standard RGB bands. Lumban-Gaol et al. (2021) stated that the performance of the CNN SDB model is influenced by the number of bands or channels included in the input data. The RGB, RGBN, and RGBNSS bands, were found to produce the same prediction accuracy for the depth range 0-7m but deviated in waters between 1-2m and waters deeper than 17m (Lumban-

Gaol, et al., 2021). In the depth interval of 7-16m, the RGBNSS combination had a higher accuracy due to the addition of the SWIR bands. With more bands available for training, the model had a greater ability to estimate water depth in this range. There was not any significant improvement in accuracy by the inclusion of all nine spectral bands. This behaviour is expected as the study by Manessa et al. (2016) using the Random Forest excluded the Red-Edge bands which contain more noise than information on the bottom reflectance. Lumban-Gaol et al. (2021) concluded that the inclusion of all bands may cause the CNN model to have a bias in a certain depth range, which affects the prediction accuracy.

From the results of training on the Ponce area, the finalized CNN4 model, with the RGBNSS band combination, was used for the rest of the study and was validated against the empirical transform method. The empirical transform method achieved an RMSE of 4.05m and R^2 of 0.41 in the Ponce study area, while the CNN4 model achieved an RMSE of 1.31m and R^2 of 0.94 and Lumban-Gaol et al. (2021) noted that this was due to the empirical model's inability to estimate depth in very shallow water of 0 to 2m, due to the variations in bottom reflectance, and in water deeper than 10m. This further proved that CNNs are more capable of modelling the complex nature of light in water. However, though deep learning methods outperform empirical methods, they are still prone to errors in certain depth ranges, such as deep water, where the accuracy in prediction decreases with an increase in depth (Lumban-Gaol, et al., 2021).

This investigation of Puerto Rico Island revealed the advantages of using more multispectral bands in the CNN model as the SWIR bands stabilize the performance of the model in the 2-16m depth range (Lumban-Gaol, et al., 2021). Including more channels allows the model to perform better as certain channels or bands contain valuable information about the bottom reflectance, such as the SWIR bands, when other bands have been scattered or absorbed due to the depth of the water column. Be that as it may, Lumban-Gaol et al. (2021) noted that including too many bands may cause the model to have a bias to a particular depth range which will in turn not result in any significant

improvement in accuracy.

3.4.3 Model Transferability

Lumban-Gaol et al. (2021) tested the transferability of their pre-trained CNN4 model on the remaining areas around Puerto Rico Island. Regarding the Southwest area of the island, the model had an accuracy of 3.16m and an R^2 of 0.82 (Lumban-Gaol, et al., 2021). This is comparable to the Ponce area as both study areas are found to the south of the island and have comparable water properties. The similarity in performance is also due to the related spectral signatures as the same image was used for both study areas (Lumban-Gaol, et al., 2021). However, the model was not successful in accurately predicting the water depth of the San Juan area. The model obtained an RMSE of 4.93m and an R^2 of -0.17 (Lumban-Gaol, et al., 2021). The CNN4 model predictions had the opposite trend to that of the ground truth data which Lumban-Gaol et al. (2021) expected since the San Jaun area has more turbid conditions caused by port activity that increase the sediments in the water. These turbid conditions make it difficult for the DL models to predict water depth especially when these models have been previously trained on much calmer bathymetric conditions. Therefore, the pre-trained CNN4 model had to be retrained using the bathymetric data from the other study areas to make better estimations. After retraining the model using data from the Southwest area, the CNN4 model achieved an RMSE of 1.78m and an R^2 of 0.92 (Lumban-Gaol, et al., 2021). This was a considerable improvement in accuracy from the previous results as the RMSE improved by approximately 1.3m. The model was also retrained using data from the San Juan area and an RMSE of 3.03m and R^2 of 0.55 was obtained.

Lumban-Gaol et al. (2021) noted that these results prove the model is not transferable as it fails to accurately predict new areas with different bathymetric conditions. To ensure the transferability of DL SDB models, the CNNs need to be trained on a wide variety of bathymetric data from areas that are widely geographically distributed. Another approach is to use multitemporal satellite imagery with the corresponding tide-corrected bathymetric data to create a diverse set of reflectance

data that can be used for training (Lumban-Gaol, et al., 2021).

3.4.4 Summary

Predicting water depth from satellite imagery is a complex task due to challenges such as atmospheric conditions, sediment interference, and variations in image brightness. ML models are limited by their reliance on individual pixels, which neglects spatial relationships between adjacent pixels. CNNs, overcome these limitations by leveraging their ability to model complex, non-linear relationships, generalize across diverse datasets, and capture spatial dependencies between pixels.

From the studies by Lumban-Gaol et al. (2021) and Wilson et al. (2020) prove that deep CNNs (DCNNs) are highly effective in satellite-derived bathymetry (SDB), where the task is often treated as a pixel-wise classification problem. By utilizing large and diverse training datasets and incorporating multi-temporal imagery, the DCNNs can accurately map pixel reflectance values to water depth and generalize well on unseen data separate from the training dataset. While DL models exhibit strong performance, their transferability to new regions with differing bathymetric conditions can be limited. Models trained on specific datasets may struggle in regions with turbid water or high sedimentation levels. However, retraining the model using localized data from diverse bathymetric conditions can significantly improve accuracy.

Chapter 4

Methodology

In this investigation, satellite-derived bathymetry is treated as a classification task rather than a regression task, as the Radiative Transfer Model reveals that due to the inherent optical properties (IOPs) and different bathymetric conditions, such as turbidity and bottom character, a range of multispectral reflectance values can belong to a single depth point. In addition, the non-linear relationship between reflectance and depth and the presence of sensor-environment noise in satellite images, even after atmospheric correction, further complicate the task. Classification is a more robust approach where a range of multispectral reflectance values is mapped to predefined one-metre interval depth classes, effectively classifying depth classes from the satellite image.

For the classification of satellite images, the machine learning methods, Random Forest (RF) and the deep learning method (DCNNs), U-Net, SegNet, and DeepLabv3+, were selected for this investigation because of their ability to capture the complex relationship between multispectral reflectance and water depth. The RF model has been used in studies by Manessa et al. (2016) and Manessa et al. (2020), while U-Net was employed by Al Najar et al. (2023), and SegNet by Wilson et al. (2020). The DeepLabv3+ was included due to its success in land cover classification in the investigation by Gharahbagh et al. (2025). The RF model adapts to complex high-dimensional data without making assumptions about the distribution of the input reflectance data. With many op-

portunities to optimize the model through hyperparameter tuning, it is ideal for handling the complex nature of the task. U-Net efficiently retains spatial details through skip connections, which is crucial for identifying subtle boundaries between depth classes. Its ability to work with small datasets further supports its use in this investigation. To make the encoder-decoder architecture more memory-efficient, SegNet uses max-pooling indices during the decoding phase to preserve spatial resolution, making it suitable for a large-scale investigation. Finally, DeepLabv3+ enhances segmentation using Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale features, essential for detecting fine-grained and broader patterns in satellite imagery.

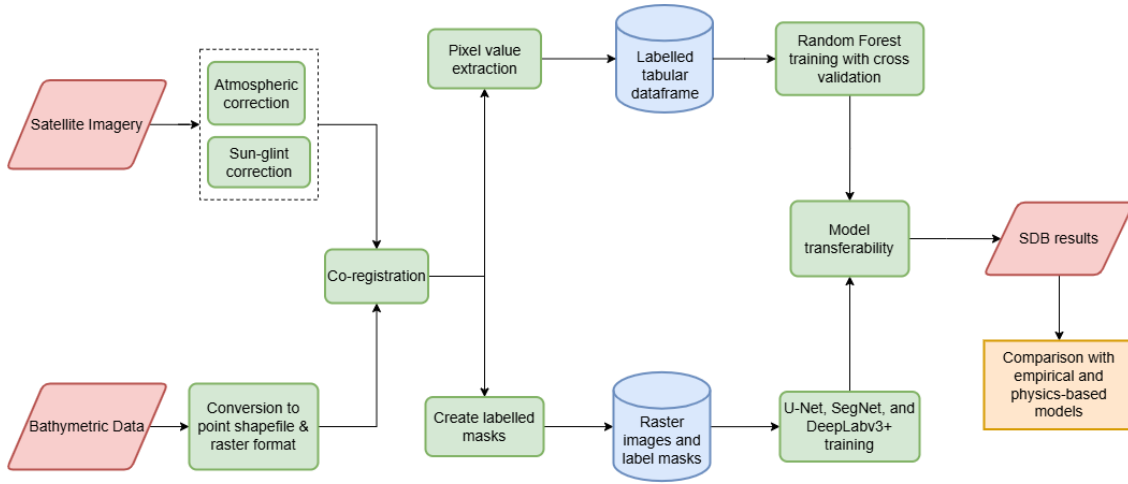


Figure 4.1: Methodology

Figure (4.1) illustrates the steps of the methodology implemented in our investigation. Firstly, bathymetric data and satellite imagery were collected and corrected for systematic errors. The bathymetric data from the Canadian Hydrographic Service (CHS) were corrected for tides (Section 4.2) and transformed into point shapefiles and raster datasets (Section 4.3). Sentinel 2A Level 1C satellite imagery was atmospherically corrected and sun-glint was removed to obtain the remote-sensing reflectance (Section 4.3). Both image and bathymetric datasets were time synchronized and then co-registered to ensure the bathymetric data aligned with the satellite image (Section 4.4). A labelled tabular data frame was created through pixel value extraction, and a dataset of 256x256 raster images and label chips was created through image mask generation (Section 4.4). These labelled datasets were then used to train the ML models to make water depth predictions from

satellite image data (Section 4.5). Once training was completed, transferability was tested for each ML model by evaluating performance on unseen data or data from three test sites not included in the training process (Section 4.6). Finally, the performance of all the ML models on the test sites was compared to the empirical and physics-based models (Section 4.7). The interactive web-based computing platform, Jupyter Notebook, was used to implement all the ML models.

4.1 Study Areas

As part of our contribution to SDB modelling, we investigated the performance of SDB modelling techniques in freshwater and saltwater and the combination of both salinity types. Freshwater models were trained and tested on freshwater areas, and saltwater models were trained and tested on saltwater areas. We selected study areas from diverse Canadian water bodies, concentrating on relatively calm coastal zones. The study areas are listed in Table 4.1.

Table 4.1: Study Areas

Freshwater	Saltwater
Wasaga Beach, Georgian Bay	Moresby Island, Hecate Strait
Long Point, Lake Erie	Graham Island, Hecate Strait
Grand Bend, Lake Huron	Prince Edward Island, Northumberland Strait
Point Clark, Lake Huron	Pugwash, Northumberland Strait
Athol, Lake Ontario	Tatamagouche Bay, Northumberland Strait
Wolfe Island, Lake Ontario	Rimouski Port, Saint Lawrence Seaway

Lake Ontario is a valuable resource for drinking water, industrial and recreational activities, and wastewater processing for the province of Ontario in Canada and New York in the USA (Makarewicz & Howell, 2012). The lake is oligotrophic and mesotrophic, with a maximum depth of 244m. Makarewicz and Howell (2012) report that the lake has moderate levels of nutrients, such as phosphorous which stimulates the growth of microscopic plants to cause algae blooms. However, due to human activity in the coastal zone of Lake Ontario, more nutrients are loaded into the water, encouraging eutrophication. Therefore, algae and CDOMs are expected to influence Lake Ontario's

spectral responses. The coastal region of Athol in Lake Ontario was selected as the study area due to the calm conditions of the lake and its distance from the source of the St. Lawrence Seaway, shown in Figure (4.2). Lake Erie is a shallow eutrophic lake, full of nutrients and has a maximum depth of 19m. Sekaluvu et al. (2018) also state that harmful algae blooms constantly characterize the lake. The seafloor comprises fine sediments stirred by strong waves and winds. The average sea surface temperature is higher than that of the other Laurentian Great Lakes, as it is the southernmost lake. These warm temperatures promote biological productivity (Sekaluvu, et al., 2018). Due to the biological and geomorphological characteristics of the lake, great variance in the spectral responses is expected in the investigation. Lake Huron is the second-largest of the Great Lakes, bordered by the province of Ontario and the state of Michigan. From measurements taken by Berst and Spangler (1972), they determined that Lake Huron is an oligotrophic lake, which means it has very low levels of nutrients. The average surface temperatures are lower than Lake Erie's and cannot support much biological productivity. It is a relatively deep lake with a maximum depth of 229m. Finally, the Georgian Bay is a bay within Lake Huron with the same characteristics. It is an oligotrophic bay, but due to human activity, the levels of nutrients are rising (Berst & Spangler, 1972). It has a maximum depth of 165m as it approaches Lake Huron. Due to the low levels of productivity and relatively calm conditions, the spectral responses from Lake Huron and the Georgian Bay are expected to be consistent.

The St. Lawrence Seaway is one of Canada's most important river systems as it supplies water to the population of Quebec, is a wildlife habitat, and is a major maritime transport corridor for the country (Vincent & Dodson, 1999). It stretches from the Laurentian Great Lakes to the Gulf of St. Lawrence. It can be categorized into the fluvial section, the fluvial estuary, the upper estuary, and the lower estuary (Vincent & Dodson, 1999). The lower estuary experiences heavy marine vessel traffic and is maintained to a depth of 13m. Vincent and Dodson (1999) found that strong tidal fluctuations and wave activity characterize the shoreline and contain some marshland, a habitat for vital aquatic life. With the combination of salt and fresh water, especially around the Rimouski

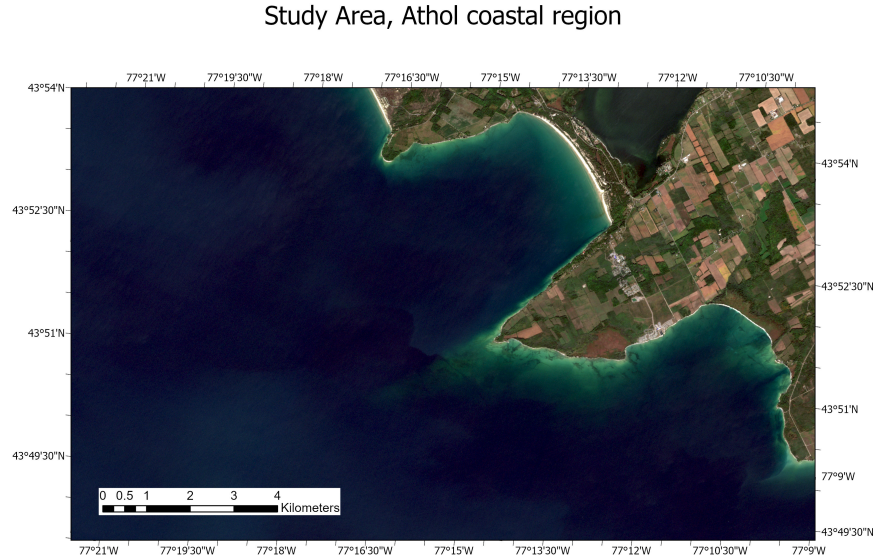


Figure 4.2: Freshwater test site in Lake Ontario

Port shown in Figure (4.3a), inconsistent spectral responses from this study area are expected. The Northumberland Strait is in the Gulf of Saint Lawrence and is susceptible to strong tidal currents and turbulence. It is a shallow strait with warm saline waters with a maximum depth of 68m. Measurements by Kranck (1972) revealed that the strait consists of red silt and clay sediments. The tidal currents stir up the sediments, so it should be expected that the spectral signatures from this study area will be highly variable. The Hecate Strait, shown in Figure (4.3b), is a shallow saline strait that borders the western province of British Columbia and the Queen Charlotte Islands in Canada. Crawford et al. (1988) found the maximum depth of the strait to be 300m, and waters reached an average surface temperature of 13°C in summer. The strait also experiences large tidal fluctuations as much as 6m and rotary tidal currents, which may affect the spectral responses.

Data was collected from all the study areas and separated for training and testing model transferability. Due to intense cloud cover over saltwater coastal zones in higher latitudes, more data was collected from the Northumberland Strait to match the quantity of data collected from freshwater environments. Table 4.2 shows the allocation of study areas for training and the testing of model transferability. Lake Ontario's coastal region of Athol was selected as the site to test the trans-

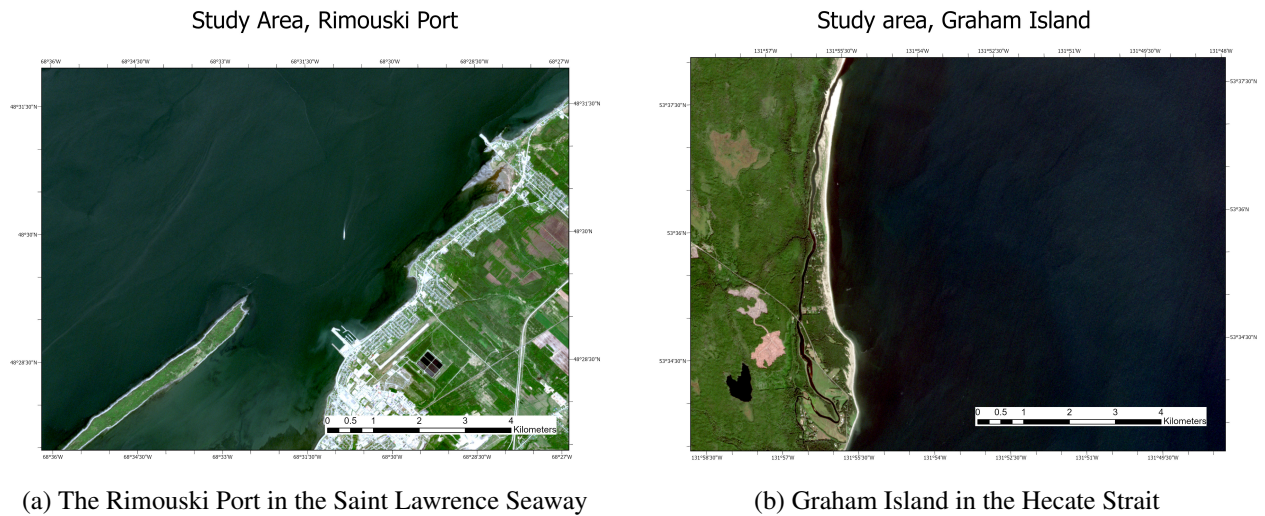


Figure 4.3: Combination (a) and pure saltwater (b) test sites

ferability of freshwater models. Graham Island, a pure saltwater environment with a steep seabed in the Hecate Strait, and the Rimouski Port, a combination of fresh and saline waters in the Saint Lawrence Seaway, were also selected as sites to test the transferability of saltwater models. Data from these areas was not included in the model training. The empirical and physics-based models were also tested on the same test sites to compare the performance of ML models against traditional methods.

Table 4.2: Study Areas for training and testing model transferability

	Freshwater	Saltwater
Modelling	Wasaga Beach, Georgian Bay Long Point, Lake Erie Grand Bend, Lake Huron Point Clark, Lake Huron Wolfe Island, Lake Ontario	Moresby Island, Hecate Strait Prince Edward Island, Northumberland Strait Pugwash, Northumberland Strait Tatamagouche Bay, Northumberland Strait
Transferability	Athol, Lake Ontario	Graham Island, Hecate Strait Rimouski Port, Saint Lawrence Seaway

4.2 Datasets

In-situ bathymetric data that covers the study areas was collected from the Canadian Hydrographic Service (2022) (CHS) Non-Navigational Bathymetric data portal. CHS NONNA-10 packages were collected, containing four data formats that can be used in various software programmes. The formats provided were 32-bit GeoTIFF, ASCII++ (XYZ), CARIS Spatial Archive (CSAR), and Bathymetric Attributed Grid (BAG). The ASCII++ (XYZ) is a gridded raster dataset that contains X and Y coordinates and Z (depth) measurements in metres, at a spatial resolution of 10m, made using traditional methods from 1976 through to the 30th of September 2019. The features of the ASCII table include Longitude and Latitude in DMS, Depth in metres, and Contributor, which is the ship ID and date of measurement. The bathymetric data provided by CHS are referenced to the Chart Datum. According to Dohler (1968), the Chart Datum is a low-water datum that ensures that depths measured rarely fall below it.

As the most recent CHS data was acquired on the 30th of September 2019, multitemporal satellite imagery captured around this time was acquired. Three satellite images from each study area were acquired during the 2016 to 2023 time span (Appendix C). Images captured between May and October were collected to avoid performing SDB on images with ice present. Sentinel imagery, which consists of 13 multispectral bands (Table 4.3), was acquired from the European Space Agency (2014a) Copernicus Open Access Hub. Level 1C imagery, which contains TOA reflectance, and Level 2A imagery, which has been atmospherically corrected and contains surface reflectance, with less than 10% cloud cover over the scene, were selected for the investigation. However, tidal corrections were applied to ensure that the in-situ data, which is referenced to the Chart Datum, corresponds to the water level observed at the time of satellite image acquisition. Tide information from ports in each study area was obtained from the library of tidal stations managed by the CHS, which stores real-time tide elevations, relative to the Chart Datum, measured every 4 to 40 minutes. To synchronize the tidal observations with the satellite acquisition time, the first tide observation measured a few minutes after image acquisition was used to correct all in-situ bathymetric data.

Table 4.3: Sentinel 2-A multispectral bands

	Band	Central Wavelength [nm]	Spatial Resolution [m]
1	Coastal blue	442.7	60
2	Blue	492.7	10
3	Green	559.8	10
4	Red	664.6	10
5	Vegetation red-edge	704.1	20
6	Vegetation red-edge	740.5	20
7	Vegetation red-edge	782.8	20
8	NIR	832.8	10
8a	Narrow NIR	864.7	20
9	Water Vapour	945.1	60
10	SWIR – Cirrus	1373.5	60
11	SWIR	1613.7	20
12	SWIR	2202.4	20

Source: European Space Agency (2014b). *Sentinel Online*. [Online]

4.3 Data preprocessing

4.3.1 Satellite Image preprocessing

The Sentinel Applications Platform (SNAP) software was used to process and atmospherically correct the Level 1C imagery with TOA radiance to obtain R_{rs} values for the Random Forest classifier. Level 2A, used for the DCNNs, was not pre-processed as it is already atmospherically corrected and contains dimensionless surface reflectance values.

The Level 1C satellite imagery was first resampled to ensure that all multispectral bands were of the same 10m resolution, which matches the spatial resolution of the CHS in-situ data. Thereafter, the image was atmospherically corrected to obtain the R_{rs} and optical properties of water using the Case 2 Regional Coastal Colour (C2RCC) algorithm. C2RCC uses neural networks trained on radiative transfer simulations of water-leaving radiance and TOA radiance. Brockmann et al. (2016) states that the algorithm works by inverting the input TOA radiance to derive the water-leaving radiance, remote-sensing reflectance in the visible and NIR wavelength regions, and inherent optical properties (Brockmann, et al., 2016). The algorithm also masks the land and clouds from the scene,

leaving the water to be analysed and studied.

The C2RCC algorithm also performs a sun-glint correction. The reflectance in the NIR band was used to determine if there were any effects of sun-glint, as any radiance over water in this band is due to the specular reflectance of the sun's light in the direction of the satellite sensor (Lyzenga, 1985). The sun-glint corrected band is obtained using Equation (4.1) (Lyzenga, 1985):

$$R'_{rs,i} = R_{rs,i} - a_i(R_{rs,NIR} - R_{rs,NIR,min}) \quad (4.1)$$

where $R'_{rs,i}$ is the sun-glint corrected band i , a_i is the slope derived from the regression line and $R_{rs,NIR,min}$ is the minimum radiance in the NIR band. A subset of the image, which only contained the study area, was created to ease management and reduce processing times.

4.3.2 Bathymetric data preprocessing

Using ArcGIS Pro, the tide-corrected bathymetric data was then transformed into two different data types for the RF model and the DCNNs. Firstly, the ASCII++ (XYZ) data was converted into point shapefiles with water depth set as the z -dimension. The point bathymetric data was grouped into 12 one-metre classes using ArcGIS Pro's Round Up tool to prepare the data for classification. This rounded all data points to the next integer and meant that depths between 0m and 1m were labelled 1, and depths greater than 10m were labelled 11. Table 4.4 shows how the depth was assigned to categorical classes. Thereafter, the point shapefiles were transformed into a tabular data format for the RF model to ensure that, through pixel value extraction, many spectral bands could be used as predictors. For the DCNNs, the data was transformed into grayscale raster datasets.

4.4 Image data annotation

Before image data annotations or labels were created, satellite imagery and in-situ bathymetric data were co-registered to ensure that both datasets were aligned spatially. This allowed the models to

Table 4.4: Depth classes

Depth range [m]	Class
< 0	0 (Land)
0 - 1	1
1 - 2	2
2 - 3	3
3 - 4	4
4 - 5	5
5 - 6	6
6 - 7	7
7 - 8	8
8 - 9	9
9 - 10	10
10 <	11 (Deep)

accurately map reflectance data to the corresponding depth classes.

Once the bathymetric data was transformed into a point shapefile for the RF model, the vector layer was reprojected to match the WGS 84 UTM coordinate reference system of the Sentinel-2A imagery for each study area and test site. For the DCNNs, the point shapefile was transformed further into a topographic surface raster by interpolating a continuous surface from the discrete data points using the Nearest Neighbour technique. The raster images were then resampled to a cell size of 10m x 10m to match the resolution of the satellite images.

4.4.1 Image data labelling for RF model

To label high-dimensional image data for the Random Forest, the European Space Agency's Sentinel Applications Platform (SNAP) was used to extract the multispectral pixel values at each discrete data point. The bathymetric point shapefile was converted into a tabular dataframe to create a list of coordinates for the pixel value extraction tool. At each coordinate location, the reflectance value from each multispectral band is extracted and populated in a comma-separated-value file to create a labelled tabular dataframe. Thereafter, data from Lake Ontario's coastal region of Athol was left out for freshwater testing. In contrast, Graham Island in the Hecate Strait and Rimouski

Port in the Saint Lawrence Seaway were left out for saltwater testing.

4.4.2 Image data labelling for DCNNs

The ArcGIS Pro software was used to create the labelled datasets because it can transform labelled thematic raster images and multispectral images into a format compatible with TensorFlow and PyTorch. To create a labelled thematic raster, the surface raster, with water depth set as the z-dimension, was converted to integer values representing the class label names. The labels were then converted from integers to categorical (text) labels. The topographic surface was finally restored to a vector polygon feature class, which could then be exported as a labelled mask. Figure (4.4) shows a labelled mask of Point Clark, Lake Huron.

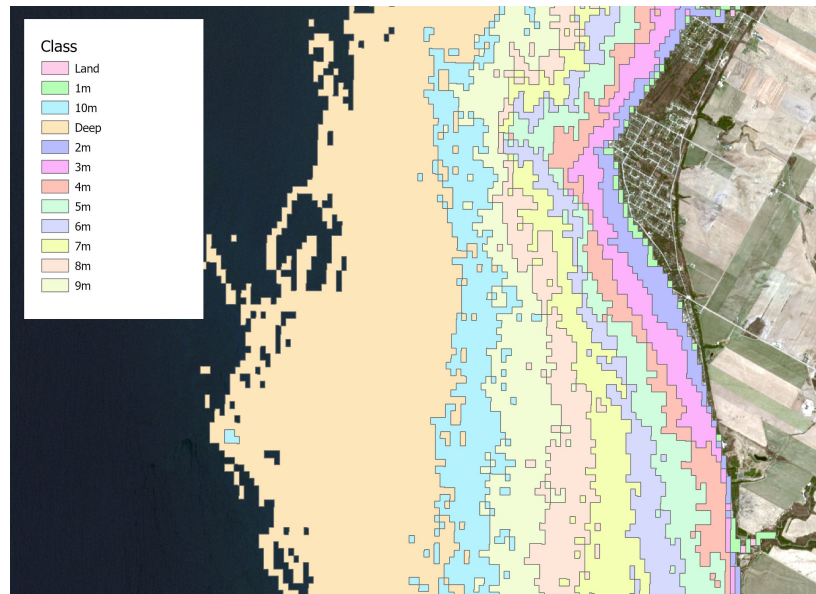


Figure 4.4: Labelled mask of Point Clark, Lake Huron for training

Due to the nature of the investigation, the masks were exported in a specific label format that was compatible with satellite imagery. The labelled mask was exported as classified tiles, with one classified image chip per input image chip. The metadata contains information about the class names, class values, and output statistics and is ideal for pixel-wise classification. To ensure that the labels were accurate, the class polygons were not simplified but remained pixelated to preserve detail

and the complex geometries of the depth classes, and limit the alteration of spatial relationships between depth classes. Rounding the point bathymetric data to the next integer also created blocks of depth classes, which explains the pixelated nature of the ground truth label masks. Semantic segmentation models require large volumes of labelled data to learn the patterns in the data. Therefore, to increase the size of our dataset, the image from each study area was subsampled to create 256x256 image chips and label chips with a stride of 128, resulting in nearly 9 000 images collected for the training, validation, and testing of the transferability of DL models. Table 4.5 shows the distribution of class features and the number of images in which each class is present for all the collected data.

Table 4.5: Distribution of class features

Class	Number of images	Number of features	Percentage of features [%]
Land	5 247	26 484	6.54
1m	6 328	32 255	7.97
2m	6 859	36 848	9.10
3m	6 987	36 411	8.99
4m	6 935	36 781	9.08
5m	6 835	35 514	8.77
6m	6 826	34 754	8.58
7m	6 819	33 242	8.21
8m	6 880	34 418	8.50
9m	6 977	35 822	8.85
10m	7 003	37 027	9.14
Deep (beyond 11m)	8 874	25 401	6.27

Nearly 10% of the images were used for validation and data from Lake Ontario’s coastal region of Athol was left out for freshwater testing, while Graham Island in the Hecate Strait and Rimouski Port in the Saint Lawrence Seaway were left out for saltwater testing.

4.5 Model Training

Our investigation used a supervised learning approach to learn a mapping function between multi-spectral reflectance and water depth. Supervised learning is well-suited for this task, as it enables

machine learning models to learn from labelled datasets and generalize patterns to new, unseen data. The Random Forest model is trained on the tabular dataset, while the deep convolutional neural networks (DCNNs) are trained using labelled image masks. The following sections describe the training process.

4.5.1 Random Forest Training

A Random Forest (RF) classifier was imported and initialized from Python’s Scikit-Learn library. The data used to train the RF consisted of 8 predictor variables: products from the C2RCC atmospheric correction and their one-metre target class variables.

Table 4.6: Features included in the Random Forest as predictors

Predictors	Description
Coastal Blue, Blue, Green, and Red	Atmospherically corrected bands containing R_{rs} values
Kd_{min}	Mean incoming radiance attenuation coefficient of three bands
Kd_{489}	Incoming radiance attenuation coefficient received at the surface at 489nm
Iop_{bpart}	Scattering coefficient of marine particles at 443nm
Iop_{apig}	Absorption coefficient of phytoplankton pigments at 443nm

This data was stored in a CSV format and loaded into Jupyter Notebook as a Pandas data frame. The eight predictor variables included in the investigation, which were described by Kyryliuk and Kratzer (2019), are given in Table 4.6. The salt and fresh water RF models were constructed by defining the $Ntree$ and Kn parameters, and setting $Mtry$ to the square root of the number of features. The models were trained on the saltwater and freshwater tabular datasets, respectively. The number of points in each class available for training or the class distributions are shown in Table 4.7. The stratified K-Fold cross-validation was incorporated into the training to overcome the effect of imbalanced classes. This involved further splitting the training dataset into 10 folds. 9 folds are selected randomly for training, and the 10th fold is used for validation. This was iterated 10 times to

Table 4.7: Class distribution of points in fresh and salt water tabular datasets

Class	Saltwater	Freshwater
Land	181 886	14 021
1m	409 347	48 489
2m	369 288	71 765
3m	326 285	166 722
4m	283 039	247 926
5m	278 949	255 541
6m	330 565	255 661
7m	382 462	237 235
8m	431 184	214 785
9m	410 271	214 495
10m	449 722	292 535
Deep	1 137 580	1 062 548

ensure the whole dataset was used for training and validation. To find the optimal hyperparameters for the fresh and saltwater models, the Randomized Search was implemented to search through a wide range of model parameters that will generate the most accurate results. Thereafter, the Grid Search was used to refine the results from the Randomized Search to find the optimal hyperparameters. The best set of hyperparameters is represented in Table 4.8.

Table 4.8: Optimal Random Forest hyperparameters for salt and freshwater

Hyperparameter	Saltwater	Freshwater
<i>Ntree</i>	433	483
<i>Mtry</i>	SQRT	SQRT
<i>Kn</i>	13	13

Once the best hyperparameters were identified, the saltwater and freshwater models were trained again. The performance of the models on the validation datasets was evaluated by comparing their predictions against the ground truth values and was visualized using a confusion matrix, as shown in Figure (4.5). The true-positive (TP), false-positive (FP), and false-negative (FN) were used to calculate the precision, recall, and F1 scores, which quantitatively measure the performance of classification models and are defined in Equation (4.2):

		Predicted Class	
		POSITIVE	NEGATIVE
Actual Class	POSITIVE	TP True-Positive (Number of positive samples correctly classified as positive)	FP False-Positive (Number of negative samples incorrectly classified as positive)
	NEGATIVE	TN True-Negative (Number of negative samples correctly classified as negative)	FN False-Negative (Number of positive samples incorrectly classified as negative)

Figure 4.5: Confusion Matrix

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (4.2)$$

where precision measures the proportion of positive predictions that are actually correct, it ranges from 0 to 1, where 0 means none of the positive predictions were correct, and 1 indicates all positive predictions were correct. Recall, also known as sensitivity, measures the proportion of actual positives correctly classified by the model. It ranges from 0 to 1, where 0 indicates that none of the actual positives were classified, and 1 indicates that all actual positives were classified. The F1 score is the harmonic mean of the precision and recall and is suited for managing imbalanced classes. It ranges from 0 to 1, where 0 indicates the worst performance (no correct positive predictions), and 1 indicates the best performance (perfect precision and recall). The metrics for each depth class were compiled into a table.

4.5.2 DCNNs Training

A Python project was developed for the investigation using several existing GitHub repositories. The PyTorch library was used to build and train the DL models, while TensorBoard from the Tensorflow library was used to visualize the models' performance metrics.

A data loader was created to efficiently load data into the DL model by converting the label and image chips into tensors. In the data loader, the ‘mode’ parameter was added to specify the model type (freshwater or saltwater) being used. Data augmentation techniques were applied to our dataset to ensure the model had sufficient diverse data to learn from and was robust in its predictions. The transformations applied to the image data included random cropping to a size of 224x224, rotation, horizontal and vertical flips, blur, and scale.

The training and performance of the DL models were monitored using the multiclass focal loss function, where a value close to zero indicates the model is confident in its predictions. In contrast, a higher value indicates poor confidence. The stochastic gradient descent (SGD) was used as the network’s optimizer to minimize the calculated loss. The optimizer updated the network weights after each batch iteration by propagating the loss backwards through the network. The SGD optimizer included several tuning parameters that control the rate at which the network’s weights are updated. These include the momentum, learning rate, and weight decay. However, to identify the optimal hyperparameters for each DL model, a hyperparameter search was conducted using the Optuna optimization framework. The tuned parameters that were tuned were the learning rate, weight decay, training batch size, and validation batch size. Table 4.9 summarises the optimal hyperparameters for each network.

Table 4.9: Optimal hyperparameters for DCNNs in fresh and salt water

Water type	Hyperparameter	U-Net	SegNet	DeepLabv3+
Freshwater	Learning rate	0.1	0.02	0.003
	Weight decay	$1e^{-10}$	$1e^{-10}$	$1e^{-12}$
	Train batch size	25	23	24
	Validation batch size	21	23	24
Saltwater	Learning rate	0.01	0.01	0.01
	Weight decay	$1e^{-9}$	$1e^{-9}$	$1e^{-8}$
	Train batch size	8	10	14
	Validation batch size	26	17	26

Once the optimal hyperparameters were identified, the models were trained on the freshwater and saltwater datasets, separately for 800 epochs. Early stopping was implemented during training with a patience of 80 epochs to prevent overfitting. By monitoring the model’s performance on the validation dataset, training was terminated if the focal loss function did not improve in 80 epochs. The model weights from the epoch with the lowest validation loss were saved and would later be used for testing model transferability. The precision, recall, and F1 scores were used as evaluation metrics for the predictive performance.

4.5.2.1 Loss Function: Focal Loss

The cross-entropy (CE) loss function is commonly used in multi-class classification tasks to quantify the error between the predicted class probabilities and the actual target labels. The function penalizes the model for confidently making incorrect predictions by increasing the loss. The loss is reduced and approaches 0 when the model makes a confident and correct prediction (Ross & Dollár, 2017). The function is defined in Equation (4.3) as:

$$CE(p_t) = -\alpha_t \log(p_t) \quad (4.3)$$

where $\alpha_t \in [0, 1]$ is the true target distribution and the weight factor that manages class imbalance, and p_t is the prediction probability distribution of class t . However, the cross-entropy cannot manage significant class imbalances as the weight factor only considers the correct and incorrect predictions and cannot distinguish between the easy and hard data instances to classify. As Table 4.5 shows, our dataset is slightly imbalanced, so the focal loss (FL) function was used in our investigation to measure the DCNN’s performance. Ross and Dollár (2017) addressed the limitations of the CE by adding a modulating factor that deteriorates to 0 as the confidence in the predicted classes increases. Equation (4.4) defines the focal loss as:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (4.4)$$

where $\gamma \geq 0$ is a tuneable parameter, and $(1 - p_t)$ is the modulating factor that assigns smaller weights to confident predictions and larger weights to misclassified data instances. Ross and Dollár (2017) explain that as p_t approaches 1, the modulating factor diminishes to 0 to decrease the weight for easily classified instances. γ controls the rate at which the weight adjusts. If $\gamma = 0$, the focal loss behaves like the cross-entropy loss, and once increased, the modulating effect increases. $\gamma = 2$ was used in the investigation to ensure the model focuses on learning the hard-to-classify data instances.

4.6 Model Transferability

Transferability is the ability to apply model weights from one location to a new, unseen area. Models use the learned relationships and patterns from the training data to predict water depth from unseen satellite imagery of different geographic locations. To ensure transferability, domain generalization was implemented, where models were trained on bathymetric data from widely geographically distributed areas. Multitemporal satellite imagery and the corresponding tide-corrected bathymetric data were also incorporated to create a diverse set of reflectance data for training. This guaranteed that models learned general and transferable patterns and were less likely to overfit. Once training was complete, the model weights from the best training epoch were imported for testing. Each trained saltwater model was used to predict water depth from unseen satellite image test data from Graham Island and Rimouski Port, and each trained freshwater model was tested on the Lake Ontario Athol site. The result produced is a classified image, where each pixel in the input image is associated with a depth class label. We can determine if ML models trained in specific geographic locations can generalize to a new geographic area by testing model transferability. The accuracy of the model prediction was evaluated by comparing it with the ground truth label and using the precision, recall, and F1 scores.

4.7 Comparison with deterministic models

To demonstrate the credibility of machine learning methods for SDB, the results were compared against well-established empirical and physics-based models. Stumpf's empirical band ratio algorithm was used due to its consideration of variable bottom types. The physics-based radiative transfer model was implemented using the Water Colour Simulator (WASI) programme developed by Gege (2005), which uses a set of analytical equations to calculate the water column and bottom properties. The approach used by Albert (2004) was followed, whereby a set of R_{rs} spectra was simulated by defining the IOPs, bottom character, solar zenith angle, and the viewing angle of the satellite sensor of each test site. The bottom character of each test site was determined and inserted into WASI. The solar zenith angle and viewing angle were set to 30° . For the IOPs, the concentration of phytoplankton and suspended matter, and the absorption of gelbstoff were estimated from the results of the C2RCC atmospheric correction. Water depth was then derived by finding and inverting the simulated spectral signature that best matches the observed spectral signature. During the inversion, all estimated parameters remained fixed while depth was set as the fit parameter and was determined iteratively using nonlinear curve fitting.

The regression predictions were grouped into one-metre classes and converted to categorical feature class shapefiles in ArcGIS Pro to compare the results of empirical and physics-based models to the ML models. The precision, recall, and F1 scores were calculated using the Tabulate Intersection tool, which returns the count of the class-wise predicted features intersecting with the ground truth features, specifying the correct and incorrect predicted features. These results were then used to compute a classification accuracy report for each class.

Chapter 5

Experiments and Results

This chapter presents the results from the machine learning model, Random Forest, and the deep learning models, U-Net, SegNet, and DeepLabv3+. These results include the performance of each model in fresh and saltwater during training and validation, and finally, testing of model transferability. The experiments and results of the Random Forest are presented in Section 5.1. Section 5.2 shows the results from the DCNNs with their associated metric scores and provides a performance comparison between the three networks. The performance of all machine learning methods is compared to the baseline empirical and physics-based models in Section 5.3. Finally, the advantages, limitations and problems of each modelling approach are summarized in section 5.4.

5.1 Random Forest classification results

The performance of the RF model in saltwater and freshwater environments during training and validation is presented, followed by its performance during model transferability testing.

5.1.1 Saltwater Random Forest model

5.1.1.1 Model Training

To prevent overfitting, the stratified K-fold cross-validation method was used to ensure the model was trained and tested on the entire saltwater training dataset. The classification report in Table 5.1 summarizes the class-wise and mean accuracy scores.

Table 5.1: Saltwater RF classification report

Class	Precision	Recall	F1
Land	0.77	0.73	0.75
1m	0.81	0.62	0.70
2m	0.47	0.44	0.46
3m	0.35	0.22	0.27
4m	0.35	0.11	0.16
5m	0.35	0.11	0.16
6m	0.33	0.10	0.16
7m	0.31	0.21	0.25
8m	0.21	0.12	0.15
9m	0.24	0.05	0.08
10m	0.21	0.04	0.06
Deep	0.36	0.96	0.53
Mean	0.40	0.31	0.31

During the validation process, the saltwater model achieved a mean F1 score of 0.31, which indicates that less than half of the image pixels were accurately predicted. This low accuracy reflects the inability of the model to perform well on an independent dataset that it was not trained on during that specific fold. The class-wise scores reveal that the model's prediction accuracy decreases with depth. The model can accurately predict the Land class to obtain an F1 score of 0.75, signifying that the RF can identify the coastline to an accurate degree. The 1m class acquired the second-highest F1 score of 0.70 and the highest precision score of 0.81. Within the 1m class, 81% of the pixels were 1m, which reveals that the saltwater model can identify and learn the patterns of the R_{rs} from the depth class. However, with a recall score of 0.62, many 1m class pixels are misclassified. The 2m class achieved an F1 score of 0.46 and was accurately predicted. The confusion matrix in Fig-

ure (5.1) reveals that the diagonal elements are only high for the Land, 1m, 2m, and Deep classes. Beyond 2m, the accuracy decreases significantly to almost 0 in the 10m class. This is likely due to the deeper water columns that absorb and scatter more light, resulting in less light reaching the bottom and reflecting in the direction of the sensor. The low R_{rs} makes it difficult for the RF model to distinguish unique patterns for classification. Figure (5.1) also shows that all depth classes were consistently assigned to the Deep class, which explains the low precision score of 0.36 but a high recall score of 0.96.

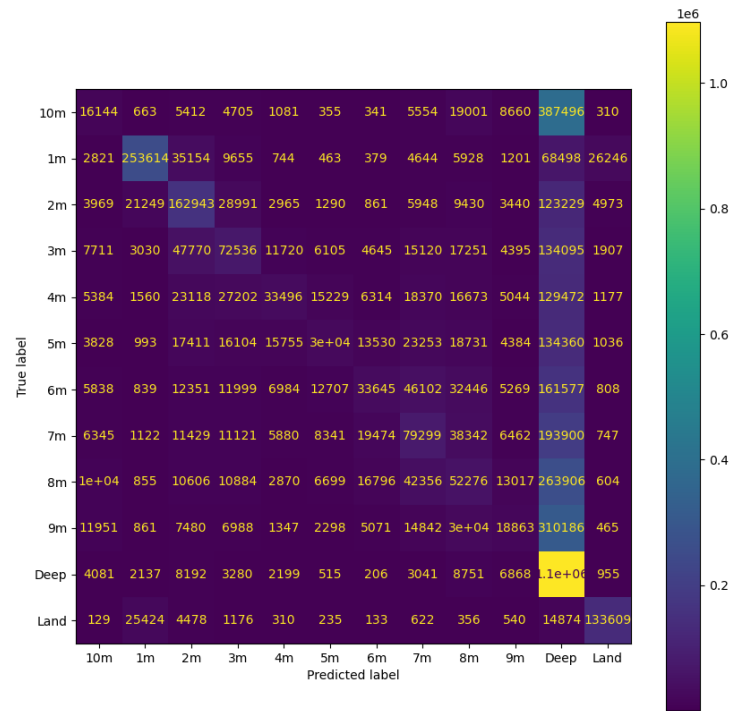


Figure 5.1: Saltwater RF model confusion matrix

The significant features during the cross-validation process are illustrated in Figure (5.3). The mean incoming radiance attenuation coefficient, Kd_{min} , contributed most to the prediction process as it has an exponential relationship with water depth. As it determines the amount of light available at each depth level, this band contains valuable information for the model to learn and use to classify image pixels. The next most significant contributor was the green band, which has the greatest penetrative power, is the least attenuated in the water column, and can reach the bottom to be

reflected. The optical property, IOP_{apig} , was the third-greatest contributor. In saline waters loaded with sediments, the estimation of optical properties based on in-situ measurements like average temperature and salinity can assist ML models in predicting water depth from satellite image data. IOPs can alleviate the negative impact of low R_{rs} values from deeper waters in SDB. The coastal blue and blue bands contributed least to the model as water molecules promptly absorb their short wavelengths. Hence, they are only valuable in extremely shallow waters but are insignificant in deep waters.

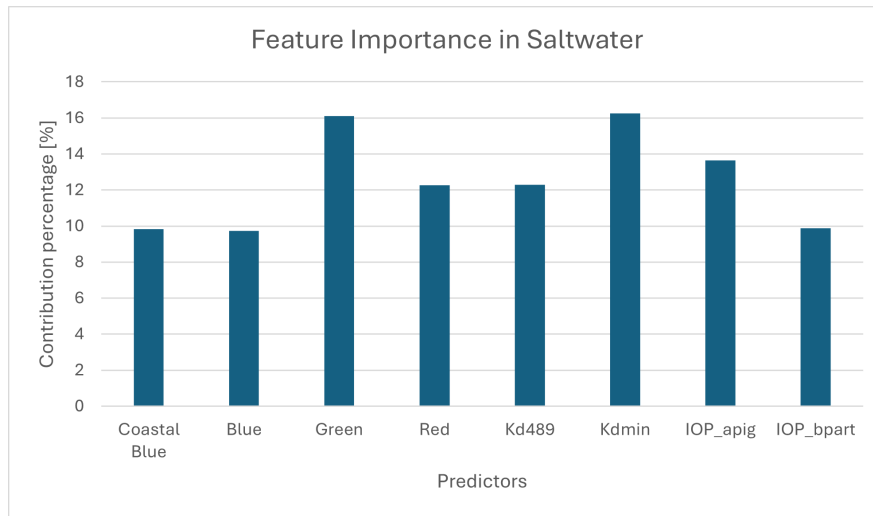


Figure 5.2: Saltwater RF model feature importance

5.1.1.2 RF Model Transferability in Graham Island

Thereafter, the model was tested on Graham Island and the plot of these predictions is shown in Figure (5.3). The model achieved a low mean F1 score of 0.07. The saltwater model missed several classes, mainly those in the 3m to 7m depth range and explains the accuracy scores of 0 for these classes. This is a clear indication that the model severely overestimates depths in saltwater. Only pixels in the 3m, 5m, and Deep classes were identified and correctly classified; however, a small percentage of these. The Land class obtained a recall score of 0.88 but a low precision score of 0.05. This indicates that many shallow waters were classified as land. The steep slope of the shores of Graham Island presents a challenge for ML models in making accurate predictions. This effect

Random Forest depth prediction, Graham Island

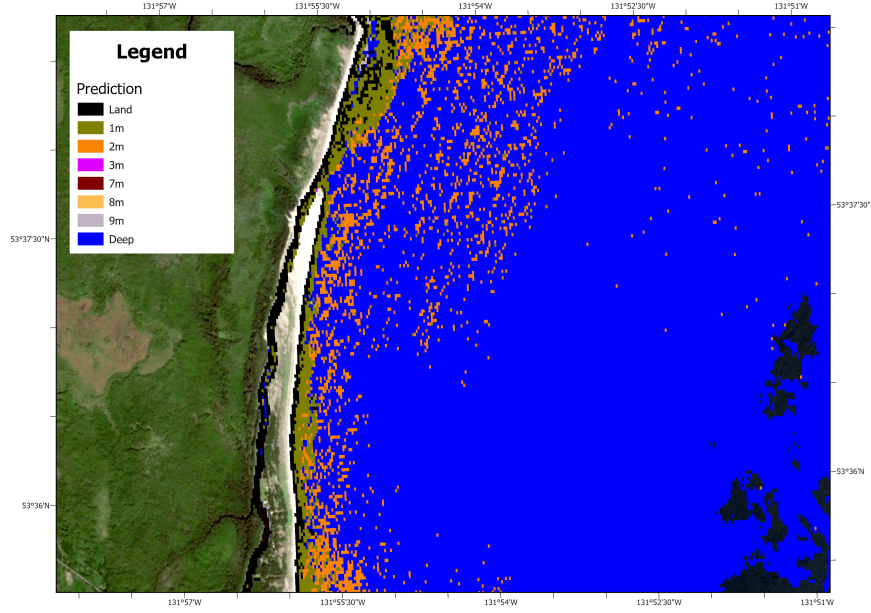


Figure 5.3: Random Forest prediction of Graham Island

is exacerbated in the RF, which does not learn spatial context. The misclassification of the entire test site also reveals that models must be retrained on the test areas.

Table 5.2: Classification report of Graham Island

Class	Precision	Recall	F1
Land	0.05	0.88	0.10
1m	0.01	0.07	0.01
2m	0.00	0.02	0.00
3m	0.25	0.00	0.00
4m	0.00	0.00	0.00
5m	0.50	0.00	0.00
6m	0.00	0.00	0.00
7m	0.00	0.00	0.00
8m	0.00	0.00	0.00
9m	0.01	0.00	0.00
10m	0.00	0.00	0.00
Deep	0.66	0.89	0.76
Mean	0.12	0.16	0.07

5.1.1.3 RF Model Transferability in Rimouski

The saltwater model was also tested on the port of Rimouski, a fresh and saltwater mixture. The classification report in Table 5.3 shows that the model performed better in Rimouski.

Random Forest depth prediction, Rimouski Port

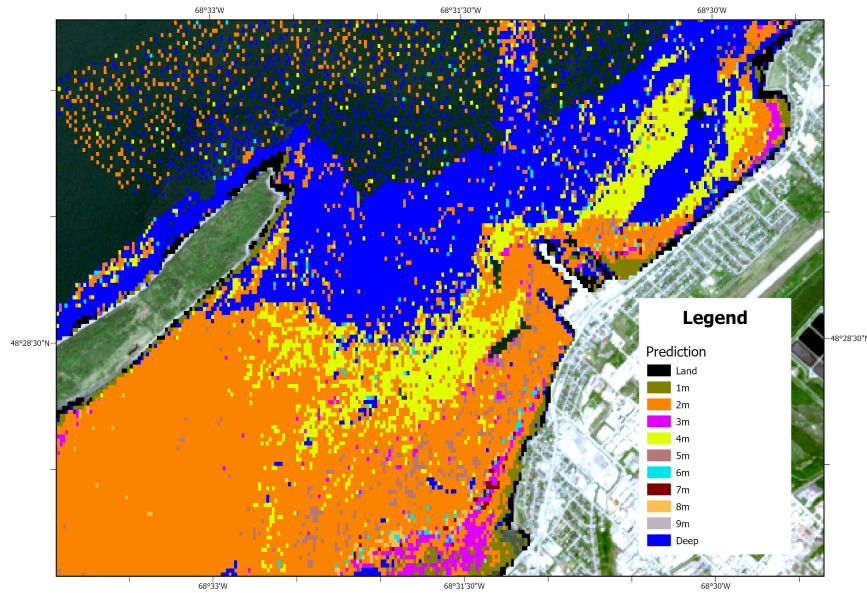


Figure 5.4: Random Forest prediction of the Rimouski Port

Table 5.3: Classification report of the Rimouski Port

Class	Precision	Recall	F1
Land	0.45	0.60	0.51
1m	0.57	0.25	0.35
2m	0.53	0.71	0.60
3m	0.20	0.03	0.05
4m	0.28	0.33	0.30
5m	0.09	0.05	0.07
6m	0.13	0.03	0.05
7m	0.00	0.00	0.00
8m	0.00	0.00	0.00
9m	0.00	0.00	0.00
10m	0.00	0.00	0.00
Deep	0.16	0.73	0.26
Mean	0.20	0.23	0.18

The model moderately predicted up to 4m, with the 2m class obtaining an F1 score of 0.60. As shown in Figure (5.4), the Rimouski test site is shallow primarily and higher R_{rs} values allow the model to distinguish the spectral signatures for each class. In the case of Graham Island, the dense, sediment-laden saltwater does not return significant reflectance values which can be differentiated. However, the model's performance decreased in deeper waters, and the 10m class was never identified. The model underestimates depth in deeper waters as Figure (5.4) shows misclassification between the Deep and 2m classes. This reflects the turbidity of the Saint Lawrence River in this region, which causes intense backscattering of light in the deeper waters.

5.1.2 Freshwater Random Forest model

5.1.2.1 Model Training

The freshwater model performed significantly better than the saltwater model and achieved a mean F1 score of 0.64. Over half of the image pixels were accurately predicted. This high accuracy reflects that SDB is more successful in freshwater, as it contains less sediment than saltwater. Because of its low internal resistance to flow, light can penetrate.

Table 5.4: Freshwater RF classification report

Class	Precision	Recall	F1
Land	0.66	0.44	0.53
1m	0.69	0.72	0.70
2m	0.75	0.65	0.69
3m	0.72	0.59	0.65
4m	0.58	0.61	0.60
5m	0.49	0.48	0.49
6m	0.47	0.48	0.47
7m	0.45	0.40	0.43
8m	0.47	0.31	0.37
9m	0.45	0.30	0.37
10m	0.51	0.38	0.44
Deep	0.76	0.95	0.84
Mean	0.58	0.53	0.55

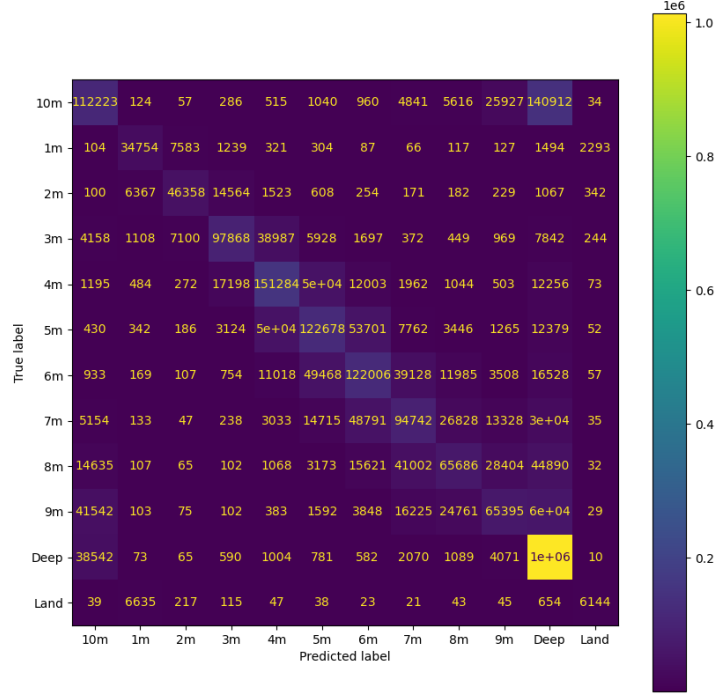


Figure 5.5: Freshwater RF model confusion matrix

The RF freshwater model maintained relatively high accuracy scores across all depth classes, implying that the transparency of freshwater allows more light to penetrate further in deep water columns to reach the bottom. Higher R_{FS} values also will enable the freshwater model to differentiate the spectral signatures for each depth class and to make accurate predictions, as Figure (5.5) shows that the diagonal elements are high. The Deep class had the highest F1 score of 0.84 with a recall of 0.96. This indicates that the model can identify and assign deep-water pixels to the Deep class. The precision score of 0.76 reveals that some 7m to 10m class pixels are mistakenly assigned to this class, as shown by the confusion matrix in Figure (5.5). The 1m class was next with the highest F1 score of 0.70. This means the model can accurately predict the dynamic shallow class, which is expected as light passes through a short water column. The Land class obtained a lower F1 score of 0.53, much lower than the score from the saltwater model. This may impact the model's ability to identify the shoreline in unseen test images precisely. From 0m to 4m, the model acquired F1 scores above 0.6, which proves that the model performs well in shallow waters. Beyond 4m, the accuracy falls to roughly 0.4, with the 7m and 8m classes obtaining even lower scores of 0.37,

showing that deeper water returns low R_{rs} values that are difficult to classify.

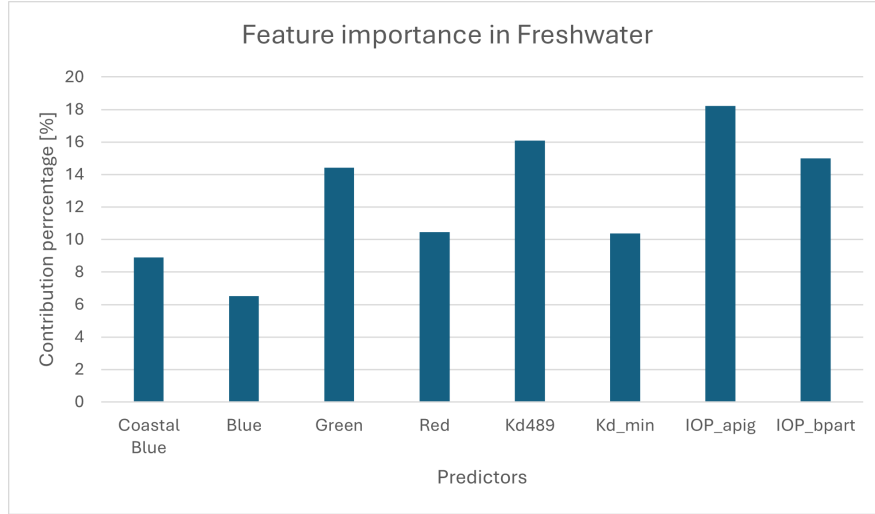


Figure 5.6: Freshwater RF model feature importance

Figure (5.6) displays the significant features during the cross-validation process, and the contributions differ from those of the saltwater model. The absorption coefficient of phytoplankton pigment, IOP_{apig} , contributed most to the prediction process as the concentration of phytoplankton has a stable behaviour in water, where there is a greater concentration in shallow waters and a lower concentration in deeper waters further away from the shoreline. This was followed by Kd_{489} and IOP_{bpart} . The green band was the next big contributor after the IOPs, which differs from the saltwater model, which had the green band as the second-greatest contributor. Like the saltwater model, the coastal blue and blue bands contributed least to the model's prediction as the pigment in phytoplankton, chlorophyll-a, strongly absorbs light in the blue, red, and NIR regions of the visible spectrum.

5.1.2.2 RF Model Transferability in Athol

Thereafter, the model was tested on Athol, and the plot of these predictions is shown in Figure (5.7). The plot shows that the RF can learn the general pattern in the contour of depths up to 5m. The 1m class acquired the highest F1 score of 0.56, followed by the 2m class with 0.49. The transparency of the water aids the freshwater model in making good predictions in waters closest to the shoreline.

Random Forest depth prediction, Athol

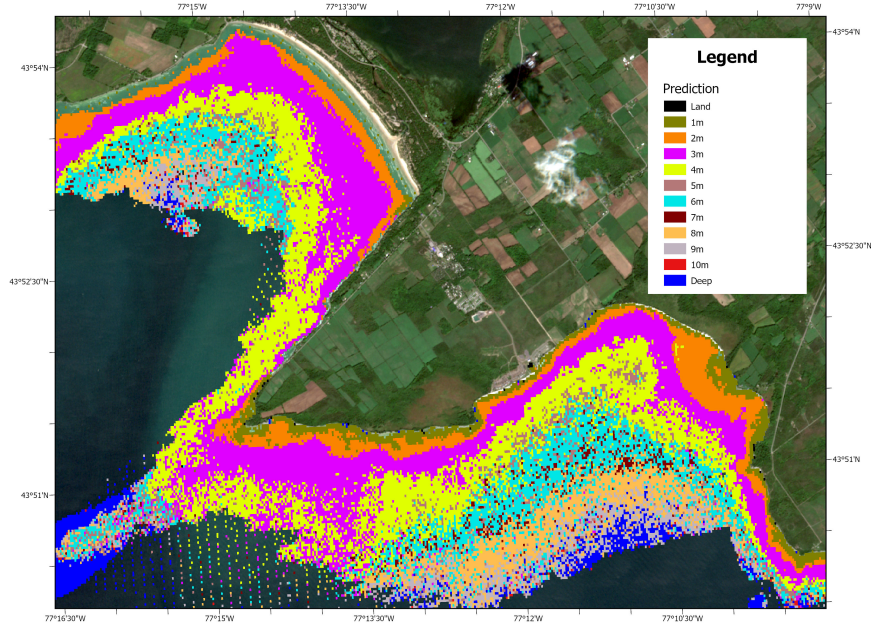


Figure 5.7: Random Forest prediction of Athol

Table 5.5: Classification report of Athol

Class	Precision	Recall	F1
Land	0.20	0.14	0.17
1m	0.41	0.88	0.56
2m	0.40	0.64	0.49
3m	0.20	0.57	0.30
4m	0.10	0.20	0.13
5m	0.08	0.05	0.06
6m	0.07	0.08	0.07
7m	0.12	0.04	0.06
8m	0.04	0.03	0.04
9m	0.01	0.01	0.01
10m	0.02	0.00	0.00
Deep	0.67	0.19	0.29
Mean	0.12	0.16	0.18

However, Table 5.5 reveals that the transparency decreases beyond 3m, complicating the prediction process. Though the decrease in transparency should not severely impact the model's performance since IOP_{apig} was identified as the most significant predictor in freshwater, clear water is still an important factor in water depth prediction. The low accuracy scores in the deep waters prove that

the Random Forest cannot successfully classify low R_{rs} values.

5.2 DCNNs semantic segmentation results

5.2.1 U-Net Results

The performance of U-Net in saltwater and freshwater environments during training and validation is presented, followed by its performance during model transferability.

5.2.1.1 Saltwater Model Training

Figure (5.8) shows the training and validation loss curves that provide insight into the model's learning behaviour during training. Wall time is the elapsed time taken to train the model. This includes the computation, reading and writing time, and time spent waiting for resources.

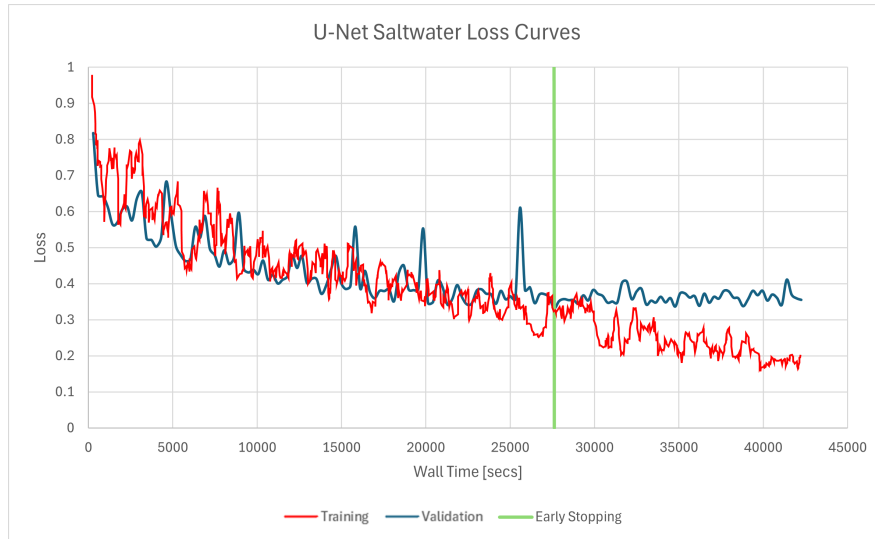


Figure 5.8: U-Net saltwater loss curves

The early stopping point indicates the time validation loss reached its minimum and training stopped.

Both loss curves should converge as they approach 0. At the start, both loss curves decreased rapidly as the model learned the patterns in the reflectance data. After some time, the validation curve began to plateau while the training curve continued to decline. Early stopping allowed the model weights from the epoch with the lowest validation loss to be saved before the model began

overfitting, indicated by the increasing validation curve. The saltwater model acquired a training loss of 0.255 and a validation loss of 0.332. By implementing the focal loss function into the training, the network could generalize better and perform well on the validation dataset. After each training and validation epoch, the recall, precision and F1 scores were calculated. The metrics from the best training and validation epoch are summarized in Table 5.6 and the class-wise scores in Table 5.7.

Table 5.6: Best training and validation epoch scores of U-Net saltwater model

	Training	Validation
Loss	0.255	0.332
Precision	0.554	0.456
Recall	0.538	0.436
F1	0.537	0.433
Pixel Accuracy	0.640	0.513

The saltwater model achieved a mean training F1 score of 0.537 and a mean validation F1 score of 0.433, indicating that the network did not memorize the training data and could generalize to new, unseen data. On the other hand, the class-wise scores show that specific depth ranges were easy to predict while others were difficult. Land had the highest precision score of 0.767. This implies that the model can identify the coastline but occasionally misclassifies some water surfaces as land and vice versa. This is expected in extremely dynamic waters with wave activity and high tides. The model could accurately classify the depth range between 0m and 3m, as class-wise F1 scores above 0.5 were obtained. This is expected as the shorter water column means there is less attenuation, which allows more light to reach the seafloor and reflect in the direction of the satellite sensor. The greater spectral response from this depth range enables U-Net to distinguish between one-metre classes. However, as depth increases, the validation accuracy decreases. The 4m and 5m classes were not classified as accurately as the shallower waters and obtained scores around 0.4. This accuracy dropped sharply beyond 5m. The 7m class had the lowest precision score of 0.303, which indicates that less than half of the predicted pixels are 6-7m deep. It is within this depth range that U-Net cannot successfully differentiate one-metre classes.

Table 5.7: U-Net saltwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.767	0.635	0.695
1m	0.572	0.602	0.586
2m	0.586	0.544	0.564
3m	0.469	0.538	0.501
4m	0.433	0.456	0.444
5m	0.416	0.427	0.421
6m	0.365	0.247	0.295
7m	0.303	0.232	0.263
8m	0.311	0.248	0.276
9m	0.358	0.33	0.344
10m	0.332	0.069	0.114
Deep	0.56	0.904	0.692

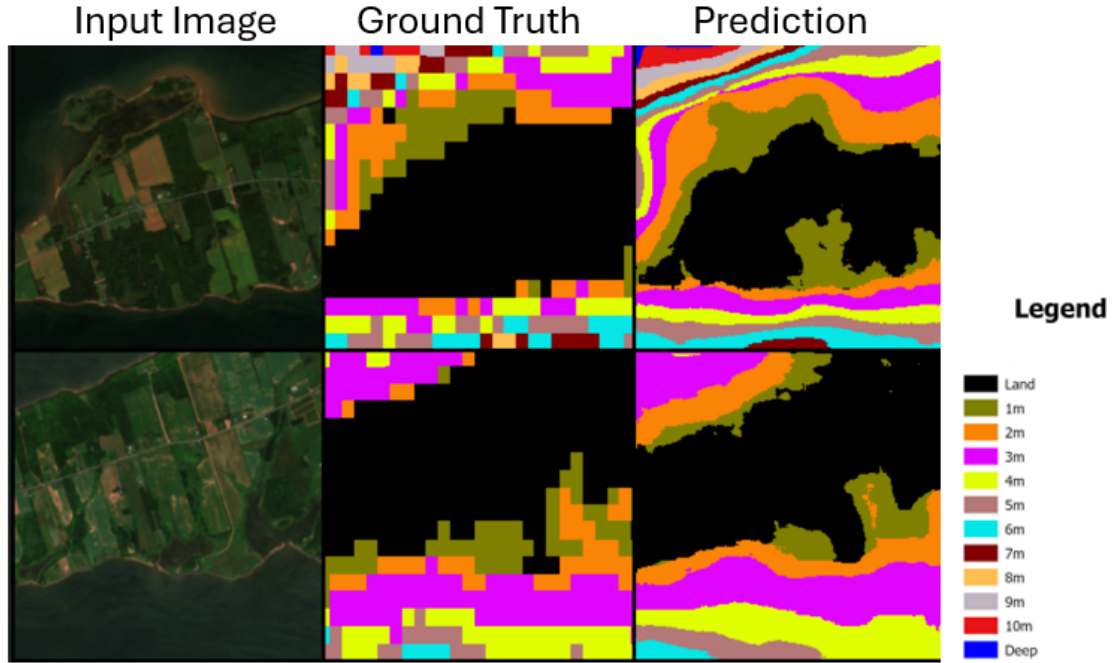


Figure 5.9: U-Net saltwater validation predictions

The 10m class had the lowest F1 and recall scores, 0.114 and 0.069, respectively, which indicates that the model finds it difficult to classify this depth range. On the other hand, the Deep class had a high F1 score of 0.692, recall of 0.904, and precision of 0.56, showing that the model is highly accurate in classifying and predicting deep waters. As the spectral response from waters deeper than 9m should not vary greatly, this score contrast implies that the 10m class is not distinguish-

able due to insufficient data or its spectral similarity to deeper classes. Figure (5.9) displays the accuracy of the model's prediction on the validation dataset. The model can identify the coastline and shallow water well and recognize the gradual water depth change. By visually inspecting the predicted images, it can be understood that the coastline is often misclassified as the 1m class, and consequently, the 1m class is misclassified as 2m. However, the model can predict the spatial relationship between each class from the input image.

5.2.1.2 Model Transferability in Graham Island

Once the validation results were inspected, the weights from the best validation epoch were saved and used during testing. The test images were fed into the network, and predictions, shown in Figure (5.10), were produced. The mean accuracy scores from the Graham Island test site are displayed in Table 5.8.

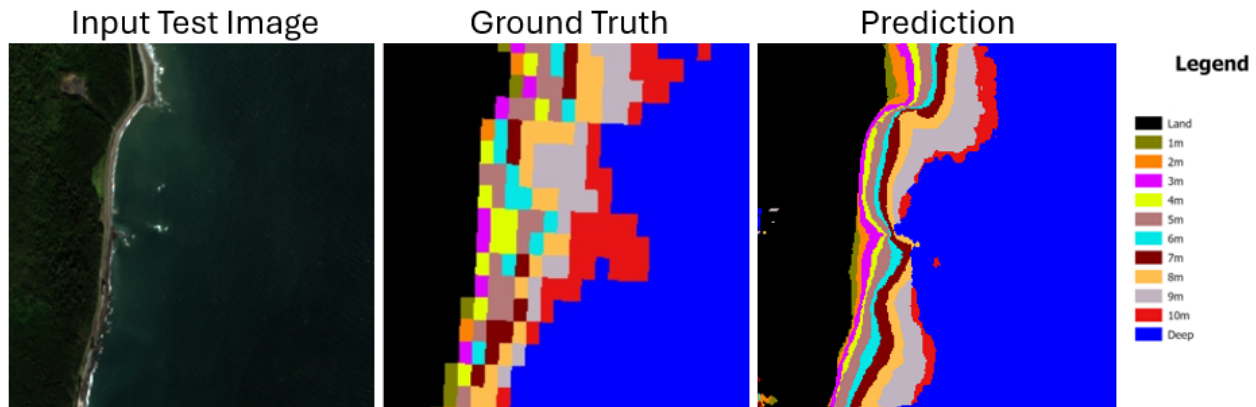


Figure 5.10: Graham Island U-Net predictions

The results in Table 5.8 convey that the U-Net saltwater model cannot accurately predict new unseen data from a different location. The model obtained an F1 score of 0.319, indicating over half of the predictions were incorrect. The class-wise scores show more information as to which classes were misclassified. The Land and Deep classes were accurately predicted, with the Land class having an F1 score of 0.938, a recall of 0.903, and a precision of 0.976. This shows that U-Net can identify the coastline as 97.6% of the pixels classified as Land were Land. The Deep class obtained an F1 score of 0.63, a recall of 0.979, but a low precision score of 0.464. The wide discrepancy in

Table 5.8: Graham Island U-Net prediction scores

Class	Precision	Recall	F1
Land	0.976	0.903	0.938
1m	0.227	0.246	0.236
2m	0.305	0.238	0.268
3m	0.228	0.32	0.266
4m	0.327	0.188	0.239
5m	0.321	0.312	0.316
6m	0.232	0.163	0.192
7m	0.386	0.123	0.186
8m	0.371	0.181	0.243
9m	0.386	0.209	0.271
10m	0.119	0.03	0.048
Deep	0.464	0.979	0.63
Mean	0.362	0.324	0.319

accuracy scores reveals that though most deep-water pixels were correctly classified, many pixels from other classes were included in the Deep class. The 10m class was most likely misclassified as deep, as revealed by the extremely low F1 score of 0.048. All depth classes between 0m and 10m had low F1 scores around 0.2, which likely implies that the model must be trained on the area to make accurate predictions.

5.2.1.3 Model Transferability in Rimouski

The saltwater model was also tested on the Port of Rimouski (Figure (5.11)), which has a mix of salt and fresh water and the accuracy scores are summarized in Table 5.9.

The model achieved slightly higher scores for Rimouski than Graham Island, with an F1 of 0.326, recall of 0.446, and precision of 0.351. The model behaved similarly to Graham Island, where the Land class was predicted accurately, while the rest were predicted poorly. However, the model performed differently in this study area as the Deep class acquired lower scores, with an F1 score of 0.042, a recall of 1.0, and a precision of 0.021. With a recall of 100%, this indicates that all the deep-water pixels were correctly predicted as deep, but other class pixels were included in the class to produce a low precision and F1 score. Just as was exhibited in the results from Graham Island,

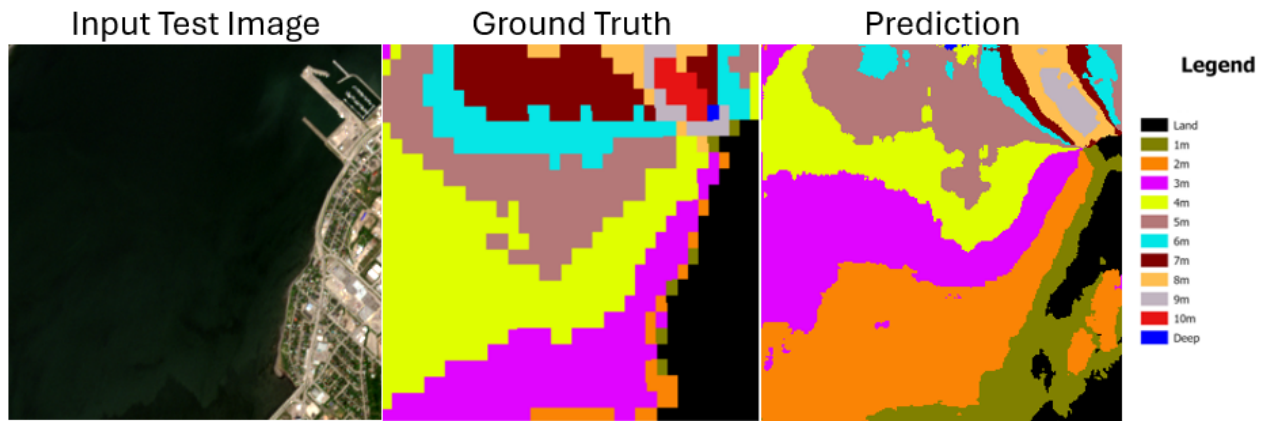


Figure 5.11: Rimouski U-Net predictions

the 10m class was inaccurately predicted and may have been classified as Deep. In addition, the 8m class displayed higher scores with an F1 score of 0.616, recall of 0.629, and precision of 0.604.

Table 5.9: Rimouski U-Net prediction scores

Class	Precision	Recall	F1
Land	0.998	0.794	0.885
1m	0.221	0.623	0.326
2m	0.142	0.544	0.225
3m	0.408	0.356	0.38
4m	0.719	0.256	0.378
5m	0.23	0.241	0.236
6m	0.193	0.171	0.181
7m	0.224	0.464	0.302
8m	0.604	0.629	0.616
9m	0.447	0.276	0.341
10m	0.002	0.0	0.0
Deep	0.021	1.0	0.042
Mean	0.351	0.446	0.326

5.2.1.4 Freshwater Model Training

Figure (5.12) shows that the freshwater model performed better than the saltwater model during training, as U-Net achieved a loss of 0.121 and an F1 score of 0.7. The freshwater model obtained a validation loss of 0.278, which reflects the model's ability to generalize and perform well on unseen data, before overfitting.

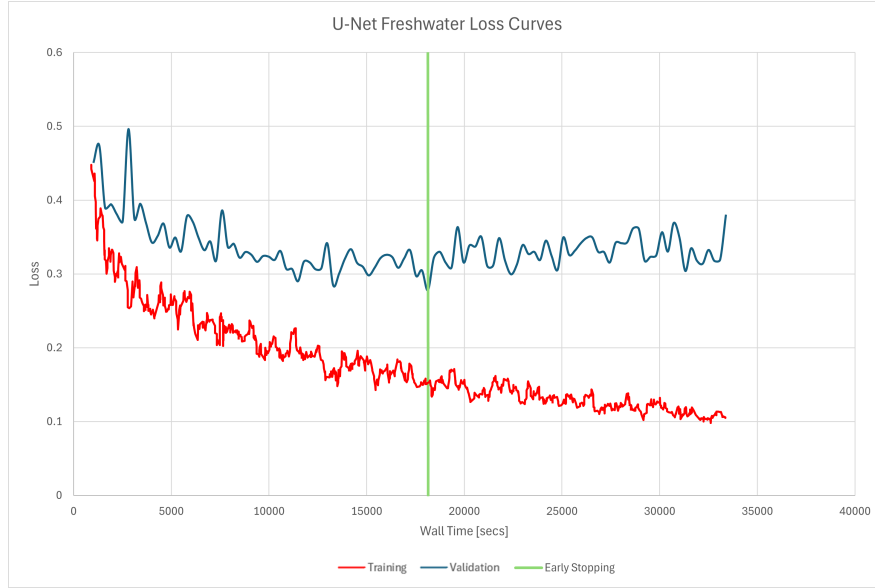


Figure 5.12: U-Net freshwater loss curves

Table 5.10: Best training and validation epoch scores of U-Net freshwater model

	Training	Validation
Loss	0.121	0.278
Precision	0.606	0.533
Recall	0.585	0.526
F1	0.593	0.528
Pixel Accuracy	0.740	0.583

An F1 score of 0.528 was obtained, nearly 10% higher than that of the saltwater model. The class-wise scores also reveal that the model maintained roughly consistent accuracy scores across most classes, as an F1 score of more than 0.4 was obtained for depths between 0m and 9m. The freshwater model accurately classified the coastline (Land) and achieved an F1 score of 0.872, nearly 20% higher than the saltwater model. The model also performed better on the 10m class and acquired an F1 score of 0.389, a recall of 0.355, and a precision of 0.43. This is notably better than the saltwater model and implies that spectral responses from freshwater are more easily distinguishable. The model could differentiate the 10m class from the Deep class and achieved an F1 score of 0.389. These high scores reveal that fresh water has spectral signatures unique to one-metre depth classes.

Figure (5.13) compares the predicted classified image to the ground truth label and reveals that

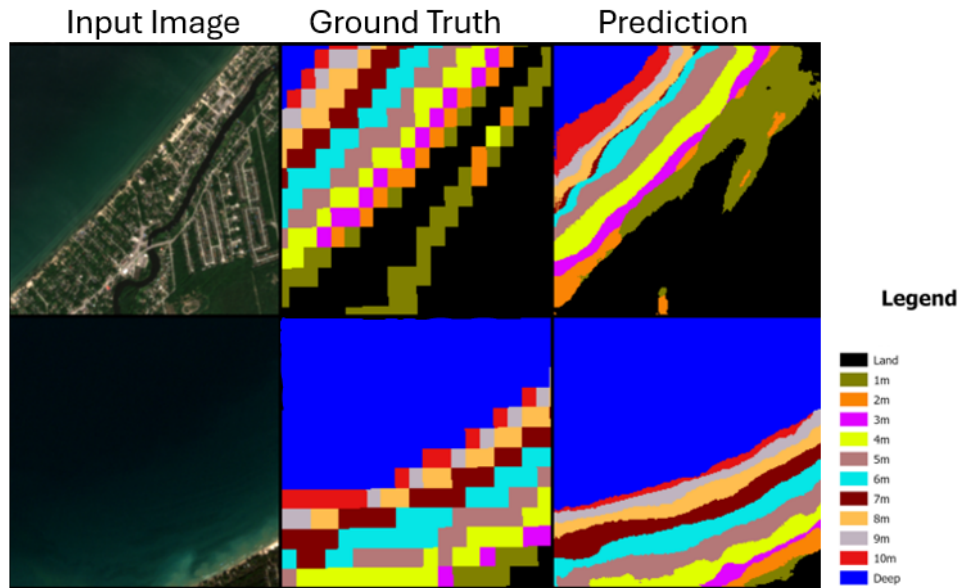


Figure 5.13: U-Net freshwater validation predictions

Table 5.11: U-Net freshwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.889	0.857	0.872
1m	0.483	0.434	0.457
2m	0.545	0.542	0.544
3m	0.524	0.459	0.489
4m	0.473	0.488	0.481
5m	0.462	0.568	0.509
6m	0.52	0.441	0.477
7m	0.434	0.456	0.445
8m	0.445	0.394	0.418
9m	0.383	0.433	0.411
10m	0.43	0.355	0.389
Deep	0.811	0.875	0.841

U-Net can predict and identify the coastline and the spatial relationships between depth classes. However, the model fails to identify inland water bodies or lagoons separated from the larger body of water. The model tends to misclassify adjacent classes, especially in deeper waters. This is expected as the spectral reflectance from deeper waters is much lower and cannot be differentiated.

5.2.1.5 Model Transferability in Athol

The freshwater model was then tested on the Athol study area, as shown in Figure (5.14), and the prediction accuracy scores are listed in Table 5.12.

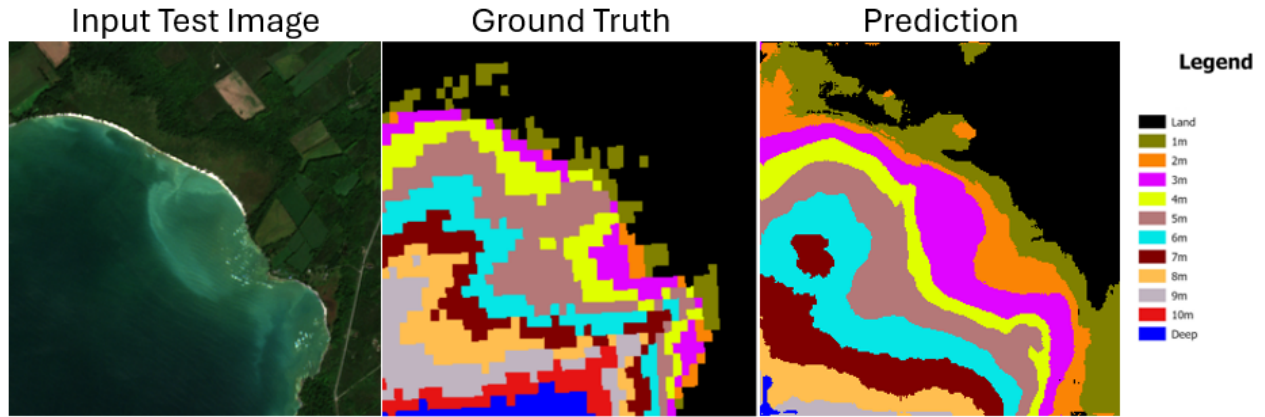


Figure 5.14: Athol U-Net predictions

Table 5.12: Athol U-Net prediction scores

Class	Precision	Recall	F1
Land	0.998	0.671	0.802
1m	0.095	0.453	0.157
2m	0.537	0.695	0.606
3m	0.57	0.67	0.616
4m	0.59	0.714	0.646
5m	0.467	0.584	0.519
6m	0.321	0.399	0.355
7m	0.263	0.307	0.283
8m	0.41	0.269	0.325
9m	0.395	0.294	0.337
10m	0.406	0.177	0.247
Deep	0.521	0.946	0.671
Mean	0.464	0.515	0.464

The performance of the freshwater model surpassed that of the saltwater model, as it achieved a mean F1 score of 0.464, recall of 0.515, and precision of 0.464. These scores, which are close to the validation scores in Table 5.10, prove that the freshwater model was well generalized, which allowed it to perform well on the unseen test data. Shallow waters below 5m were predicted accu-

rately, and F1 scores above 0.5 were obtained. This is due to the distinguishable spectral signatures from shallow waters, which allowed the model to classify image pixels into the right classes. The 4m class was successfully classified as it acquired a recall score of 0.714, meaning that 71% of all 4m class pixels were correctly classified. However, the precision score of 0.59 reveals that other pixels were incorrectly misclassified as 4m. However, the model failed to predict the 1m class meticulously and was the least accurate class. This is contrary to the validation scores and even the saltwater model results as the class obtained an F1 score of 0.157. With a precision score of 0.095, it is evident that more pixels were incorrectly classified as 1m than correctly classified 1m pixels. This is likely due to the lack of data from this class for the model to predict. Beyond 5m, the performance of the model deteriorated as the classes achieved F1 scores around 0.3. This was expected as the strength of light propagating through the water columns decreases with depth due to attenuation. The Deep class acquired an F1 score of 0.671 which indicates that the model could identify most deep-water pixels, with slight misclassifications with pixels belonging to the 9m and 10m classes.

5.2.2 SegNet Results

This section presents the results of the performance of SegNet in saltwater and freshwater environments during training and validation, and model transferability.

5.2.2.1 Saltwater Model Training

Figure (5.15) shows the training and validation loss curves for the SegNet saltwater model, which is like that of the U-Net model, with a training loss of 0.293 and a validation loss of 0.343.

Tables 5.13 and 5.14 summarize the training and validation metrics and the class-wise scores, respectively. SegNet achieved higher validation metrics than U-Net, which also conveys that the model was generalized and performed successfully on unseen data. The network also ran for fewer epochs to achieve a comparable validation loss.

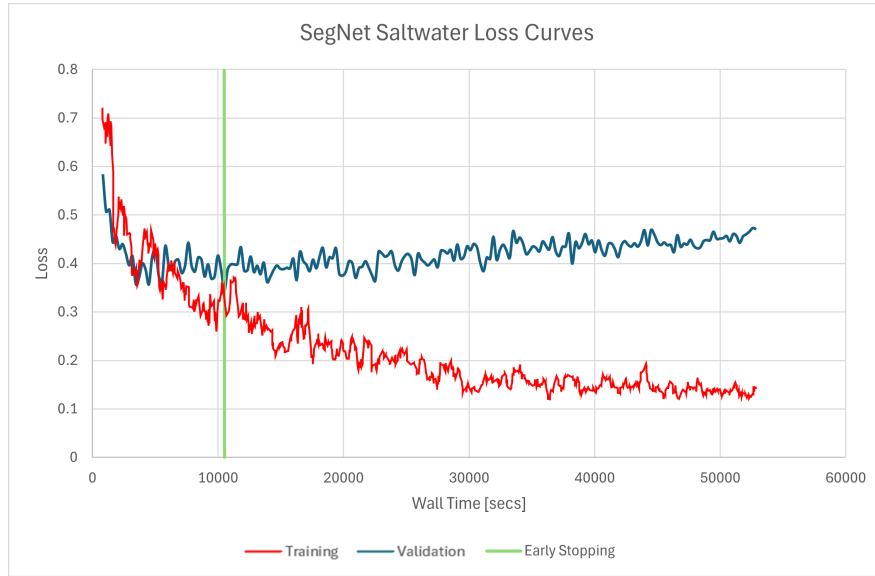


Figure 5.15: SegNet saltwater loss curves

Table 5.13: Best training and validation epoch scores of SegNet saltwater

	Training	Validation
Loss	0.293	0.343
Precision	0.548	0.449
Recall	0.534	0.443
F1	0.532	0.441
Pixel Accuracy	0.633	0.509

The saltwater model achieved a mean training F1 score of 0.532 and a mean validation F1 score of 0.441, slightly higher than the U-Net saltwater model. In terms of the class-wise scores, the difference between SegNet and U-Net is inconsequential. The model attained higher accuracy scores in shallow waters, with F1 scores approximately equal to 0.5 for depths between 0m and 4m. The shorter water column enables more light to reflect off the seafloor and due to the intensity, the segmentation network can differentiate one-metre depth classes. The 1m class obtained a considerably high F1 score of 0.519, implying that the model can classify and identify 1m depth pixels. The 3m class also obtained a high F1 score and recall of 0.63, indicating that 63% of the 3m pixels were correctly classified. The Land class achieved high scores which strengthens the notion that these

models can consistently and accurately identify the coastline.

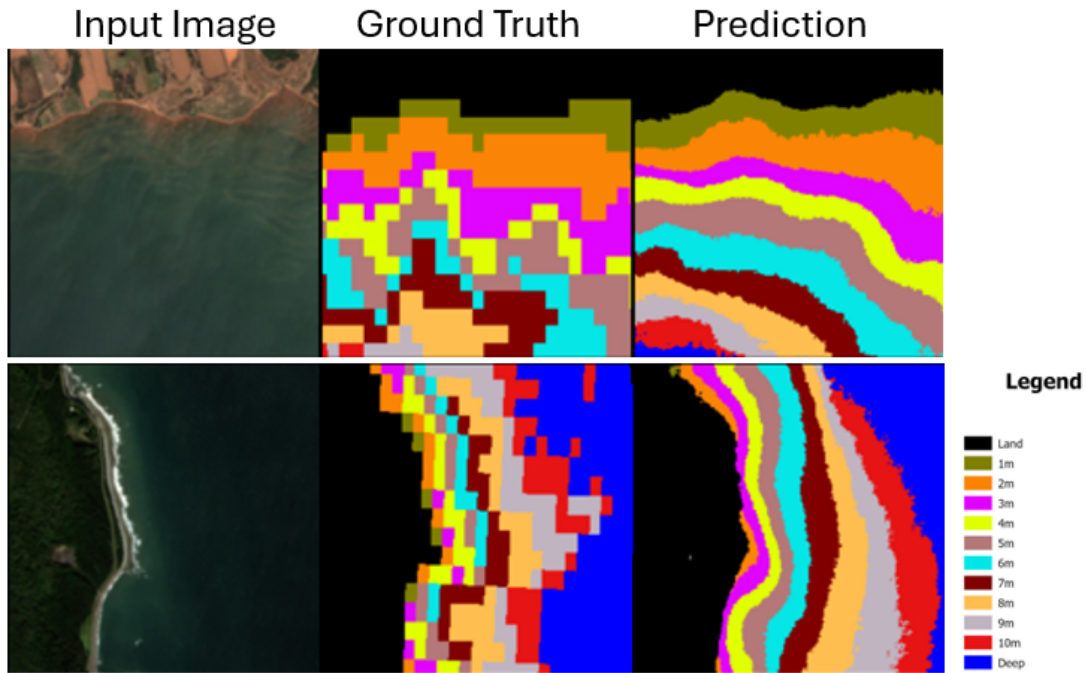


Figure 5.16: SegNet saltwater validation predictions

Table 5.14: SegNet saltwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.727	0.695	0.711
1m	0.568	0.477	0.519
2m	0.551	0.569	0.56
3m	0.472	0.63	0.539
4m	0.41	0.464	0.435
5m	0.397	0.388	0.392
6m	0.358	0.27	0.308
7m	0.335	0.274	0.301
8m	0.327	0.245	0.28
9m	0.323	0.269	0.294
10m	0.274	0.197	0.229
Deep	0.64	0.833	0.724

As the water depth increased, the validation accuracy decreased slightly. F1 scores of roughly 0.3 were obtained for depths between 5m and 9m. These lower scores represent the spectral similarities within this depth range, which cannot be differentiated. The precision scores ranging from 0.245

to 0.338 further explain that there are misclassifications among these classes. The 10m class had the lowest F1, recall, and precision scores, 0.229, 0.197, and 0.274, respectively, which, once again implies that approximately 80% of this depth class is misclassified. Figure (5.16), the 10m class is predicted to cover a greater area than the label. The Deep class had the highest F1 score of 0.724 and the highest recall of 0.833, revealing that the SegNet saltwater model finds it easy to classify deep waters. This may be due to the volume of data containing deep water pixels available for training. Figure (5.16) reveals that SegNet can accurately identify the coastline better than U-Net, with few occurrences of misclassifying the 1m class as Land. In addition, the saltwater model tends to underpredict in deeper waters, as the 9m and 10m classes appear to encroach into the Deep class. However, it is noteworthy that the model can learn the spatial relationships between each class and use this knowledge in the prediction process.

5.2.2.2 Model Transferability in Graham Island

To test the model's performance, the test images from Graham Island were introduced to the network, and the predictions are shown in Figure (5.17). The class-wise accuracy scores are also summarized in Table 5.15.

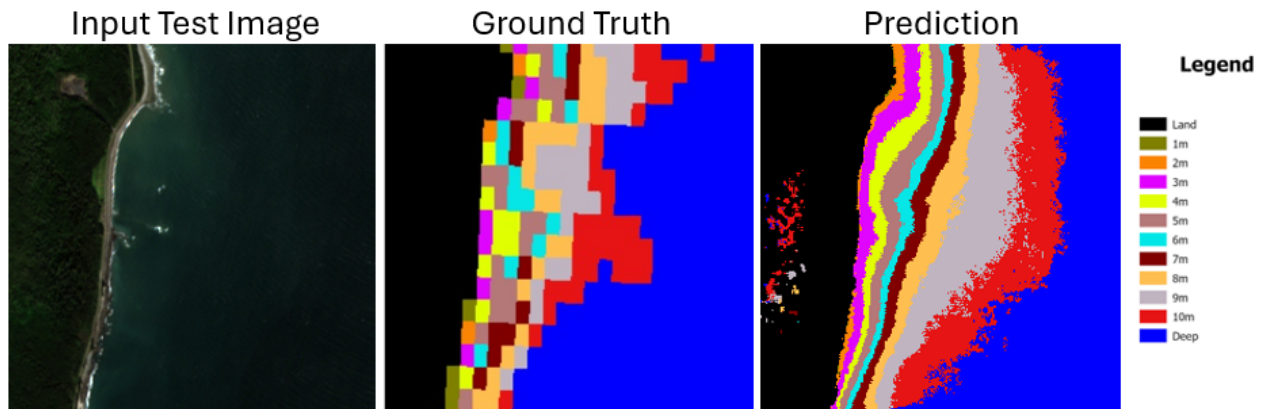


Figure 5.17: Graham Island SegNet predictions

As was evident from the U-Net saltwater model, the SegNet saltwater model fails to make accurate predictions on new, unseen data from a different location. The model obtained an F1 score of 0.332, like U-Net, stipulating that most image pixels were incorrectly predicted. The Land and

Deep classes were accurately predicted with the Land class having an F1 score of 0.931 and recall of 0.895. As Figure (5.17) shows, the coastline is predicted accurately with a precision of 0.97, meaning that very few pixels, roughly 3%, were misclassified as Land. The Deep class obtained an F1 score of 0.688, recall of 0.886, and precision of 0.562.

Table 5.15: Graham Island SegNet prediction scores

Class	Precision	Recall	F1
Land	0.97	0.895	0.931
1m	0.11	0.112	0.111
2m	0.381	0.182	0.247
3m	0.235	0.304	0.265
4m	0.373	0.25	0.299
5m	0.295	0.261	0.277
6m	0.093	0.107	0.1
7m	0.304	0.217	0.253
8m	0.328	0.277	0.301
9m	0.284	0.558	0.376
10m	0.125	0.144	0.134
Deep	0.562	0.886	0.688
Mean	0.338	0.350	0.332

The high accuracy scores signify that trained DCNNs can identify deep-water pixels from different bathymetric environments to a certain degree. However, the precision score reveals that more classes are mistakenly classified as Deep. This is likely the adjacent 10m class, which acquired a low F1 score of 0.134, recall of 0.144, and precision of 0.125. The 10m class is consistently misclassified, with the model only successfully predicting 14% of the pixels. The 1m and 6m classes also obtained low accuracy scores, while all other classes acquired scores close to 0.3. The 6m class had the lowest precision score of 0.093, which implies that the model could not differentiate 6m from 5m or 7m. This is likely the case for the 1m class, which obtained an F1 score of 0.111, recall of 0.112, and precision of 0.11. The model fails to distinguish this class from the other shallow classes. This consistent misclassification of the 1m class is likely due to the steep sloping seafloor of Graham Island. As shown in Figure (5.17), the slope of the seafloor changes rapidly and with the spatial resolution of the satellite image, various depth classes may occur within

a 10m x 10m pixel.

5.2.2.3 Model Transferability in Rimouski

As for the performance of the saltwater model on the Port of Rimouski, there was no improvement but a significant drop in accuracy, as shown in Table 5.16 and Figure (5.18).

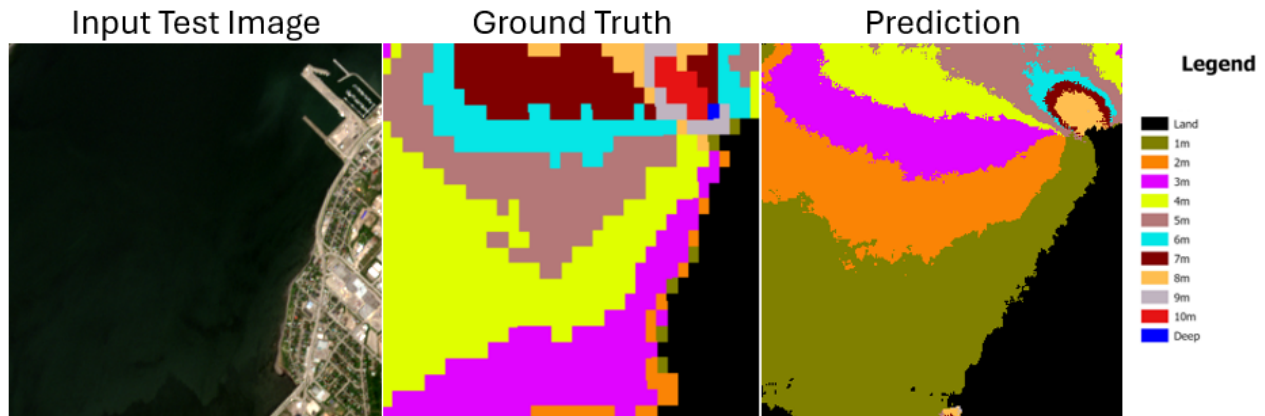


Figure 5.18: Rimouski SegNet predictions

The model achieved lower scores for Rimouski than Graham Island, with an F1 of 0.198, recall of 0.273, and precision of 0.229. This is contrary to the behaviour of the U-Net saltwater model which performed better in Rimouski. All the deep-water classes were poorly predicted with F1 scores rarely exceeding 0.1. In particular, the 10m class obtained scores close to 0. The Deep class acquired an F1 of 0.049 and precision of 0.025, but a recall of 1.0, which signifies that 100% of the Deep class pixels were correctly predicted but many other classes were mistakenly assigned to this class. The shallow classes were predicted more accurately, with the 3m class having the highest F1 score of 0.413, recall of 0.388, and precision of 0.441. These higher scores in the shallower waters show that the spectral responses from these classes are high and intense for the segmentation network to distinguish. The lower scores of the 1m class also reflect the dynamic nature of Rimouski as the model cannot accurately predict this depth range. This implies that the model must be trained in the study area to make accurate predictions that fit the environmental conditions present. The decrease in accuracy as depth increases indicates that the spectral intensity reflected from deeper waters is extremely low for the network to adequately classify.

Table 5.16: Rimouski SegNet prediction scores

Class	Precision	Recall	F1
Land	0.975	0.677	0.799
1m	0.186	0.302	0.23
2m	0.204	0.292	0.24
3m	0.441	0.388	0.413
4m	0.456	0.102	0.166
5m	0.091	0.096	0.094
6m	0.075	0.068	0.071
7m	0.083	0.148	0.106
8m	0.152	0.153	0.152
9m	0.063	0.044	0.052
10m	0.002	0.002	0.002
Deep	0.025	1.0	0.049
Mean	0.229	0.273	0.198

5.2.2.4 Freshwater Model Training

Figure (5.19) shows that during training, SegNet achieved a loss of 0.193 and an F1 score of 0.5. The freshwater model obtained a validation loss of 0.256, much lower than U-Net, which reflects the model's ability to generalize on unseen data. However, the validation metrics are lower than those obtained by U-Net.

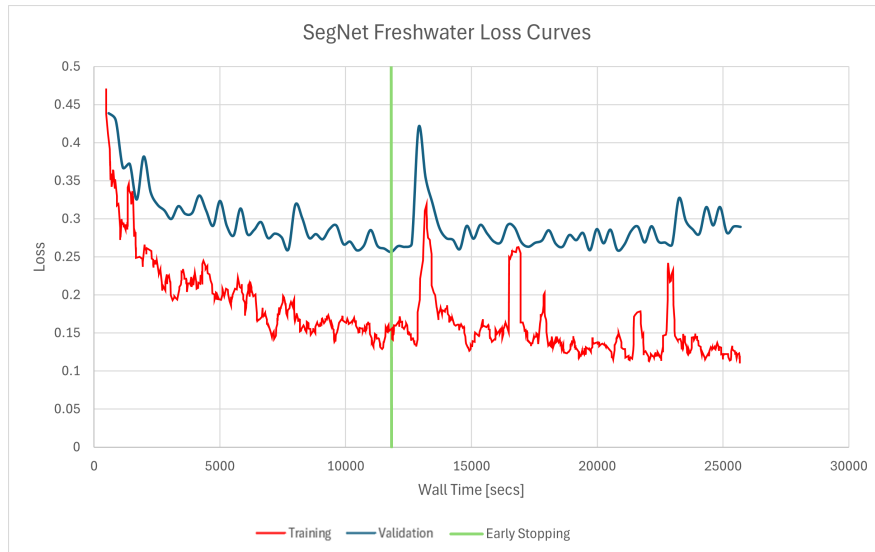


Figure 5.19: SegNet freshwater loss curves predictions

Table 5.17: Best training and validation epoch scores of SegNet freshwater

	Training	Validation
Loss	0.193	0.256
Precision	0.522	0.515
Recall	0.496	0.503
F1	0.502	0.503
Pixel Accuracy	0.684	0.576

An F1 score of 0.503 was obtained, higher than that of the saltwater model. The higher performance of the freshwater model is due to all classes acquiring a score greater than 0.3. The Land and Deep classes have F1 scores greater than 0.8, indicating that the model can predict the coastline and deep waters accurately. There is little difference in accuracy between the shallow waters and the deeper water, as depths below 3m and beyond 6m have approximately the same accuracy scores. This reveals that the SegNet freshwater model can learn the specific patterns and properties of the spectral reflectance that belong to each depth class and use this knowledge in its prediction. This behaviour is unique to fresh water as the saltwater model exhibited a decrease in accuracy with depth. Though the 10m class does have the lowest recall score of 0.245, it maintains a precision score (0.457) equivalent to the rest of the classes, showing that nearly 50% of the pixels classified as 10m did belong to that class.

The 4m and 5m classes have high F1 scores of 0.528 and 0.507, respectively. This conveys that sufficient training data from these classes was available for training to enable the model to make better predictions. With the 1m class expected to be successfully predicted, the lower F1 score of 0.355 and recall of 0.29 imply that insufficient data in this class was collected. As the 1m class is important for identifying the shoreline and is the most dynamic part of the coastal zone, the DCNNs must receive adequate data to successfully classify this class. It is evident from Figure (5.20) that poorly classifying the 1m class does impact the prediction of the shoreline, as Land is incorrectly predicted as 1m. In deeper waters, the model tends to overpredict and greater parts of the image are classified as Deep.

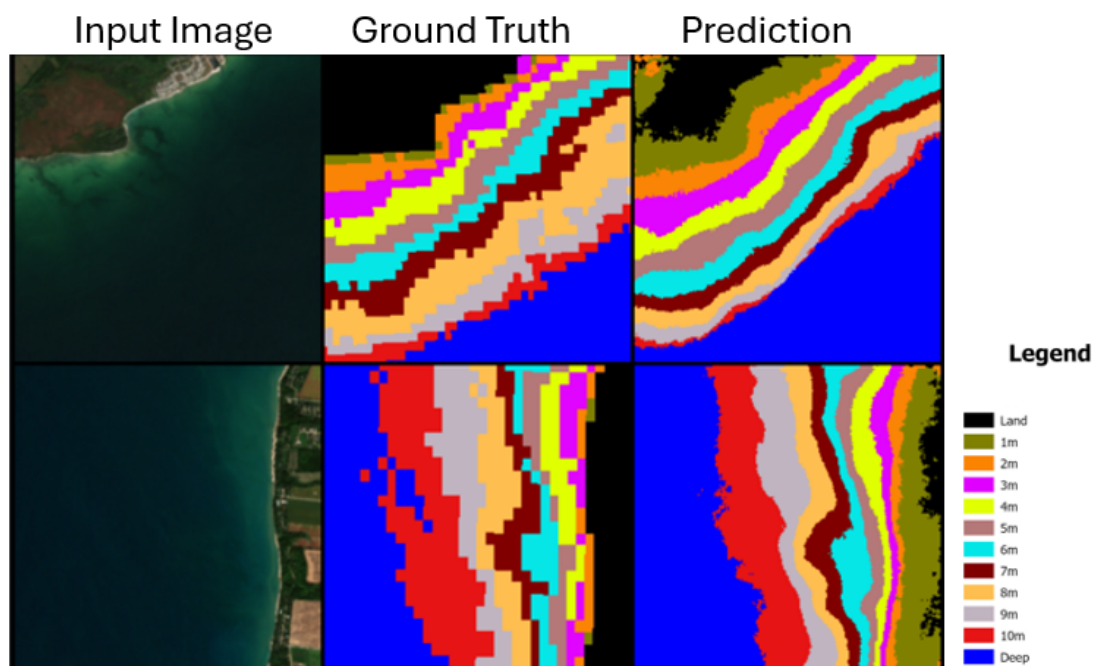


Figure 5.20: SegNet freshwater validation predictions

Table 5.18: SegNet freshwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.836	0.898	0.866
1m	0.458	0.29	0.355
2m	0.488	0.447	0.467
3m	0.473	0.523	0.497
4m	0.501	0.557	0.528
5m	0.479	0.538	0.507
6m	0.5	0.449	0.478
7m	0.429	0.373	0.399
8m	0.409	0.386	0.397
9m	0.364	0.399	0.38
10m	0.457	0.245	0.319
Deep	0.775	0.925	0.843

5.2.2.5 Model Transferability in Athol

Figure (5.21) shows the freshwater model prediction of the Athol study area and Table 5.19 lists the class-wise scores.

The performance of the SegNet freshwater model is like that of the U-Net freshwater model, as it achieved a mean F1 score of 0.461, a recall of 0.499, and a precision of 0.469. The scores reveal

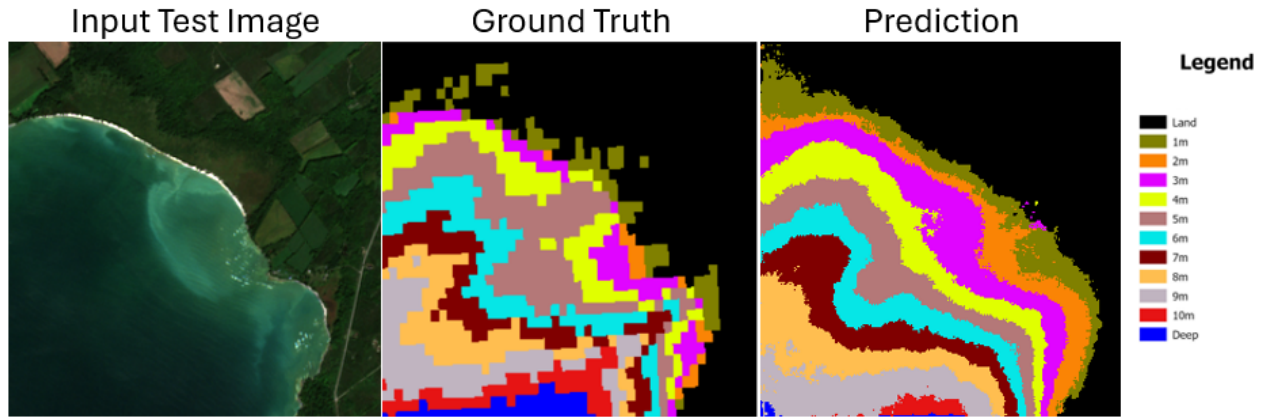


Figure 5.21: Athol SegNet predictions

Table 5.19: Athol SegNet prediction scores

Class	Precision	Recall	F1
Land	1.0	0.673	0.805
1m	0.131	0.371	0.194
2m	0.448	0.644	0.528
3m	0.506	0.657	0.572
4m	0.587	0.413	0.485
5m	0.503	0.492	0.497
6m	0.441	0.429	0.435
7m	0.453	0.349	0.394
8m	0.409	0.315	0.356
9m	0.303	0.488	0.374
10m	0.407	0.223	0.288
Deep	0.445	0.934	0.602
Mean	0.469	0.499	0.461

that the freshwater model could generalize to unseen data and acquired consistent scores across the entire depth range, with a slight decrease in accuracy in the deeper waters. This proves that fresh water is more transparent than salt water and generates distinguishable spectral signatures, allowing SegNet to classify image pixels into the right classes. In the Athol region, the 2m and 3m predictions were fairly accurate, with F1 scores of 0.528 and 0.572, which show that the model could differentiate the spectral responses from this depth range and classify them. By having recall scores above 0.6, most of the pixels were correctly classified. As expected from the validation results, the model failed to accurately predict the 1m class, which had the lowest F1 and precision

scores of 0.194 and 0.131, respectively. However, Land class was predicted accurately with 100% precision, meaning that the model knows how to differentiate between land and water and did not mistakenly classify any water pixels as Land. This is evident in Figure (5.21), where the coastline is sharply outlined. The Deep class acquired an F1 score of 0.602, with a recall of 0.934, which indicates that the model was extremely capable of identifying this class.

5.2.3 DeepLabv3+ Results

The performance of DeepLabv3+ in saltwater and freshwater environments during training and validation is presented, followed by its performance during model transferability.

5.2.3.1 Saltwater Model Training

The training and validation loss curves for the DeepLab saltwater model are displayed in Figure (5.22). DeepLabv3+ obtained the lowest loss of the three DCNNs, as summarized in Table 5.20. With a training loss of 0.122 and a validation loss of 0.323, the model was trained adequately to predict depth from unseen satellite image data.

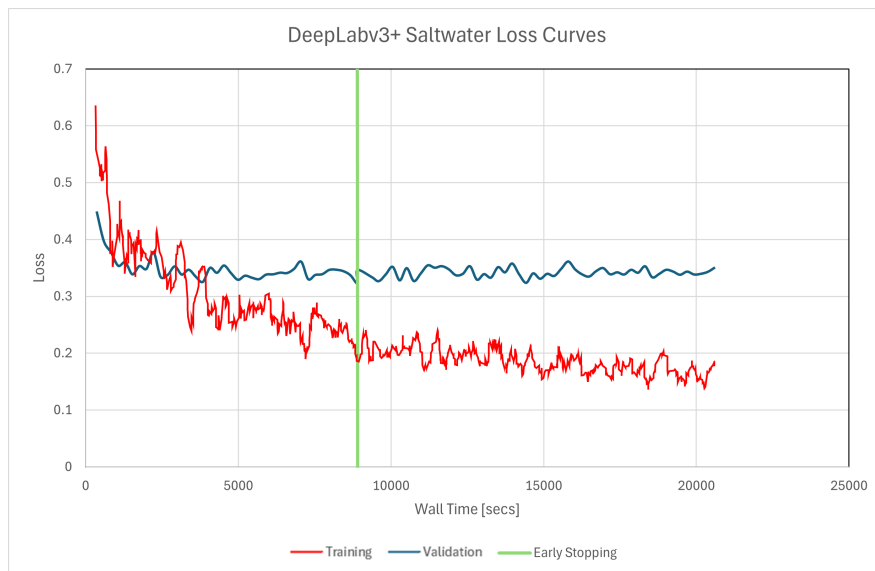


Figure 5.22: DeepLabv3+ saltwater loss curves

During validation, the DeepLabv3+ saltwater model achieved accuracy scores comparable to those

Table 5.20: Best training and validation epoch scores of DeepLabv3+ saltwater

	Training	Validation
Loss	0.122	0.323
Precision	0.618	0.463
Recall	0.608	0.457
F1	0.610	0.454
Pixel Accuracy	0.691	0.532

from U-Net and SegNet, with a training F1 score of 0.610 and a validation score of 0.454. The similarity in validation performance reveals that all DCNNs can be trained to avoid overfitting and generalize to unseen data. However, the low accuracy metrics indicate that the models could benefit from more training data and regularization techniques.

Table 5.21: DeepLabv3+ saltwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.741	0.754	0.747
1m	0.633	0.52	0.571
2m	0.569	0.642	0.603
3m	0.511	0.583	0.544
4m	0.451	0.464	0.457
5m	0.456	0.459	0.458
6m	0.364	0.366	0.365
7m	0.305	0.27	0.286
8m	0.331	0.211	0.258
9m	0.327	0.223	0.265
10m	0.266	0.135	0.179
Deep	0.608	0.851	0.709

The class-wise F1 scores decrease with depth, making the shallow waters easier to predict. This behaviour was exhibited by the U-Net and SegNet models, which reinforce the notion that salt water does not emit and reflect sufficient spectral reflectance that can be differentiated into one-metre classes. Salt water contains more sediment which absorbs more light resulting in dark water surfaces that are difficult to classify, as shown in Figure (5.23). Higher accuracy scores were acquired in shallow waters from 0m to 3m, with F1 scores greater than 0.5. Despite saltwater being dense and sediment-loaded, the shallow depths enable enough light to reflect off the seafloor in the di-

rection of the satellite sensor to be classified. Land obtained the highest F1 score of 0.747, while the 1m class obtained a score of 0.571. This signifies the model's ability to classify the coastline by differentiating it from the dynamic 1m class and possibly identifying marshland as shown in Figure (5.23). The 4m and 5m classes were not classified as accurately as the shallower waters and obtained scores around 0.45.

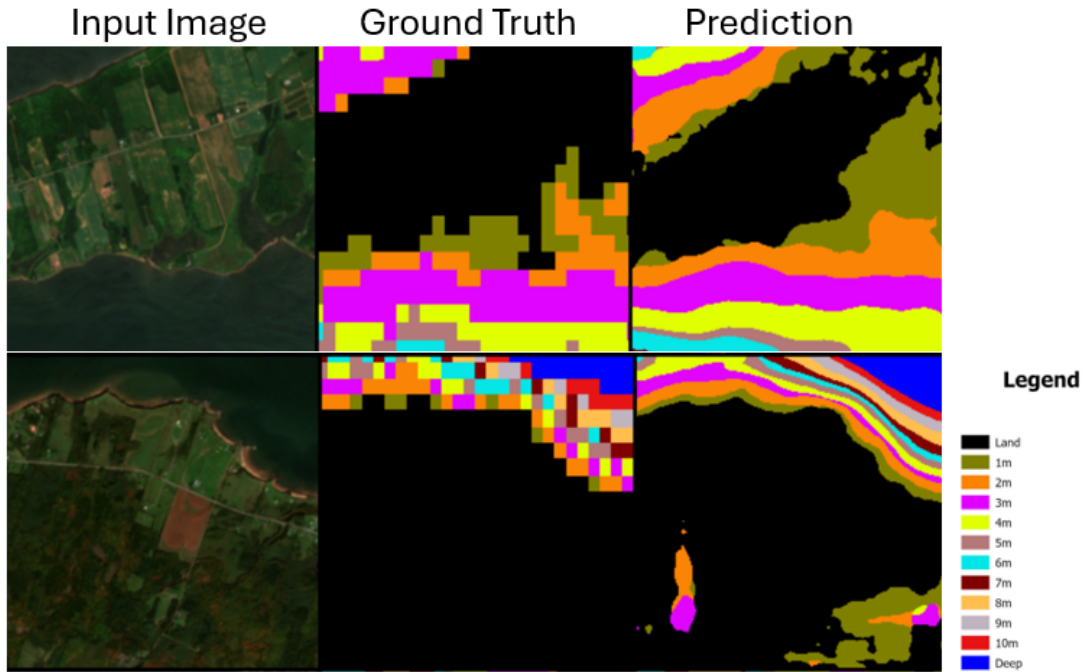


Figure 5.23: DeepLabv3+ saltwater validation predictions

Beyond 5m, the F1 and recall scores dropped below 0.3, while precision for the deeper classes remained above 0.3. The accuracy scores reveal a tendency for the saltwater model to misclassify severely in deep waters. The 10m class had the lowest scores overall, with an F1 of 0.179, recall of 0.135, and precision of 0.229. Less than 14% of the 10m pixels are correctly predicted, which implies that this class has spectral properties that are too identical to the Deep class and other deeper water depths. On the other hand, the Deep class was accurately predicted with an F1 score of 0.709, further reflecting that larger volumes of training data aid the model in making better predictions. As depicted in Figure (5.23), the atrous convolution approach of DeepLabv3+ allows the models to use spatial context to make accurate predictions. The saltwater model can simulate the exact change in slope presented in the ground truth labels. However, due to a lack of training data in

shallower depths, there are instances of misclassification, where parts of land are predicted to be water.

5.2.3.2 Model Transferability in Graham Island

The trained saltwater model was then used to predict the depth from the test images from Graham Island, as shown in Figure (5.24). The model's predictions were inaccurate, as a mean F1 score of 0.296 was acquired. The recall score of 0.321 reveals that roughly 32% of the study area was classified correctly, like the performance of U-Net and SegNet. As all the DCNNs performed poorly in the Hecate Strait, off the coast of Graham Island, it is evident that prediction is difficult in high-latitude saline waters. Coupled with the steep slope of the seafloor, as shown in Figure (5.24), abundant misclassifications are prevalent.

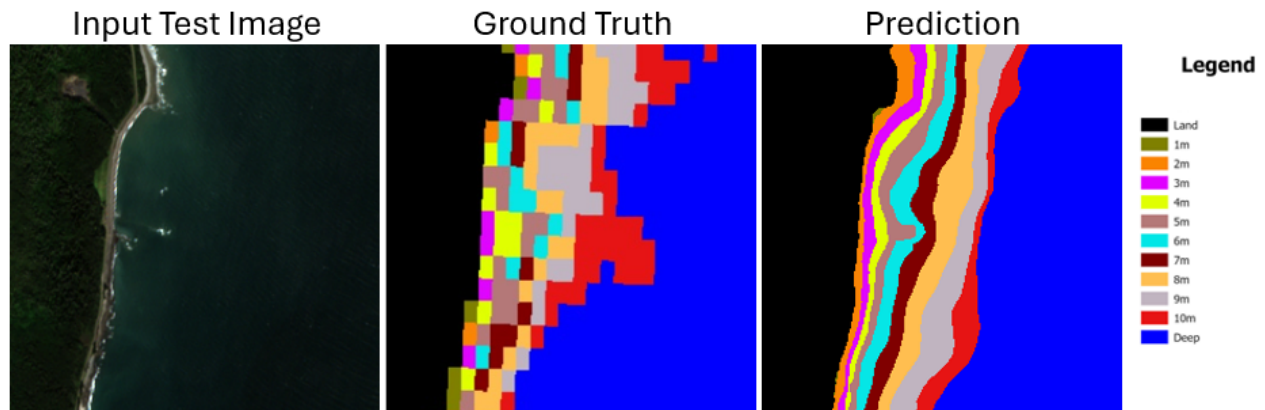


Figure 5.24: Graham Island DeepLabv3+ predictions

From Table 5.22, the DeepLabv3+ saltwater model cannot make accurate predictions from unseen data it has not been trained on, especially data from a location as challenging as Graham Island. The Land and Deep classes were expected to obtain high scores. However, an F1 score of 0.937 from the Land class indicates that the model accurately separates land from water and identifies the coastline. The challenge lies with the 1m class, which obtained F1 and recall scores close to 0. This is expected in a study area with steep slopes. The accuracy reaches 0.3 in deeper waters, with the 6m class achieving higher scores than the adjacent classes. A recall score of 0.525 shows that the saltwater model can identify specific multispectral properties of certain classes and use these

Table 5.22: Graham Island DeepLabv3+ prediction scores

Class	Precision	Recall	F1
Land	0.964	0.912	0.937
1m	0.298	0.04	0.07
2m	0.33	0.216	0.261
3m	0.243	0.362	0.291
4m	0.345	0.331	0.338
5m	0.358	0.21	0.265
6m	0.259	0.525	0.347
7m	0.347	0.097	0.152
8m	0.491	0.111	0.181
9m	0.174	0.038	0.063
10m	0.229	0.004	0.009
Deep	0.47	1.0	0.639
Mean	0.376	0.321	0.296

to make predictions. The Deep class obtained an F1 score of 0.639 and 100% recall. However, depths closely associated with the Deep class, 9m and 10m, received the lowest scores, with an F1 of 0.063 and 0.009, respectively. This consistent misclassification of the deeper classes proves that SDB can only be reliable to a certain depth, such as 6m, with data limitations. Beyond this depth, deep water depths can be grouped into a single class.

5.2.3.3 Model Transferability in Rimouski

Table 5.23 shows that the model did not perform better in Rimouski. The F1 score decreased to 0.201, as did the recall and precision. The DeepLabv3+ model behaved just as the SegNet model by performing poorly in Rimouski, while U-Net maintained its accuracy and even performed slightly better in this test site.

The model behaved similarly with Graham Island, where the Land class achieved a high F1 score of 0.861, while the rest were poorly predicted. However, the Deep class obtained an F1 score of 0.039, while having 100% recall, indicating that deep-water pixels were correctly classified but all other depth classes were mistakenly assigned to this class. This implies that Rimouski, with a mixture of fresh and salt water, has multispectral properties in its deeper waters that are inherently

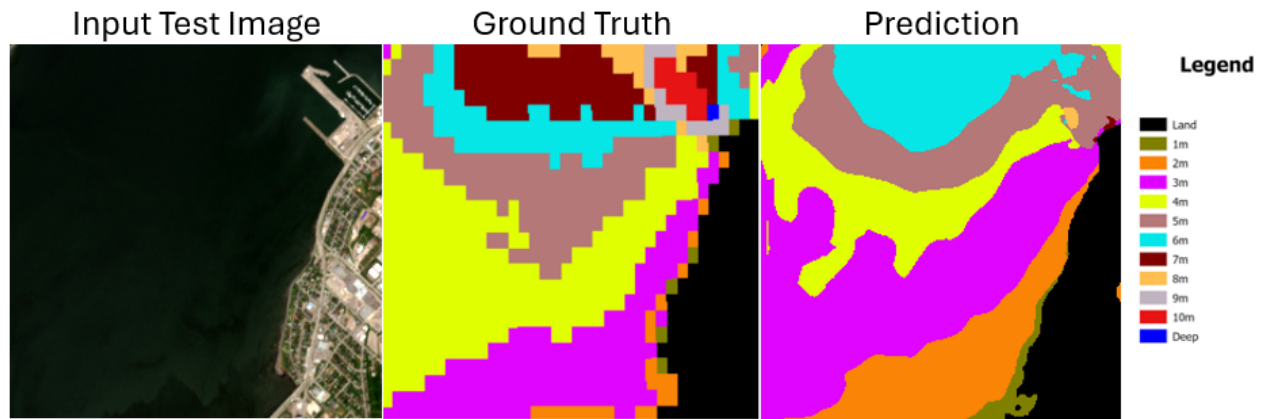


Figure 5.25: Rimouski DeepLabv3+ predictions

Table 5.23: Rimouski DeepLabv3+ prediction scores

Class	Precision	Recall	F1
Land	0.994	0.759	0.861
1m	0.118	0.277	0.165
2m	0.109	0.488	0.179
3m	0.678	0.364	0.473
4m	0.48	0.135	0.21
5m	0.078	0.076	0.077
6m	0.027	0.028	0.027
7m	0.049	0.094	0.065
8m	0.243	0.24	0.242
9m	0.079	0.076	0.078
10m	0.0	0.0	0.0
Deep	0.02	1.0	0.039
Mean	0.240	0.295	0.201

different from purely saline deep waters, like those of Graham Island. This is evident from the 10m class, which was entirely misclassified with accuracy scores of 0. Classes deeper than 4m were also severely misclassified, with F1 scores below 0.1, apart from the 8m class with a score of 0.242. In the shallow waters, a higher accuracy was achieved with classes between 0m and 4m having an F1 score greater than 0.1. Only the 3m class obtained considerably higher scores, with an F1 of 0.473, a recall of 0.364, and a precision score of 0.678. After the Land class, the 3m class was the most accurately predicted class. The precision score indicates that over 67% of the pixels assigned to the 3m class were true positives. This high score is likely due to the test site having large areas with its

depth range, as shown in Figure (5.25). These scores emphasize that the saltwater model must be retrained on the test sites to generate accurate predictions due to each site’s unique environmental and bathymetric conditions.

5.2.3.4 Freshwater Model Training

Figure (5.26) shows that the DeepLabv3+ freshwater model performs better than the saltwater model, with a training loss of 0.172 and a validation loss of 0.279. DeepLabv3+ achieved the lowest training loss of all DCNNs but the highest validation loss. However, the validation metrics are high and comparable to those obtained by U-Net.

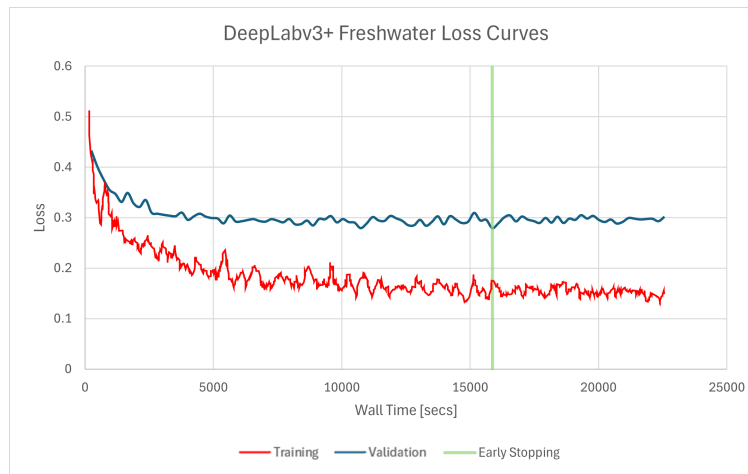


Figure 5.26: DeepLabv3+ freshwater loss curves predictions

Table 5.24: Best training and validation epoch scores of DeepLabv3+ freshwater

	Training	Validation
Loss	0.172	0.279
Precision	0.597	0.526
Recall	0.580	0.522
F1	0.586	0.521
Pixel Accuracy	0.739	0.585

A mean F1 score of 0.521 was acquired, higher than that of the saltwater model. The class-wise scores also reveal that the model maintained consistent accuracy across most classes, as an F1 score of more than 0.4 was obtained for depths between 1m and 9m. The 1m class, which represents

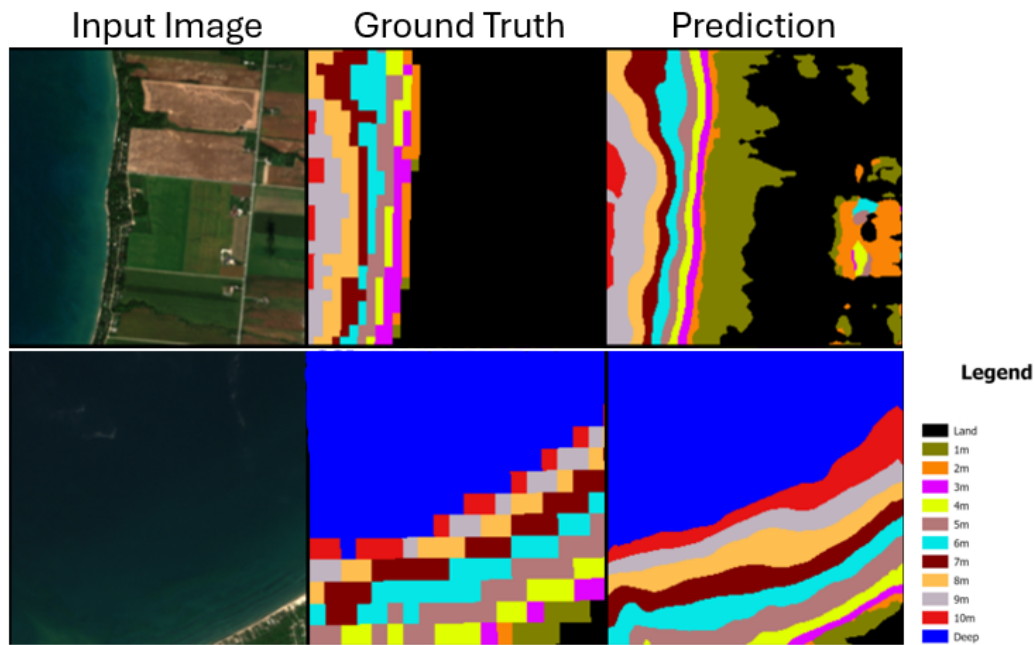


Figure 5.27: DeepLabv3+ freshwater validation predictions

depths from 0m to 1m, had an F1 score below 0.4 and follows the trend that has been observed, which is that this dynamic class is challenging to predict, especially with insufficient training data. The higher performance of the freshwater model is due to all classes acquiring a score greater than 0.3. Due to the transparency of fresh water and the high spectral reflectance from shallow waters with sandy bottoms, the segmentation network struggles to differentiate the coastline from shallow waters of 1m or less as shown in Figure (5.27). There is a slight increase in accuracy in the 3m and 4m classes with F1 scores of 0.516 and 0.529, respectively. Beyond 5m, the model maintains an accuracy of 0.4 for the deep classes, revealing that freshwater allows more light to penetrate the water column and reach the bottom. The model's accuracy drops slightly for the 10m class to an F1 score of 0.366 and increases for the Deep class to 0.846. Though the difference between these two adjacent classes is great, the model can still classify these two classes much better than the saltwater model.

Table 5.25: DeepLabv3+ freshwater class-wise validation scores

Class	Precision	Recall	F1
Land	0.822	0.926	0.871
1m	0.473	0.305	0.371
2m	0.465	0.461	0.463
3m	0.52	0.511	0.516
4m	0.489	0.577	0.529
5m	0.477	0.496	0.486
6m	0.524	0.416	0.464
7m	0.458	0.473	0.465
8m	0.469	0.477	0.473
9m	0.392	0.42	0.405
10m	0.39	0.344	0.366
Deep	0.834	0.858	0.846

5.2.3.5 Model Transferability in Athol

The freshwater model was then tested on the Athol test site, as shown in Figure (5.27), and the prediction accuracy scores are listed in Table 5.26.

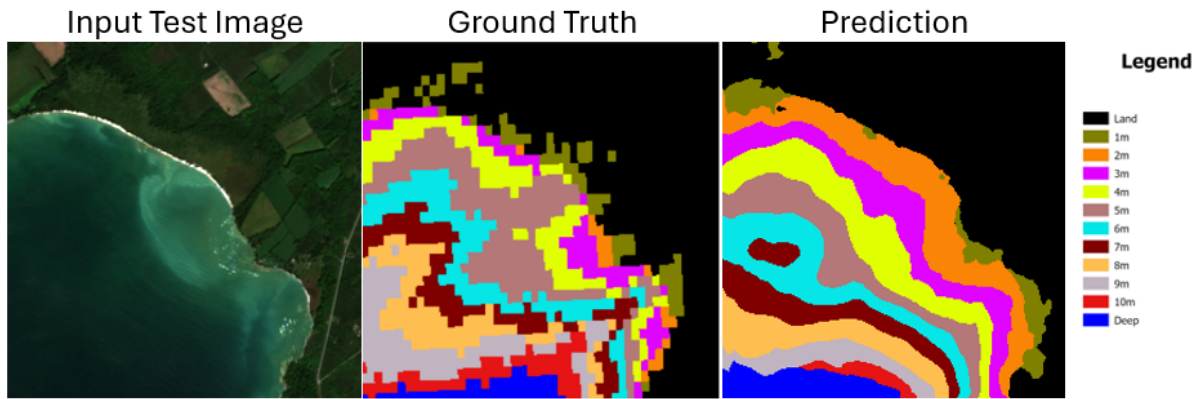


Figure 5.28: Athol SegNet predictions

Table 5.26 shows that the DeepLabv3+ performed the best among all the DCNNs in the Athol coastal region. With an F1 score of 0.517, over half of the image was correctly predicted. The mean F1 score from the validation process was equal to 0.521, proving that the DeepLabv3+ freshwater model was well generalized to perform accurately on new, unseen data. This also reinforces that more variable training data could improve performance and accuracy.

The freshwater model maintained high F1 scores above 0.5 throughout all classes between 2m and

Table 5.26: Athol DeepLabv3+ prediction scores

Class	Precision	Recall	F1
Land	0.996	0.686	0.813
1m	0.198	0.649	0.303
2m	0.515	0.611	0.558
3m	0.58	0.567	0.573
4m	0.661	0.701	0.68
5m	0.65	0.574	0.61
6m	0.591	0.478	0.529
7m	0.552	0.456	0.5
8m	0.529	0.619	0.57
9m	0.374	0.320	0.345
10m	0.252	0.068	0.107
Deep	0.452	0.989	0.62
Mean	0.529	0.560	0.517

8m. This indicates that DeepLabv3+ is very powerful in differentiating the multispectral reflectance from a wide range of depths in fresh waters. This also reveals that SDB can be reliable up to a depth of 8m in transparent, freshwater conditions. In this range, the 4m class obtained the highest accuracy scores with a recall of 0.701. This implies that 70% of the 4m pixels were accurately classified. The 1m class was not successfully predicted, with a precision of 0.198, indicating that many depth classes, possibly even Land, are assigned to this class. This affects the definition of the shoreline, as shown in Figure (5.27). Beyond 8m, the accuracy decreases, where the 9m class has an F1 score of 0.345 and the 10m class has a score of 0.107. This drop in accuracy reflects the decrease in light intensity that reaches the bottom to be reflected. The spectral reflectance from these classes is low, resulting in the model making inaccurate predictions. This is also revealed in the Deep class, which has a low recall score of 0.452, signifying that many of the deep-water pixels are misclassified due to the low spectral reflectance being detected.

5.3 Comparison with Empirical and Physics-based models

The ML models were compared against the empirical and physics-based models, with their class-wise F1 scores summarized in Table 5.27 and 5.28, respectively, across the three test sites.

Table 5.27: Empirical Model F1 scores

Class	Graham Island	Rimouski	Athol
Land	$8.462e^{-6}$	0.0001	$1.308e^{-5}$
1m	0.0	0.084	0.080
2m	0.001	0.091	0.257
3m	0.003	0.071	0.315
4m	0.043	0.037	0.277
5m	0.081	0.012	0.225
6m	0.090	0.011	0.183
7m	0.062	0.013	0.149
8m	0.068	0.019	0.106
9m	0.062	0.033	0.057
10m	0.065	0.038	0.021
Deep	0.609	0.362	0.067
Mean	0.090	0.064	0.145

Table 5.28: Physics-based model F1 scores

Class	Graham Island	Rimouski	Athol
Land	0.0	0.0	0.0
1m	0.046	0.368	0.343
2m	0.010	0.693	0.372
3m	0.002	0.414	0.380
4m	0.004	0.127	0.280
5m	0.006	0.118	0.226
6m	0.009	0.123	0.189
7m	0.010	0.099	0.153
8m	0.004	0.040	0.119
9m	0.002	0.033	0.093
10m	0.0	0.015	0.096
Deep	$6.56e^{-5}$	0.068	0.285
Mean	0.008	0.175	0.211

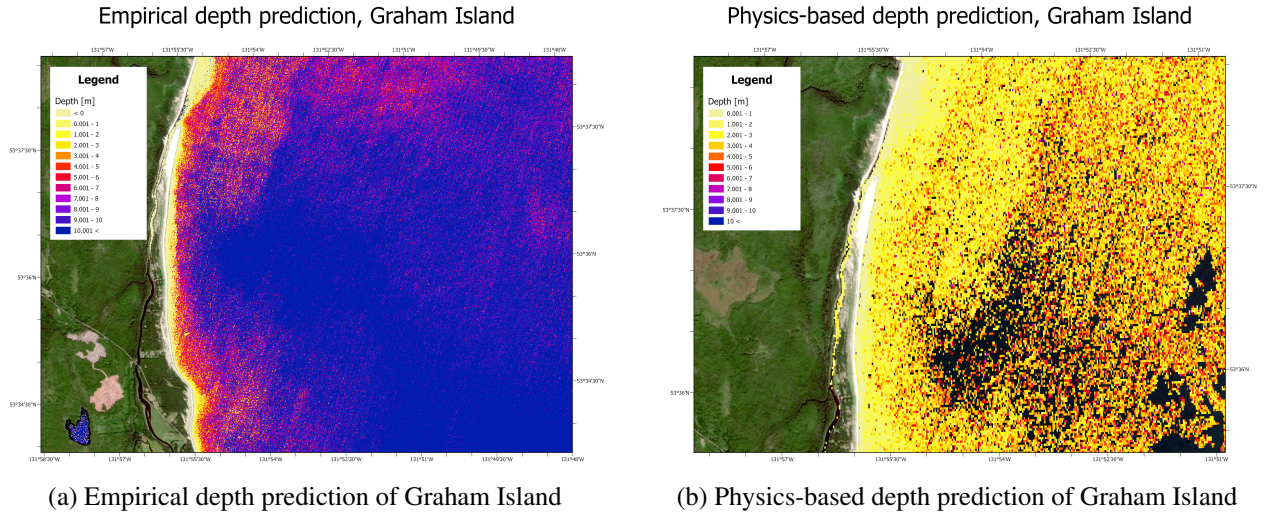


Figure 5.29: Validation of SDB using deterministic models in Graham Island

Figure (5.29) shows the satellite-derived bathymetry of Graham Island using Stumpf's empirical band ratio model and the physics-based model. The physics-based model severely underestimated the deeper waters, while the empirical model correctly identified these waters and obtained a mean F1 score of 0.09. Figure (5.29b) also shows gaps in the image, which reflect the regions where inversion failed. This tends to occur in deep waters where the observed R_{rs} is less than the deep-sea signal. This underestimation is due to the complexity of the saline waters of Graham Island, which emit and reflect higher R_{rs} because of backscattering. The sediments in the water intensely scatter light, preventing it from penetrating deeper depths.

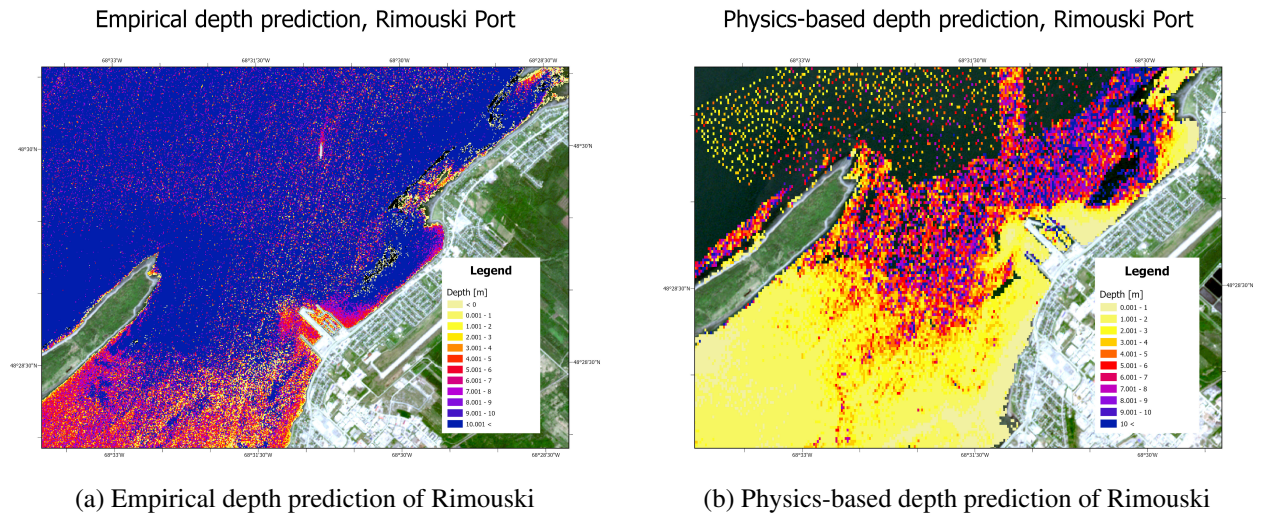


Figure 5.30: Validation of SDB using deterministic models in Rimouski

The physics-based model behaved similarly in Rimouski, underpredicting the deeper waters. Just as the Random Forest underpredicted the deep waters in Rimouski, this reveals that the reflectance from deeper water includes specular reflectance. The model still managed to predict the shallow waters to obtain a mean F1 score of 0.174. However, the empirical model overestimated the shallow water and acquired an F1 score of 0.06.

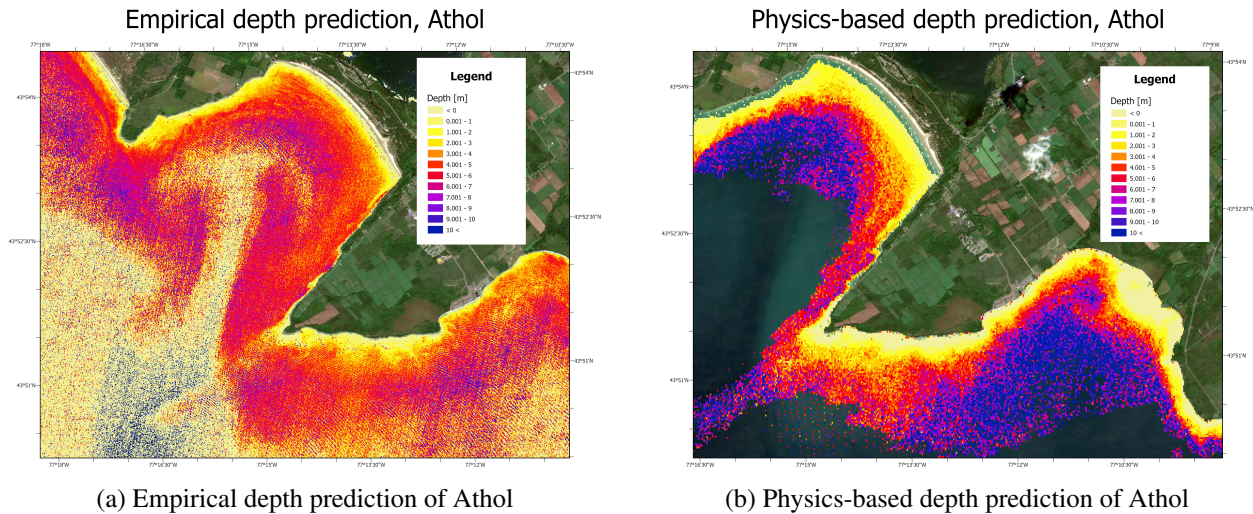


Figure 5.31: Validation of SDB using deterministic models in Athol

The physics-based model succeeded in predicting the gradual change in depth in Athol from the coastline to the deep open waters. It obtained a mean F1 score of 0.211, higher than the freshwater RF model. This proves that by considering the IOPs and sea bottom character, the prediction of depth is more accurate. The empirical model underpredicted the deep waters. This is because the modelling approach requires vertical calibration of the pseudo-depths to the actual depth, predictions are extremely dependent on the correlation between these variables. Despite obtaining an R^2 of 0.82 between the pseudo-depths and in-situ depth, only calibrating to a depth of 13m does not allow the model to make accurate predictions beyond this depth. As water depth increases and most light has been absorbed and cannot reach the bottom, light is backscattered into the atmosphere by sediments and generates a false bathymetry.

5.4 Performance Comparison

In this section, the performance of each model was compared according to its performance during training, validation, and testing. Finally, the impacts of fresh and saltwater on SDB modelling were evaluated.

5.4.1 Training and Validation

During training and validation, as shown in Figure (5.32), the freshwater RF model obtained the highest mean F1 score of 0.55 amongst all ML models. This was due to the number of predictive variables the Random Forest used and their various combinations. This proves that the performance of machine learning models in SDB could improve by including more spectral bands and IOPs. The freshwater DCNNs did not perform as well as the RF during training and validation, with U-Net achieving a mean F1 score of 0.528, while SegNet and DeepLabv3+ obtained scores just above 0.5. This reveals that the transparency of freshwater allows light to penetrate deeper in the water column, which is essential for better prediction across the entire depth range. This lower performance is likely because the DCNNs were limited to the three RGB channels for learning and prediction and hence cannot exploit the available multispectral and IOP channels.

However, the performance of all ML models dropped in saltwater, due to the concentration of sediments and density, which prevents light from penetrating and reaching the seafloor. The DCNNs obtained F1 scores above 0.4, with DeepLabv3+ obtaining the highest mean F1 score of 0.454. Since the DeepLabv3+ model uses multiscale spatial information to learn intricate patterns from the satellite image, it can make better predictions from even low-reflectance saltwater pixels. As U-Net and SegNet achieved similar scores, this reveals that considering the spatial contextual information from satellite imagery allows models to predict water depth in sediment-laden dark salt water. The RF model acquired the lowest score of 0.310 as it only considers the individual pixel and not the spatial information, which affected its predictive ability in saltwater.

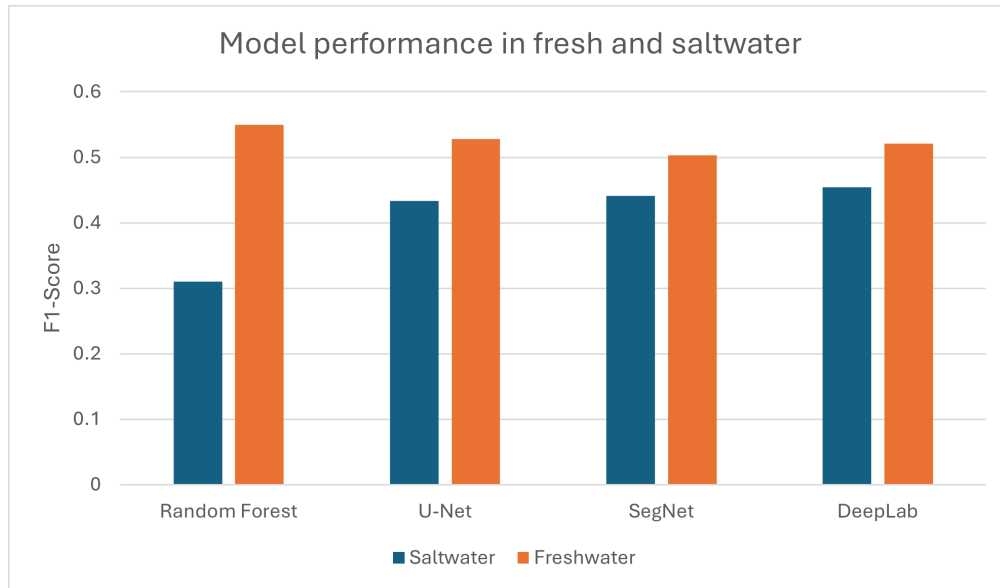


Figure 5.32: Machine learning model validation performance in fresh and saltwater

Regarding consistency, SegNet exhibited the smallest difference in its performance in freshwater and saltwater during training and validation, with a score of 0.503 in freshwater and 0.441 in saltwater. This proves there is little difference between freshwater and saltwater properties, and SegNet can understand and learn these differences. DeepLabv3+ and U-Net also display relatively minor differences, reinforcing the evidence that DCNNs can learn unique patterns in each environment. More saltwater data is likely required for saltwater models to perform well. However, the RF displayed the most significant difference in performance. By neglecting spatial contextual information, the RF model cannot learn the unique patterns in saltwater. This proves that spatial information is essential in predicting water depth in saltwater coastal zones. The strong performance of the RF in freshwater also proves that IOPs can assist in making better predictions; therefore, a physics-based machine learning model that considers spatial contextual information could perform better in saltwater.

5.4.2 Testing

Figure (5.33) compares the F1 scores obtained in each test site by each SDB model.

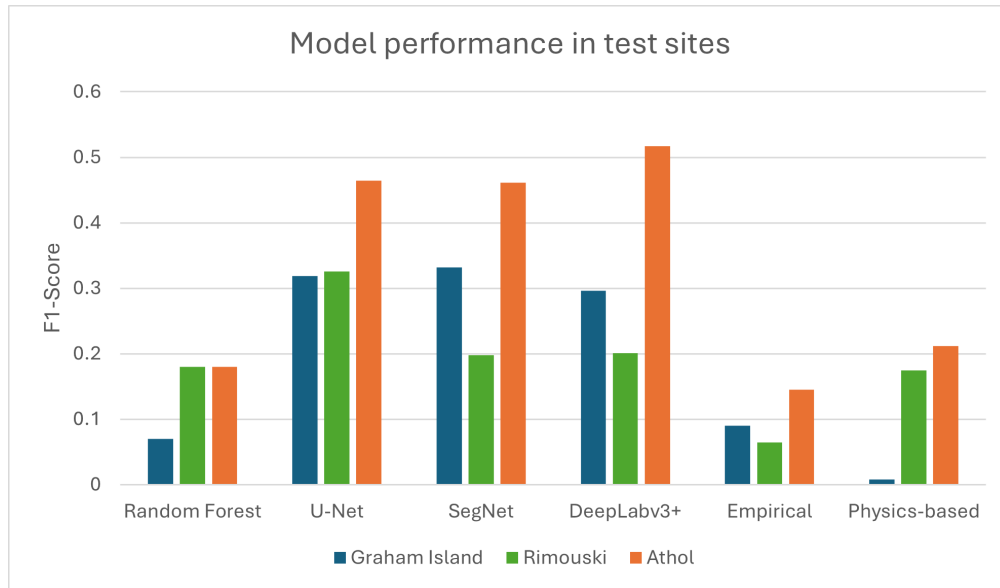


Figure 5.33: SDB model test performance comparison using F1-score

The models performed poorly in all test sites, with only DeepLabv3+ achieving an F1 score above 0.5 in Athol. This reflects the challenge of predicting water depth in high-latitude regions with no optically clear waters. This proves that models cannot generalize to new, unseen data, particularly the RF and deterministic models. As the maximum F1 score achieved during training and validation was only 0.55, these test scores indicate that more varied data and regularization techniques could improve the performance of all models. Models can learn more general and transferable features by including data from a wide range of bathymetric environments. The lower performance in saltwater test sites is due to higher concentrations of salts and dissolved ions, which increase the viscosity of the water column and inhibit light from penetrating further to reach the seafloor. Models performed better in the freshwater test sites, due to the lower concentration of salts and sediments in the water column. With fewer sediments, the water column is more transparent, and light is not severely attenuated and can penetrate further.

The test results also show model behaviour contrary to that observed during training and validation. The empirical model obtained the lowest F1 score of 0.144 for Athol, and the RF model had the second-lowest score of 0.18, though it had the highest validation F1 score. The DeepLabv3+ had

the highest F1 score of 0.517 for Athol and a validation score of 0.521. This indicates that the freshwater model was generalized well and maintained its prediction accuracy in Athol. The water transparency in Athol allowed the DCNNs to perform well as scores above 0.45 were achieved. However, the prediction accuracy of DeepLabv3+ surpassed that of U-Net and SegNet because of the ASSP module that incorporates multiscale feature learning. The poor prediction accuracy by the RF also reveals that water clarity and transparency are not solely responsible for accurate predictions, but capturing spatial information is necessary for predicting water depth. SegNet produced the most accurate predictions for Graham Island with an F1 score of 0.332, while the Physics-based model produced the lowest score of 0.007. This reinforces the notion that saltwater depth is complicated to predict, especially without spatial contextual information. U-Net performed best in Rimouski with a score of 0.326, while all other models obtained lower scores, which means that U-Net's skip connections preserve information from the original input image to use in the final prediction result. As SegNet and DeepLabv3+ achieved roughly similar scores in Graham Island but vastly different scores for Rimouski, this reveals that models trained in purely saltwater conditions cannot perform well in mixed turbid conditions like Rimouski.

5.5 Impact of fresh and saltwater conditions on SDB modelling

The results of the different ML methods indicate that models trained in saltwater conditions perform differently from models trained in freshwater conditions due to their inherent optical properties.

Machine learning saltwater models generate accurate predictions in shallow waters up to 3m deep and can achieve high F1 validation scores of 0.5. Classes such as the 1m class can be predicted with high precision. However, because saline waters have a stronger internal resistance to flow, the deeper waters are poorly predicted because very little light penetrates through the water column. This decreased the saltwater RF model's accuracy to as low as 0.06. Spectral reflectance is too low for the model to differentiate spectral signatures and learn distinct patterns for each class.

All ML models achieved a mean validation F1 score over 0.5 in freshwater and maintained consistent scores across the entire depth range. Freshwater models can accurately predict shallow waters during testing to achieve a 2m class F1 score of 0.572 by SegNet. Similarly, the models can identify deep water with scores as high as 0.671 for U-Net, which reveals that fresh deep waters have specific spectral patterns that can be distinguished from other depths. However, certain depth ranges, such as the 10m class, were poorly predicted, which acquired a test F1 score of 0.107 by DeepLabv3+. This reflects the lack of varied data in this depth range and most likely is misclassified as Deep. The constant misclassification of the 10m class potentially signifies that SDB cannot be reliable up to this depth, even in transparent fresh waters.

This performance comparison reveals the depth ranges where the freshwater and saltwater models are reliable. In freshwater conditions, the model can be reliable to a depth of 8m, as the DeepLabv3+ results from Athol show that F1 dropped from 0.57 for the 8m class to 0.345 for the 9m class. However, U-Net and SegNet convey that 5m may be the cut-off depth as F1 scores drop from 0.5 at 5m to 0.3 – 0.4 beyond this depth. More data and regularization techniques can extend this cut-off depth beyond 8m. In saltwater conditions, models are only reliable to 3m. Due to the greater density of saltwater, in addition to the fact that the Random Forest revealed that the green band contributes most to the prediction process, it becomes difficult to predict water depth beyond 3m, as light cannot penetrate deeper through the water column.

Chapter 6

Conclusions and Future Work

This research investigated the performance of various machine learning models in predicting water depth in both freshwater and saltwater environments using Sentinel-2A multi-spectral satellite imagery. Study areas from diverse Canadian water bodies were selected to train the ML models and test their transferability on unseen data from three test sites. The ML results were compared to those obtained using empirical and physics-based satellite-derived bathymetry models. The results demonstrate that ML model performance can vary significantly, especially across varying depths, depending on water transparency, inherent optical properties, and the presence of suspended sediments.

ML models generally performed well in freshwater environments, with the freshwater DeepLabv3+ model achieving the highest F1 score of 0.517 when tested on the Athol site, while having a validation score of 0.521. The freshwater U-Net and SegNet models also obtained F1 scores above 0.450, proving that DCNNs can generalize well in transparent freshwater environments and can be transferable. However, the freshwater Random Forest model obtained the lowest F1 score of 0.180 for Athol among the ML models, despite having the highest validation score of 0.550, and was surpassed by the physics-based model with a score of 0.211. Neglecting spatial contextual information and relying on individual pixels to learn the relationship between spectral reflectance and

water depth affected the model's ability to generalize to unseen data. This increased the model's sensitivity to noise and led to overfitting.

In contrast, saltwater environments posed significant challenges for the models. The performance of models trained in saltwater conditions was considerably compromised, especially for deeper depths, as light penetration is significantly reduced in turbid, sediment-laden waters. The saltwater SegNet model obtained the highest F1 score of 0.332 for Graham Island, but acquired a low score of 0.198 for Rimouski. U-Net maintained consistent scores for both Graham Island and Rimouski of 0.362 and 0.351, respectively, proving that skip connections allow the model to retain spatial information that might be lost during downsampling. While DeepLabv3+ performed the best during training and validation, achieving a mean F1 score of 0.454, its ability to generate accurate predictions was restricted by the inherent complexity and density of saltwater. The lack of light penetration beyond the depth of 3m led to poor performance in deeper water, particularly for the RF model, which achieved the lowest F1 score of 0.07 among all the ML models. These findings highlight the importance of incorporating spatial information in saltwater SDB modelling. Model performance in Graham Island was particularly low due to the seabed morphology. As the area has a steep sloping seabed, multiple depth classes, which return variable reflectance values, occur in a single 10m x 10m pixel. The effects of spectral mixing crippled the performance of the ML models. High-resolution images, greater than 10m, could improve water depth prediction in regions like Graham Island.

The results show that deep learning models are more successful in predicting water depth in sites they were not trained on than the machine learning Random Forest model. This suggests that the number of bands included during training is not a significant factor. The three RGB bands are sufficient for SDB modelling, provided that spatial contextual information is considered. However, the NIR band could assist in clearly identifying the shoreline as NIR is almost completely absorbed by water while it is reflected by land, therefore, this information could improve model performance in

the Land and 1m class. In addition, the RF model's validation results indicate that including IOPs could enhance the model's predictive capabilities. This implies that a physics-based machine learning model that considers water's IOPs and physical properties could perform well in high-latitude waters.

Testing model transferability revealed the generalization limitations of all models, as performance on unseen data from different environments, particularly saltwater, often differed from the initial training and validation results. This underscores the need for more diverse and representative training datasets from diverse environmental conditions. By introducing satellite imagery with variations in brightness and from more Canadian coastal zones, ML models can learn more general patterns in the data. Furthermore, the research explored the impact of water conditions on model transferability. In freshwater conditions, the models were reliable up to depths of approximately 8 meters, with DeepLabv3+ achieving an F1 score of 0.57 at 8m in the Athol test site. However, in saltwater, models were only reliable for shallow depths up to 3 meters due to poor light penetration. These findings suggest that while freshwater SDB models show promise for deeper depths, saltwater models require further refinement, especially in high-latitude regions. The depth results indicate they meet the Category C and D Zones of Confidence (ZOC) of the International Hydrographic Organization (IHO) standards.

Although the study reports varying F1 scores in different models and environments caused by the intrinsic variability in water transparency, bottom character, and sediment concentration, the results obtained are reliable due to the measures implemented. From May to October, satellite images acquired during the warmer seasons were collected for training and testing to limit the variation in IOPs. The images were visually inspected to verify they were free of clouds, haze, and intense wave activity to avoid significant variability in reflectance values. We also ensured that images captured a few years from 2019, when the latest measurements were taken, were collected to minimize the effects of significant changes in seabed morphology. Overall, this study empha-

sizes the critical role of water clarity and spatial contextual information in SDB modelling. While considerable progress has been made in predicting water depth in high-latitude regions using ML models, challenges remain in environments with complex sediment-laden waters. Future research should focus on improving the robustness of models through enhanced data collection, advanced regularization techniques, and multi-channel data integration.

In conclusion, this thesis contributes to a better understanding of how environmental factors influence SDB ML/DL model performance and highlights the need for continued research for accurate and versatile machine learning methods for satellite-derived bathymetry.

6.1 Future Work

While this research provides insights into the performance of machine learning methods for satellite-derived bathymetry (SDB) in freshwater and saltwater environments, several areas remain for further exploration and improvement. The limitations identified in this study, particularly in terms of generalization to unseen data and the challenges of accurately modelling deeper waters, provide numerous opportunities for future work.

By incorporating more data from various bathymetric environments, models could become more adaptable to unseen environments. Data augmentation techniques, such as simulating different water turbidity levels and depth ranges using physics-based models and algorithms, may help with model robustness.

Advanced regularization methods, such as dropout rates, could help prevent overfitting and improve model performance on unseen test data. This works by randomly dropping out neurons during training to force the network to rely on the remaining neurons and learn more generalizable patterns and features.

A physics-based machine learning approach that considers spatial contextual information could be investigated, where the physical properties of water and IOPs are integrated into existing image classification models. Raster image datasets with the physical properties as the pixel value could be introduced to the machine learning models.

Although this study utilized multispectral bands, integrating hyperspectral imagery, which covers a broader range of wavelengths, could enhance the model's ability to distinguish spectral signatures from different depth classes. Identifying which specific wavelengths or combinations of bands provide the most relevant information for SDB modelling, especially for saltwater, could lead to more accurate predictions. Beyond the RGB bands, incorporating the NIR band and IOPs could significantly improve model performance.

References

- Ai, B., Wen, Z., Wang, Z., Wang, R., Su, D., Li, C., and Yang, F. (2020). Convolutional neural network to retrieve water depth in marine shallow water area from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2888–2898.
- Akar, Ö. and Güngör, O. (2012). Classification of multispectral images using random forest algorithm. *Journal of Geodesy and Geoinformation*, 1(2):105–112.
- Al Najar, M., Thoumyre, G., Bergsma, E. W., Almar, R., Benshila, R., and Wilson, D. G. (2023). Satellite derived bathymetry using deep learning. *Machine Learning*, pages 1–24.
- Albert, A. (2004). *Inversion technique for optical remote sensing in shallow water*. PhD thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky.
- Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.
- Belgiu, M. and Draguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31.
- Berst, A. and Spangler, G. (1972). Lake huron: effects of exploitation, introductions, and eutrophication on the salmonid community. *Journal of the Fisheries Board of Canada*, 29(6):877–887.
- Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

- Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., and Ruescas, A. (2016). 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. In *Living Planet Symposium*, volume 740, page 54.
- Caballero, I. and Stumpf, R. P. (2019). Retrieval of nearshore bathymetry from sentinel-2a and 2b satellites in south florida coastal waters. *Estuarine, Coastal and Shelf Science*, 226:106277.
- Canadian Hydrographic Service (2022). Fisheries and oceans canada. Retrieved January 16, 2023.
- Çelik, O. İ., Büyüksalih, G., and Gaziouglu, C. (2023). Improving the accuracy of satellite-derived bathymetry using multi-layer perceptron and random forest regression methods: A case study of tavşan island. *Journal of Marine Science and Engineering*, 11(11):2090.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.
- Cheng, J. and Liang, S. (2011). Radiative transfer modeling.
- Crawford, W. R., Huggett, W. S., and Woodward, M. J. (1988). Water transport through hecate strait, british columbia. *Atmosphere-ocean*, 26(3):301–320.
- Dekker, A. G., Phinn, S. R., Anstee, J., Bissett, P., Brando, V. E., Casey, B., Fearn, P., Hedley, J., Klonowski, W., Lee, Z. P., et al. (2011). Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in australian and caribbean coastal environments. *Limnology and Oceanography: Methods*, 9(9):396–425.
- Dohler, G. (1968). Chart datum. *The International Hydrographic Review*.

- Duan, Z., Chu, S., Cheng, L., Ji, C., Li, M., and Shen, W. (2022). Satellite-derived bathymetry using landsat-8 and sentinel-2a images: Assessment of atmospheric correction algorithms and depth derivation models in shallow waters. *Optics Express*, 30(3):3238–3261.
- Duroux, R. and Scornet, E. (2018). Impact of subsampling and tree depth on random forests. *ESAIM: Probability and Statistics*, 22:96–128.
- European Space Agency (2014a). Copernicus open access hub. Retrieved January 16, 2023.
- European Space Agency (2014b). Sentinel online. Retrieved January 16, 2023.
- Gege, P. (2005). The water colour simulator wasi. *User manual for version*, 4:564–01.
- Gharahbagh, A. A., Hajihashemi, V., Machado, J. J., and Tavares, J. M. R. (2025). Land cover classification model using multispectral satellite images based on a deep learning synergistic semantic segmentation network. *Sensors*, 25(7):1988.
- Gholamalifard, M., Kutser, T., Esmaili-Sari, A., Abkar, A. A., and Naimi, B. (2013). Remotely sensed empirical modeling of bathymetry in the southeastern caspian sea. *Remote sensing*, 5(6):2746–2762.
- Hamylton, S. M., Hedley, J. D., and Beaman, R. J. (2015). Derivation of high-resolution bathymetry from multispectral satellite imagery: A comparison of empirical and optimisation methods through geographical error analysis. *Remote Sensing*, 7(12):16257–16273.
- Hedley, J. (2008). A three-dimensional radiative transfer model for shallow water environments. *Optics Express*, 16(26):21887–21902.
- Ilori, C. (2020). Physics-based satellite-derived bathymetry for nearshore coastal waters in north america.
- Ilori, C. O. and Knudby, A. (2020). An approach to minimize atmospheric correction error and improve physics-based satellite-derived bathymetry in a coastal environment. *Remote Sensing*, 12(17):2752.

- International Hydrographic Bureau (2008). Iho standards for hydrographic surveys. Technical report, Monaco: 5th Edition, Special Publication No. 44. International Hydrographic Organization.
- International Hydrographic Organization (2023). S-68 guidelines and recommendations for hydrographic offices for the allocation of catzoc/qobd values from survey data. Technical report, International Hydrographic Organization.
- Kranck, K. (1972). Tidal current control of sediment distribution in northumberland strait, maritime provinces. *Journal of sedimentary research*, 42(3):596–601.
- Kyryliuk, D. and Kratzer, S. (2019). Evaluation of sentinel-3a olci products derived using the case-2 regional coastcolour processor over the baltic sea. *Sensors*, 19(16):3609.
- Lee, W.-L. and Liou, K.-N. (2007). A coupled atmosphere–ocean radiative transfer system using the analytic four-stream approximation. *Journal of the atmospheric sciences*, 64(10):3681–3694.
- Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., and Patch, J. S. (1999). Hyperspectral remote sensing for shallow waters: 2. deriving bottom depths and water properties by optimization. *Applied optics*, 38(18):3831–3843.
- Liu, F., Zhou, Z., Jang, H., McMillan, A., and Kijowski, R. (2017). Deep convolutional auto-encoder and 3d deformable approach for tissue segmentation in magnetic resonance imaging. *International Society for Magnetic Resonance in Medicine*, 25.
- Lumban-Gaol, Y., Otori, K., and Peters, R. (2021). Satellite-derived bathymetry using convolutional neural networks and multispectral sentinel-2 images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:201–207.
- Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied optics*, 17(3):379–383.
- Lyzenga, D. R. (1985). Shallow-water bathymetry using combined lidar and passive multispectral scanner data. *International journal of remote sensing*, 6(1):115–125.

- Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. (2016). Fully convolutional neural networks for remote sensing image classification. In *2016 IEEE international geoscience and remote sensing symposium (IGARSS)*, pages 5071–5074. IEEE.
- Makarewicz, J. C. and Howell, E. T. (2012). The lake ontario nearshore study: introduction and summary. *Journal of Great Lakes Research*, 38:2–9.
- Manessa, M., Setiawan, K., Haidar, M., Supriatna, S., Pataropura, A., Supardjo, A., and Supardjo, A. (2020). Optimization of the random forest algorithm for multispectral derived bathymetry. *International Journal of Geoinformatics*, 16(3):1–6.
- Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T., and Higuchi, T. (2016). Satellite-derived bathymetry using random forest algorithm and worldview-2 imagery. *Geoplanning: Journal of Geomatics and Planning*, 3(2):117–126.
- Mascarenhas, S. and Agarwal, M. (2021). A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, volume 1, pages 96–99. IEEE.
- Mobley, C. (2022). *The oceanic optics book*. International Ocean Colour Coordinating Group (IOCCG).
- Mobley, C. D. (2001). Radiative transfer in the ocean. *Encyclopedia of ocean sciences*, 4:2321–2330.
- Olayinka, I. C. and Knudby, A. (2019). Satellite-derived bathymetry using a radiative transfer-based method: A comparison of different atmospheric correction methods. In *OCEANS 2019 MTS/IEEE SEATTLE*, pages 1–4. IEEE.
- O’Shea, K. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

- Reimann, L., Vafeidis, A. T., and Honsel, L. E. (2023). Population development as a driver of coastal risk: Current trends and future pathways. *Cambridge Prisms: Coastal Futures*, 1:e14.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.
- Ross, T.-Y. and Dollár, G. (2017). Focal loss for dense object detection. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2980–2988.
- Sayer, A. (2007). A sea surface reflectance model suitable for use with aatsr aerosol retrieval. *AOPP Memorandum University of Oxford*.
- Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and surveys*, 60:144–162.
- Sekaluvu, L., Zhang, L., and Gitau, M. (2018). Evaluation of constraints to water quality improvements in the western lake erie basin. *Journal of environmental management*, 205:85–98.
- Selamoglu, M. (2021). The effects of the ports and water transportation on the aquatic ecosystem. *Op Acc J Bio Sci & Res*, 10(1):2021.
- Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., and Dar, S. H. (2021). Satellite and scene image classification based on transfer learning and fine tuning of resnet50. *Mathematical Problems in Engineering*, 2021(1):5843816.
- Stumpf, R. P., Holderied, K., and Sinclair, M. (2003). Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnology and Oceanography*, 48(1part2):547–556.
- Talley, L. D., Pickard, G. L., Emery, W. J., and Swift, J. H. (2011). Chapter 3 - physical properties of seawater. In Talley, L. D., Pickard, G. L., Emery, W. J., and Swift, J. H., editors, *Descriptive*

- Physical Oceanography (Sixth Edition)*, pages 29–65. Academic Press, Boston, sixth edition edition.
- Traganos, D., Poursanidis, D., Aggarwal, B., Chrysoulakis, N., and Reinartz, P. (2018). Estimating satellite-derived bathymetry (sdb) with the google earth engine and sentinel-2. *Remote Sensing*, 10(6):859.
- Verhoef, W. (1998). *Theory of radiative transfer models applied in optical remote sensing of vegetation canopies*. Wageningen University and Research.
- Vincent, W. F. and Dodson, J. J. (1999). The st. lawrence river, canada-usa: the need for an ecosystem-level understanding of large rivers. *Japanese Journal of Limnology (Rikusuigaku Zasshi)*, 60(1):29–50.
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018(1):7068349.
- Webb, P. (2021). *Introduction to Oceanography*, chapter Light. Roger Williams University.
- Westley, K. (2021). Satellite-derived bathymetry for maritime archaeology: Testing its effectiveness at two ancient harbours in the eastern mediterranean. *Journal of Archaeological Science: Reports*, 38:103030.
- Wilson, B., Kurian, N. C., Singh, A., and Sethi, A. (2020). Satellite-derived bathymetry using deep convolutional neural network. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2280–2283. IEEE.
- Yunus, A. P., Jie, D., Song, X., and Avtar, R. (2019). High resolution sentinel-2 images for improved bathymetric mapping of coastal and lake environments.

Appendices

Appendix A: International Hydrographic Organization standards

The International Hydrographic Organization (IHO) sets standards for all hydrographic surveys that are used in electronic navigational charts to ensure safety and the conservation of the marine environment (International Hydrographic Bureau, 2008). As electronic navigational charts (ENC) contain many independent surveys with different levels of quality, the IHO has created a system to aid hydrographers or marine personnel in evaluating the limitations of the measurements in the ENC (International Hydrographic Bureau, 2008). This system is known as the Zones of Confidence (ZOC) system which portrays the degree of confidence hydrographers can place in a set of measurements. The various ZOC categories are summarized in Table A.1 (International Hydrographic Organization, 2023).

Table A.1. IHO ZOC Categories

ZOC Category	Depth Uncertainty	Horizontal Uncertainty
A1	0.5m + 1% depth	5.0m + 1% depth
A2	1.0m + 2% depth	20m
B	1.0m + 2% depth	50m
C	2.0m + 5% depth	500m
D	>2.0m + 5% depth	>500m
U	Unassessed	Unassessed

Source: International Hydrographic Organization. 2023. *S-68 Guidelines and Recommendations for Hydrographic Offices for the Allocation of CATZOC/QOBD values from survey data*. Monaco: International Hydrographic Organization.

These six ZOC levels depict the degree of confidence that can be placed in corresponding measurements, with category A1 or CATZOC A1 having the highest accuracy and hence confidence in measurements, and CATZOC U having the poorest accuracy (International Hydrographic Bureau, 2008). The International Hydrographic Organization (2023) has assigned SDB in CATZOC C, which has a vertical uncertainty or tolerance of 2m and horizontal uncertainty of 500m. This category describes the depth measurements as having some anomalies (International Hydrographic Organization, 2023), typical of SDB predictions. However, the International Hydrographic Bureau (2008) states that SDB can only be considered as CATZOC C if the metadata of the satellite imagery are known, and the predictions are validated against in-situ bathymetric data, if not, SDB falls under CATZOC D.

These standards provide a guideline for the accuracy expected from this investigation. This also supports the earlier claims that SDB cannot replace the traditional bathymetric survey methods but is sufficient to fill in the gaps in existing bathymetric maps and predict depth in extremely remote areas that are not frequently navigated.

Appendix B: Random Forest Hyperparameters

B.1 *Mtry* Parameter

With the CART algorithm, each decision tree is grown without pruning to its full depth by progressively asking a series of Boolean ‘True/False’ questions to split each node into two leaf nodes. This ends once each node in the tree contains a single input (Breiman, 2001). However, the RF model avoids overfitting by defining a set of parameters. One such parameter is the *Mtry*, which is the number of randomly selected features to split a node (Breiman, 2001). Randomly selecting a subset of the features with a low intercorrelation but high predictive power to split nodes prevents the model from selecting the same features for each split. Scornet (2017) mentions that this form of randomization creates a more diverse Forest and encourages the model to generalize and not fit a mapping function to the noise in the satellite image.

The variable *Mtry* is usually equal to the square root of the total number of predictors or features, which allows the RF algorithm to search through fewer predictors for the best split. As our investigation involves high-dimensional multispectral data, setting a lower *Mtry* will reduce the computational cost. For the best split, the CART algorithm uses a metric to measure the impurity or class homogeneity of a node split. The metric used is the GINI index, which is defined in Equation (1) (Akar & Gungor, 2012):

$$GINI = \sum_{j \neq i} \sum (f(C_i, T)/|T|)(f(C_j, T)/|T|) \quad (1)$$

where C_i is a randomly selected target class, T is the training set, and $(f(C_i, T)/|T|)$ is the probability that *Mtry* belongs to the class C_i (Akar & Gungor, 2012).

Through the GINI index, each possible split point is tested to find the best set of *Mtry* features to split the parent node and produce more homogeneous child nodes. Akar and Güngör (2012) emphasize that a node split is effective if the GINI index of the child node is less than the parent node. Node splitting ends once the GINI index is equal to zero, implying that all the data points in the

node belong to the same class. The GINI index also prevents overfitting by ensuring that the tree does not have too many levels as only splits that lead to an information gain are selected (Belgiu & Dragut, 2016).

B.2 *Ntree* Parameter

Another parameter set is the *Ntree* or n estimators, the number of trees grown in the Forest. The study by Scornet (2017) explained that *Ntree* controls the variance of the model as more trees create a better average of the prediction. Therefore, a value of *Ntree* that is large enough to produce precise predictions and low to reduce computational cost needs to be determined.

Scornet (2017) determines the optimal value by measuring the risk of finite and infinite Random Forests. As the accuracy of the RF increases with *Ntree*, herein referred to as M , the risk of the infinite Forest is lowest while the risk of the finite Forest is higher (Scornet, 2017). Therefore, the aim is to minimize the difference between the risk of the infinite and finite Forests by finding a large enough value of M . Scornet (2017) summarizes the bounds of this difference using Equation (2):

$$0 \leq R(m_M, n) - R(m_\infty, n) \leq \frac{8}{M} \times (\|m\|_\infty^2 + \sigma^2(1 + 4 \log n)) \quad (2)$$

where

$$\|m\|_\infty = \sup_{x \in [0,1]^d} |m(x)| < \infty \quad (3)$$

and R is the risk, m is the function mapping the inputs to the outputs, n is the size of the training dataset, σ^2 is the variance, and $\log n$ is the Gaussian noise.

However, Breiman (2001) notes that though the RF model is robust to noise, the predictions of the model are not completely free of noise. This means that the difference in accuracy between the finite and infinite Forest is not only dependent on M but on the distribution of the Gaussian noise through the models too. Scornet (2017) then concludes that the accuracy of the finite Forest can converge to that of the infinite Forest as M approaches infinity with noise being the limiting factor.

Therefore, the risk of the finite and infinite Forests converges to 0 when the optimum number of trees is defined as in Equation (4):

$$M = (\log n)^{1+\rho} \quad (4)$$

where ρ is a nonnegative number.

B.3 Kn Parameter

Finally, the Random Forest is initiated and grown to a user-defined $Ntree$ with $Mtry$ (Belgiu & Dragut, 2016). However, with just these two parameters, the iterative splitting of nodes will end once each node contains samples from the same class, creating deep trees and complex models. Therefore, as with many machine learning algorithms, early stopping is implemented to stop the iterative splitting of nodes and prevent the model from overfitting. By defining the tree depth (Kn), the node-splitting process is terminated once the maximum depth has been reached.

For a single decision tree that uses the entire training sample, its predictive capabilities increase with more levels. As the sample n increases towards infinity, Kn must also increase towards infinity to minimize the prediction error (Scornet, 2017). However, as the RF uses the bagging approach, where a random subset of the sample is used to train each tree, the prediction error is lower than that of a single tree trained on the entire sample, implying that Kn of the RF is lower than the single tree.

Scornet (2017) explains that by setting Kn , the tree construction stops before each node contains one sample point and does not affect the computational cost of the model. In a later study by Duroux and Scornet (2018), it was found that the $Ntree$ does not significantly improve the accuracy of the RF as much as tree construction does. Therefore, $Ntree$ should not be prioritized during optimization, but the parameters involved in construction, such as Kn and $Mtry$ (Duroux & Scornet, 2018).

Appendix C: Satellite Image Data

The multitemporal satellite imagery collected for each study area are given in Table C.1 and Table C.2 below.

Table C.1. Saltwater Study Areas

Site	Date of image acquisition
Moresby Island, Hecate Strait	2018-09-30 2018-10-10 2020-09-09
Prince Edward Island, Northumberland Strait	2017-09-21 2018-05-09 2019-08-02
Pugwash, Northumberland Strait	2017-09-21 2018-10-06 2019-08-02
Tatamagouche Bay, Northumberland Strait	2017-09-21 2018-10-06 2019-08-02
Graham Island, Hecate Strait (Test site)	2019-08-29
Rimouski Port, Saint Lawrence Seaway (Test site)	2023-05-29

Table C.2. Freshwater Study Areas

Site	Date of image acquisition
Wasaga Beach, Georgian Bay	2020-06-05 2020-08-14 2020-11-12
Long Point, Lake Erie	2018-05-07 2018-05-17 2018-07-06
Grand Bend, Lake Huron	2018-04-30 2018-06-29 2018-09-17
Point Clark, Lake Huron	2018-04-30 2018-06-29 2018-09-17
Wolfe Island, Lake Ontario	2021-04-08 2021-04-18 2023-04-08
Athol, Lake Ontario (Test site)	2016-09-11