Non-Cooperative Vessel Detection Using Multispectral Satellite Imaging and Machine Learning Applied Towards Onboard Satellite Dark Vessel Detection

by

Keelan James Millar Earle

Submitted in partial fulfilment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
April 2024

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

Table of Contents

List of Tables

List of Figures

Abstract

Dark vessels are a major source of economic and environmental damage, estimated to cost the global economy tens of billions (USD) a year. These vessels hide or obscure their presence on the ocean to conduct unlawful activity. Current satellite non-cooperative vessel detection methods rely on terrestrial processing of data from large synthetic aperture radar satellites, and small visible imaging satellites. Multispectral imaging could bridge gaps in the capabilities offered by these methods. Furthermore, the development of powerful onboard computers for small satellites has enabled applied onboard machine vision to automate and optimize the vessel detection process.

The study investigates whether the additional spectral bands in satellite multispectral imagery, beyond traditional RGB ones, provide detection benefits for marine vessels in the seas or oceans. Additionally, it also examines the suitability of multispectral imagery to detect vessels with object detection convolutional neural network (CNN) models and evaluating onboard processing of the models on representative single board computers. The process included the data fusion of multispectral satellite imagery and automatic identification system (AIS) data, to create a dataset which was used to train machine vision object detection models. The models were trained upon 39 permutations of spectral bands and model size and evaluated within the context of four simulated orbits and imaging constraints based on terrain elevation and coordinates.

For the trained CNN models, it was found that using multispectral imaging improved the ability to detect vessels present by up to 10% compared to using only RGB imaging. This improvement was not uniform with different spectral band permutations varying considerably. The inference time penalty for using multispectral imaging was found to be no more than 4 ms per image compared to RGB inferencing time. Despite this, the multispectral models were found to be suitable for near real-time processing when imaging constraints relevant to vessel detection were utilized.

List of Abbreviations and Symbols Used

AIS	Automatic Identification System
AUC	Area Under Curve
CFAR	Constant False-Alarm Rate
CNES	Centre National D’Etudes Spatiales
CNN	Convolutional Neural Network
CPU	Computer Processing Unit
CUDA	Compute Unified Device Architecture
EADS	European Aeronautic Defence and Space Company
ESA	European Space Agency
GEBCO	General Bathymetric Chart of the Oceans (GEBCO)
GNSS	Global Navigation Satellite System
GPU	Graphics Processing Unit
GSD	Ground Sample Distance
ISS	International Space Station
IUU	Illegal, Unreported, and Unregulated
LEO	Low Earth Orbit
MAC	Minimum Accepted Confidence
MEO	Medium Earth Orbit
ML	Machine Learning
MMSI	Maritime Mobile Service Identities
MS	Multispectral
MSI	Multi-Spectral Instrument
NASA	National Aeronautics and Space Administration
NED	North-East-Down
NIR	Near-Infrared
OBC	On Board Computer
OLI	Operational Land Imager
OTS	Off The Shelf
PAN	Panchromatic

RGB	Red, Green, and Blue
SAR	Synthetic Aperture Radar
SBC	Single Board Computer
SLAR	Side-Looking Airborne Radar
SOG	Speed Over Group
SSO	Sun Synchronous Orbit
SWIR	Short-Wave Infrared
TIFF	Tag Image File Format
USD	United States Dollar
VMS	Vessel Monitoring System
VNIR	Visible Near Infrared

Acknowledgements

Dr. Mae Seto

Thank you for offering me the opportunity to complete this thesis. I am immensely grateful for your time, patience, and generosity over the course of my time in your lab.

Dr. Clifton Johston & Dr. Rob Adamson

Thank you for taking the time to serve on my graduate committee and engaging my thesis with valuable feedback and suggestions.

ISL Lab Members

Thank you for all the assistance you provided during my thesis, whether it was helping troubleshoot my code with ideas in the lab or chatting over cards at lunch. It wouldn't have been the same experience without you.

Galaxia Mission Systems

Thank you for providing funding for my thesis and offering me the opportunity to engage with the satellite industry.

My Friends

Thank you for being there for me in between the hours of research and writing. It made dull periods enjoyable and hard times manageable.

My Family

I would have never been able to complete this thesis without your support.

Chapter 1 Introduction

Remote sensing is defined by the Canada Centre for Remote Science as "… the science (and to some extent, art) of acquiring information about the Earth's surface without being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information." [1] Remote sensing can be completed with both aerial and orbital platforms, but the decreasing cost, high endurance, and the combined long-term persistence of satellite constellations has seen a dramatic increase in the utilization of the latter for specific applications. The endurance and revisit rate of orbital platforms is particularly important for monitoring oceans, whether for natural phenomena such as algae blooms or detecting vessels at sea. The detection and monitoring of these situations is important when algae blooms can have dramatic effects on the affected waters [2], and vessels can engage in illegal or unlawful activities and inflict significant economic and environmental damage.

Within the context of illegal or unlawful activities, vessel detection has become a particularly compelling application for satellite remote sensing considering vessels may disengage, tamper, or spoof existing transponders (e.g., Automatic Identification System) to hide or mask their presence. These vessels, commonly referred to as 'dark vessels', contribute to illegal activity estimated to cost the global economy tens of billions per year [3] and are a major contributor to unsustainable fishing [4]. Currently, dark vessel detection from satellites utilizes optical – such as panchromatic or visible light – modalities or synthetic aperture radar (SAR) [5]. Both these methods have their strengths and weaknesses, particularly susceptibility to atmospheric obscurants and the size and subsequent cost of satellites required.

The progressive development of small satellite technologies has led to a widescale adoption and deployment of small inexpensive satellites as their capabilities have grown to match complex mission requirements [6]. More capable communications systems permit higher data transmission rates with less power. More capable attitude determination and control system components permit greater pointing and determination accuracy with more compact components. More precise cameras yield better imaging swaths and ground sample distance (GSD) in smaller form factors. Currently, multispectral cameras with GSD of less than 20 m at a 500 km altitude can be integrated aboard satellites as small as a 2U CubeSat [7]: a satellite measuring 20 × 10 × 10 cm.

Special attention is drawn to on-board processing power available to satellites which has seen significant improvements. Whether a result of specially designed and hardened on-board computers (OBC) or commercial-off-the-shelf (OTS) components with radiation tolerant designs, the algorithms and programs satellites can operate have become progressively more capable and complex [8]. These developments suggest possibilities to perform on-board machine learning. The first satellite with an onboard machine learning algorithm for imaging was launched in 2020 with a specialized Intel processor to detect clouds present in the images captured [9]. The satellite reduced the volume of data transmitted thereby lowering the data transmission requirements, while also reducing the delay in receiving processed information. If the same could be replicated for marine vessel detection, it could automate and optimize a large portion of the dark vessel detection process.

Given these considerations, this thesis aims to develop a vessel detection machine learning (ML) model applied to multispectral satellite imagery on emulated satellite hardware. It aims to develop a multispectral imagery vessel dataset to develop and train a model to process images with more spectral bands. Different spectral band combinations will be trained and evaluated, to assess the impact of different spectral bands on detection performance as measured by F1-score, recall and precision. Finally, the model will be trialed aboard emulated satellite hardware suitable for a CubeSat to contextualize the model inferencing speed within different orbital and operation conditions.

The outcome of the research is a contribution to the development and application of non-cooperative vessel detection. Current studies of (MS) ML satellite vessel detection are few in number, and either focus on the application of visible light or select locations in favorable weather conditions and environments. The dataset produced will be a resource to conduct further research and development of MS ML models. The completed research itself should help the process of quantifying the value of including additional spectral bands for sensing tasks, especially in computation limited circumstances of low Earth orbit.

1.1 Thesis Structure

This thesis is comprised of five chapters: Introduction, Background, Methodology, Results and Discussion, followed by Conclusions and Future Work. The remaining chapters are described as follows.

The Background, chapter 2, covers much of the same topics covered in the introduction to greater depth. Additional attention is given to the economic and environmental impact of dark vessels, remote sensing modalities, small satellites, and a more thorough discussion of ML. The section ends with a recount of the research statement based on information discussed.

Chapter 3 presents the methodology of the thesis work. This covers the steps to create a non-cooperative vessel detection method through application of ML. The process of creating a multispectral dataset is detailed here – with steps laid out in the interest of transparency and repeatability – including the processing of Sentinel-2 satellite imagery and AIS information. Then, the developed ML model is briefly discussed, along with the tested spectral band combinations and hardware configurations. Finally, a MATLAB satellite simulation is presented to offer context to the image processing rate.

Chapter 4 presents the results of the thesis research. Covered here are the model performance, the model inferencing rate aboard approximate hardware, and the MATLAB satellite simulation for inferencing rate context. The results for model performance and model inferencing rate are broken down into groupings based on data relevance to one another, and in the interest of legibility.

The conclusion, chapter 5, summarizes the main research outcomes. This includes obvious conclusions, as well as less substantiated positions that are noted for the inability to answer or draw conclusions from. The final portion of this thesis discusses both gaps in the research at a methodology level, as well as possible future research directions due to resource or scope limitations.

Chapter 2 Background

2.1 Dark Vessel and Location Transponders

The term “dark vessel” is a catch-all generally used to describe vessels which willfully disengage an equipped transponder system [10]. The term may also be used to describe vessels which manipulate their position signal via spoofing [11], do not have an equipped location transponder system, or their system is out of range of any terrestrial or orbital system. Given the range of conditions and circumstances for transponder outages, dark vessels are not inherently unlawful but instead are difficult to monitor for unlawful activity. The systems disengaged for dark vessels may be closed source systems like VMS which open only to specific bodies, but generally the term refers to the disengagement of automatic identification system (AIS).

AIS is a transceiver-based system originally implemented by the International Maritime Organization for collision avoidance [12], and the closest thing to a universal marine location system. When the regulations first passed [13], ships falling under certain criteria – primarily displacement, cargo, whether a voyage was international or domestic, and passengers – were required to integrate AIS, but this later expanded to include a wider range of applications such as fishing. An example of a full AIS message can be observed in Table 1. Some countries also require more strict use of AIS for vessels operating inside of territorial waters than those in international waters.

Table 1 AIS message example of the tugboat SEASPAN CORSAIR [14]. AIS messages include location, information about the vessel’s course, vessel type (cargo), and size.

description	**value**	**description**	**value**
MMSI	316003663	CallSign	CZ6447
BaseDateTime	2021-12-04T00:00:02	VesselType	31
LAT	48.87658	Status	0
LON	-123.635	Length	25
SOG	2.3	Width	7
COG	124.9	Draft	3.9

description	value	description	value
Heading	124	Cargo	52
VesselName	SEASPAN CORSAIR	TransceiverClass	A
IMO	IMO74347		

However, as a method to monitor unlawful or illegal behavior, AIS has its drawbacks. Critically, the system requires vessel participation and can be willfully disengaged [12]. Disengagements are permitted under conditions of security and safety, or simply because the vessel is out-of-range of any receiver. However, with the advent of satellite-based AIS-receivers lapses in monitored position have become rarer. Another manner the system has been exploited is through the spoofing of AIS positions. In observed cases, a ship broadcasts its position to be tens, if not hundreds, of kilometers from its actual position [15]. Either obscured or obfuscated movement can hide malicious actions from authorities, which is discussed next in Section 2.2.

2.2 Sustainability and Economic Impact of Dark Vessels

The oceans, seas and waterways of Earth are host to an abundance of resources and serve as active trade routes. It is estimated by the United Nations Conference on Trade and Development that 80% of the global trade by volume, and 70% by value – constituting 10.7 billion tons by mass – was conducted by trade at sea by a commercial fleet of 94,171 vessels in 2017 [16]. With regards to fisheries, in 2020 78.8 million tonnes of live weight were caught on the ocean by 4.1 million fishing vessels [17]. However, complicating the use of the marine domain is illegal or unlawful activity. This activity may be conducted in plain view, but the activity may be hidden or obfuscated. Examples may include illegal, unreported, and unregulated (IUU) fishing, smuggling, forced labor [18], and piracy.

Of these activities, IUU is among the most documented and has strong connections to dark vessels. Knowledge of the extent of IUU fishing is limited given its nature, but estimates believe it to account for a notable portion of the world's catch. For instance, one paper estimated IUU fishing accounts for 20% of the world's catch, reaching up to

50% in certain areas [19]. By comparison the Food and Agriculture Organization of the United Nations estimate that over 15% of the annual global catch is IUU [20]. Figure 1 shows a map of the fraction of fishing activity obscured by suspected disabling of AIS, which is to say: vessels suspected of willfully disengaging their systems to fish illegally. As a more focussed regional IUU example attributable to dark vessels, it is estimated between 2017 and 2018, $440M USD worth of squid (more than 164,000 tonnes) was caught by dark vessels off the coast of North Korea [10]. Compounding the problem is that many areas of the world's fisheries are exploited at an unsustainable level, placing pressure on legal and regulated fishing as well as the long-term economic future of these industries.

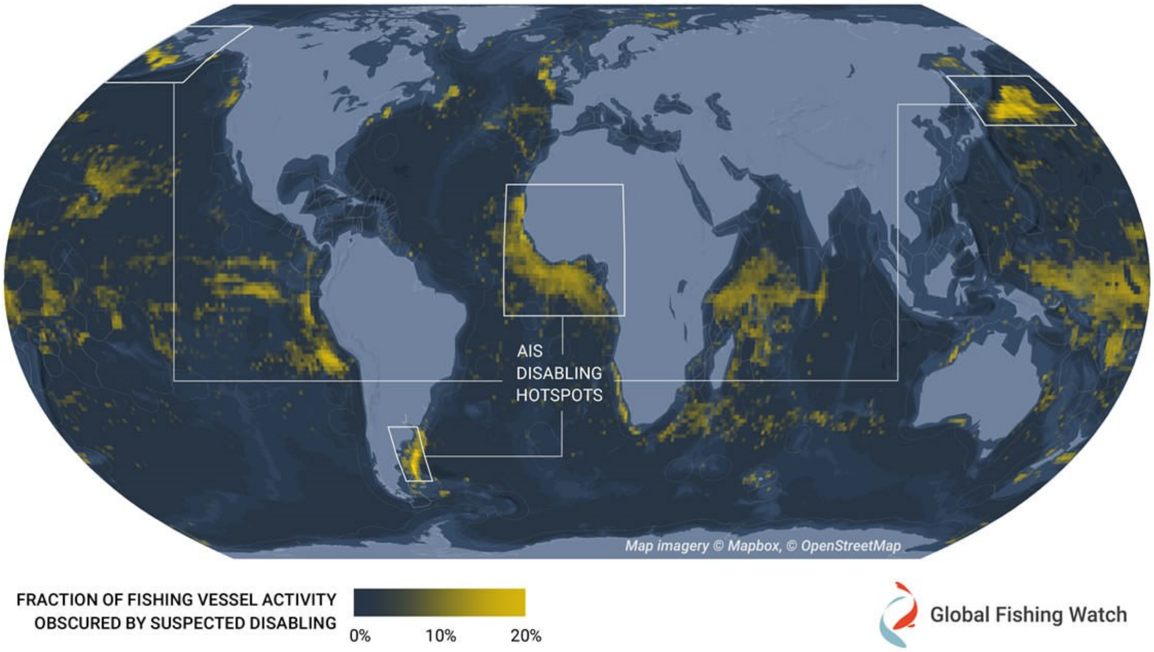


Figure 1 Regions of fishing vessel activity obscured by suspected disabling of AIS [21] highlighting the widespread use of dark vessels in fishing.

Complicating the discussion on IUU fishing is that tracking the source of catches is difficult due to transshipment: the unloading of goods from one vessel to another while at sea. This can obfuscate and make it difficult to trace the source of catches. A 2018 study examined 32 billion automatic identification system (AIS) messages for ocean vessels from 2012 to the end of 2017 and tracked 694 cargo vessels capable of transshipping fish.

According to their criteria, the research found 46,570 instances where these vessels exhibited signs of potential transshipment (Figure 2) [20]. Ships engaging in transshipment may use flags of convenience to skirt fishing laws, but also may rendezvous with dark vessels at sea.

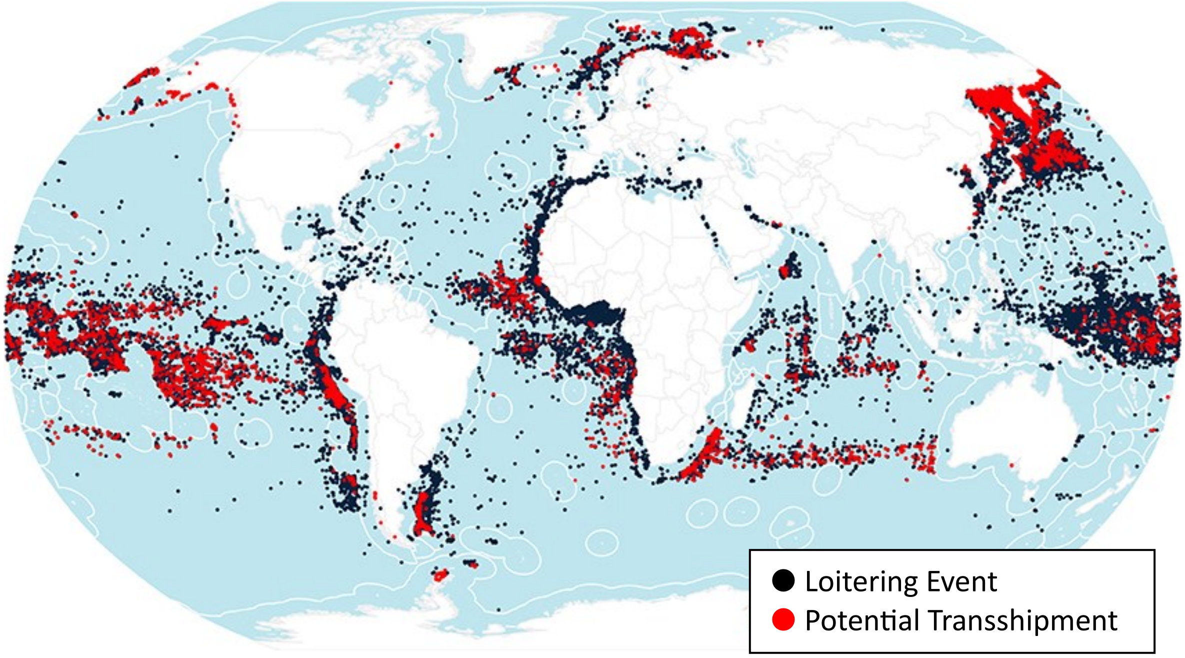


Figure 2 Global patterns illustrating encounters with other vessels(red) and loitering events where transhipment may occur (black) [20]. Loitering events may be non-dark vessels meeting with dark vessels to transship cargo.

Unlawful operation and evasion of sanctions is another area of exploitation for dark vessels. Disengagement of tracking systems – commonly AIS – is a tactic to avoid sanctions while operating despite sanctions or other activity. While not strictly falling under the category of dark vessels, it is an increasingly common tactic to spoof AIS data – particularly location manipulation – to make a ship appear elsewhere from its actual position [15, 22]. Verifying this information can be challenging under time-sensitive conditions. Oil is a common commodity for this nature of subterfuge, but may also include dredging of sand, or smuggling of other commodities [11].

With the limitations of land and air-based solutions, the continued development of satellite remote sensing methods are critical to monitoring and identifying dark vessels.

While some dark vessels may be benign, or a result of unintentional loss of signal, the inability to identify dark vessels precludes any monitoring. The detection methods available for satellite-based methods are discussed in section 2.3, broken up into cooperative (section 2.3.1) and non-cooperative (section 2.3.2) methods. The latter is further broken down and detailed by sensing modalities for SAR, optical, and multispectral detection (sections 2.3.3 to 2.3.5).

2.3 Existing Detection Methods

While methods to detect dark vessels vary, they can be broadly divided into two categories: cooperative, and non-cooperative. As the term suggests, cooperative sensing relies on the target vessel's participation to monitor it – such as using an active AIS signal – while non-cooperative methods can detect the vessel regardless of its willingness to participate. Most countries deploy a combination of these methods to secure and enforce their sovereign waters.

As a developed country with large coastal regions, Canada is an example of a country with both the reason and means to monitor their coastlines. The methods used by Canada to detect vessels include AIS tracking – either terrestrial or satellite based – vessel monitor systems (VMS), patrols, aerial surveillance, radar satellites and violation records. Of course, there are limitations with each of these methods. AIS signals can be masked, are not required for vessels of all displacements and the signal can (un)naturally drop out. VMS systems are expensive, with requirements that vary by region, and can also be masked. Aerial surveillance is limited by the operator and/or craft endurance and cloud cover. Radar satellites are expensive to design and launch which limits the number that can be launched and thereby limits the revisit rate of a constellation [23].

2.3.1 Cooperative Detection Methods

Cooperative detection methods for dark vessels can seem like a contradiction at first. To a degree this contradiction holds as cooperative detection methods are unable to detect dark vessels at a specific instant. However, dark vessel operation is rarely so binary over the course of a voyage and patterns appear over their voyage [24]. For instance, a foreign fishing vessel that approaches protected waters, then disengages their AIS to operate within the restricted area, only to re-engage it several hours later travelling away from territorial waters produces telling data patterns. Repeated patterns of this behaviour or comparing voyage patterns against voyages of other violating vessels can be used to identify high risk vessels and cue other detection methods such as patrols, or prompt investigations. It is also through this type of analysis that hot spots for dark vessel fishing can be determined [25, 21].

2.3.2 Non-Cooperative Detection Methods

For the detection of dark vessels at a given instant, non-cooperative detection methods are the only measure. These methods can range from the most rudimentary (e.g. patrols for visible surface sightings) and to the complex (ML applied to satellite imagery). Within the context of satellite remote sensing, however, ship detection is generally achieved with either SAR or optical sensors – generally either panchromatic or RGB. Other sensing modalities like hyperspectral, thermal imaging, and reflectometry systems, such as reflected SAR or GNSS reflectometry, have seen limited use for dark ship detection with satellites due to the low GSD or spatial resolution – the minimum size of a feature that can be fully occupy a pixel [5, 26, 27]. The only other sensing modality worth drawing attention to is multispectral imaging, which generally utilizes RGB spectral bands in addition to several NIR or SWIR spectral bands. Multispectral observations have had attention in recent years for vessel detection given improved spatial resolution and subsequently greater spectral information density per pixel [5, 7].

Figure 3 compares the average spatial resolution of multispectral and panchromatic satellites achieved over time.

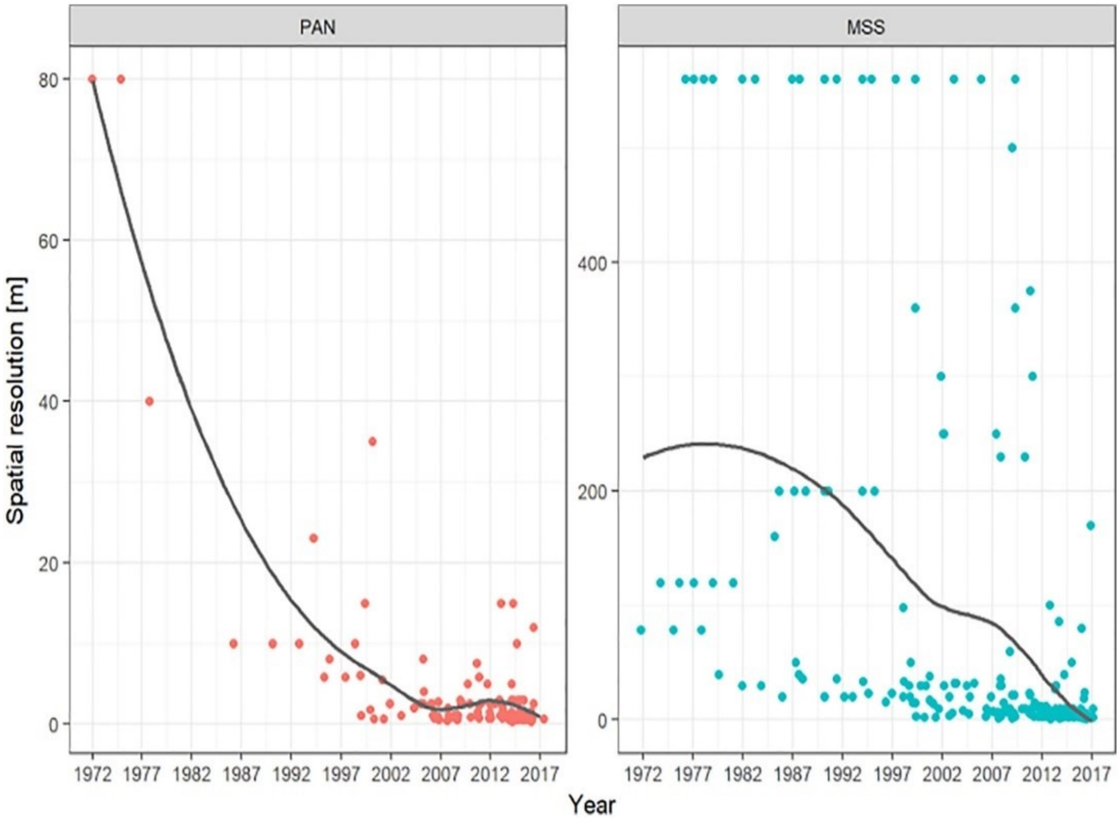


Figure 3 Increase in spatial resolution for panchromatic (PAN) and multispectral (MSS) satellite imaging over time. Higher spatial resolution means more information is contained in an image [5].

2.3.3 Synthetic Aperture Radar Vessel Detection

Among the different modalities for vessel detection within satellite remote sensing, SAR is the current leading technology [23] [5] [28] [29]. Unlike true aperture radars, such as side-looking airborne radar (SLAR) which requires a long physical antenna, SAR uses the motion of the platform (satellite) to synthesize its antenna. In remote detection, SAR surpasses other radar types in its resolution. Of the sensor modalities discussed in this section, it is also the only sensing modality unaffected by cloud cover and absence of external illumination, a factor which has seen their widespread use. Figure 4 is an example of SAR imagery taken by the Sentiel-1C Satellite. In the image, a variety of vessels can be observed as white objects against a black background.

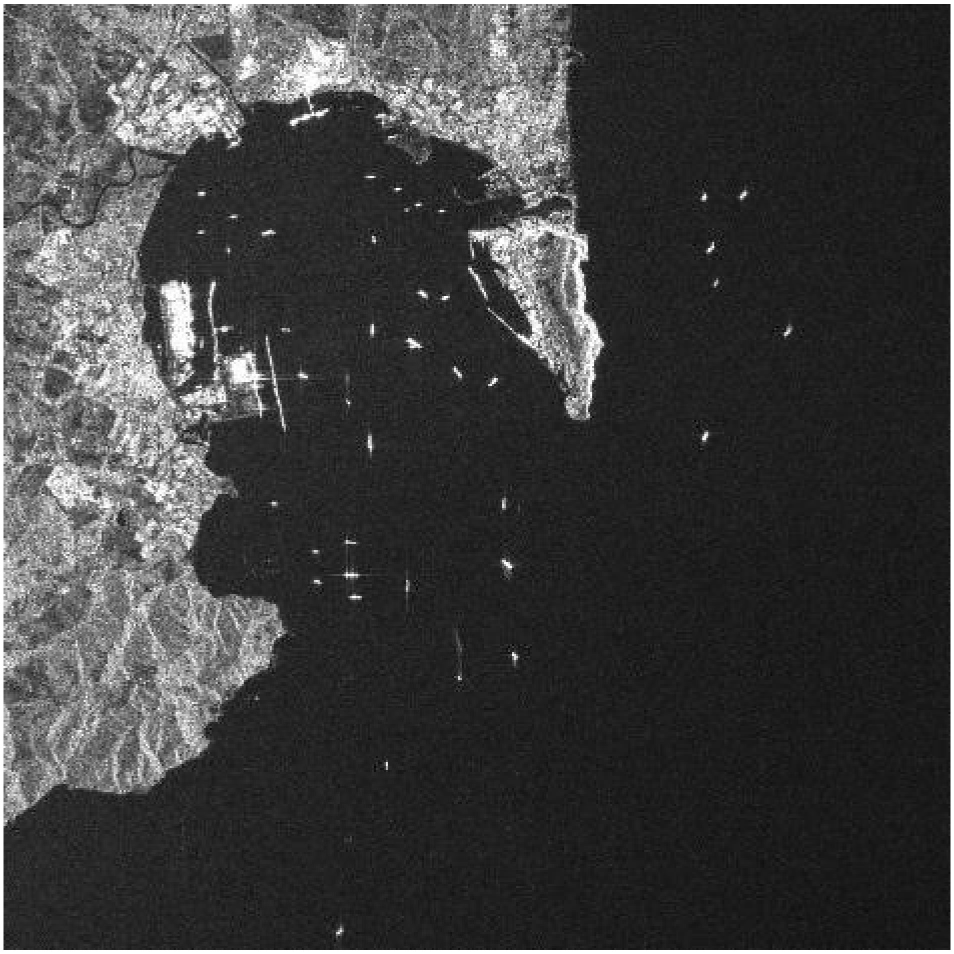


Figure 4 An example of marine vessels imaged by SAR observed between Gibraltar and Algeciras in September 2017, Copernicus Sentinel Data [30]. Marine vessels in SAR images appear as bright objects against the dark background of the water assuming sufficient imaging capabilities.

Traditionally, ships are identified from such SAR images with techniques such as constant false-alarm rate (CFAR) which searches for “bright” objects (vessels) against a “dark background”, or generalized likelihood ratio tests which measures the probability of two hypotheses (there is a target/change, and there is no target) from captured data against a minimum threshold, and uses artificially designed features for identification [31] [32]. The recent application of ML (convolutional neural networks, CNN’s) to SAR imagery, along with the improved spatial resolution of satellites, has increased the speed and accuracy of object identification and classification [31] [33] [34].

The primary challenge with SAR is its high-power draw demand of the platform. As an active sensor, SAR emits energy and observes the Earth through the energy’s returns, as opposed to passive sensors which observe reflected energy the sensor did not generate. The first SAR satellite below 100 kg mass was launched by ICEYE Ltd. in 2018 with a peak radiated power of 3.2 kW [35] for non-continuous use. Another example is the development and design of a Japanese 100 kg mass X-Band SAR satellite [36]. This design was well below the three hundred to few thousand kilograms for traditional SAR satellite but was estimated to still have an active sensor power draw of 1.3 kW. The design and launch of these massive and power-hungry satellites consequently drive their costs far above passive sensor satellites.

2.3.4 Optical Satellite Vessel Detection

For the purposes of subsequent sections, optical sensors will refer to panchromatic and red-green-blue (RGB) visible light imaging sensors. Multispectral sensors – which are often included under the classification of optical sensors in plenty of situations – will be discussed separately in the following section (2.3.5). Optical sensors generally possess the highest spatial resolution of all passive sensing modalities in use on satellites today [5]. Their high spatial resolution cannot be understated, as very-high resolution optical images can not only detect vessels, but even identify the class of vessel with a reasonable level of accuracy [26]. As a passive sensor, optical sensors enjoy advantages of low-power draw and easy integration aboard even CubeSat satellites, substantially decreasing deployment cost [7].

While optical images have been used to detect vessels, the application of ML applied terrestrially is still a relatively recent development. ML methods such as CNNs surpass classical computer vision algorithms in performance. The memory limitations of common neural networks, along with the absence of publicly available annotated datasets, have delayed development of these more advanced methods. A paper by Dr. Sergey Voinov demonstrates the capability of CNNs applied to vessel detection from optical images [26]. Figure 5 shows the results of his work for medium resolution imagery from Landsat-8, where 62% of 30 m (or larger) vessels were detected with 15 m GSD images while excluding coastal areas, considered suitable as a tool for sea traffic monitoring by the author.

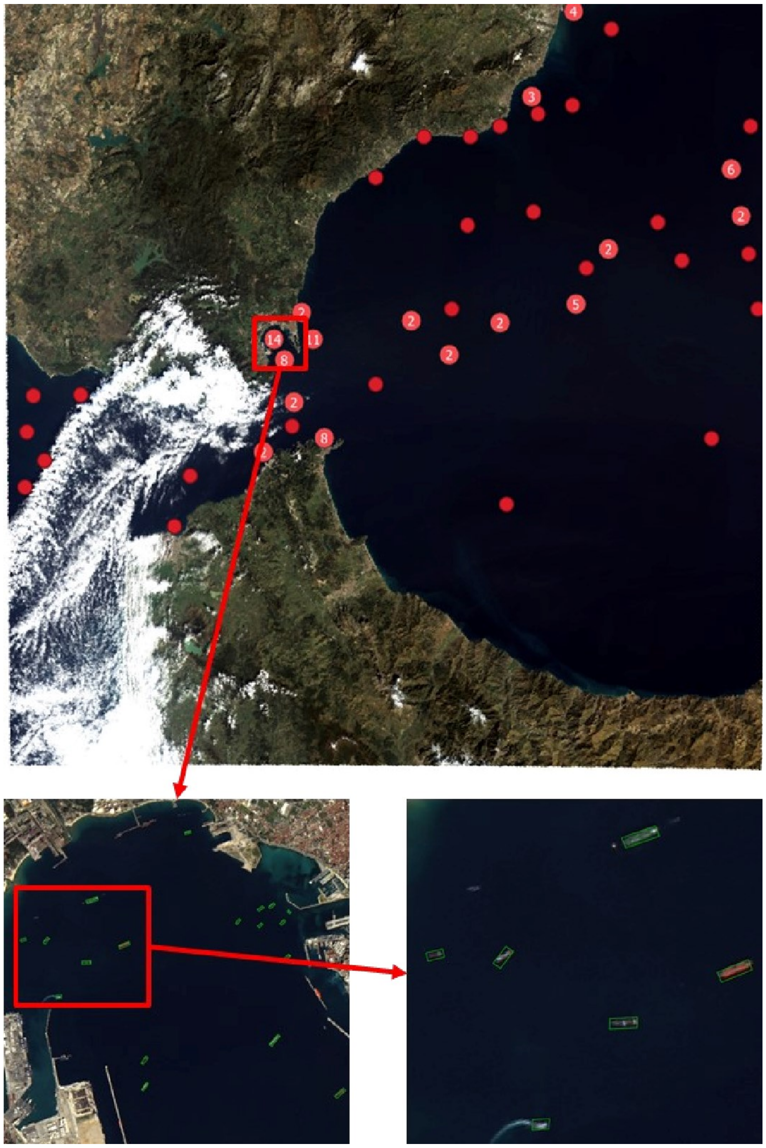


Figure 5 Example of medium resolution (Landsat-8) optical image ML vessel detection results [26]. 62% of vessels 30m or longer were detected using 15 m GSD imagery using the RGB trained model. This was considered suitable for suitable as a tool for sea traffic monitoring.

The drawbacks of optical sensors stem from their interactions with atmospheric obscurants and them being passive sensors. Atmospheric obscurants range from minor to complete reductions in the necessary transparency to observe the ground/sea level. These primarily include clouds, but also fog, smoke, smog or simply absence of light [37]. As they are passive, optical sensors can only observe reflected or emitted target illumination, limiting application in very low light or nighttime applications. Most optical detections at night are achieved with optical systems that integrate medium and long-wave infrared wavelengths [38, 5].

2.3.5 Multispectral Satellite Vessel Detection

A multispectral sensor is an extension of an optical senor. MS sensors capture additional spectral bands during the imaging period, beyond the often-included RGB colour bands, generally ranging from ultraviolet light to short wave infrared light. Each additional spectral band contains more information about what is captured in the image [5]. For instance, a red-edge spectral band – often described as a vegetation band due to chlorophyl reflecting red-edge light – can be included to monitor plant health [39] or better detect algae blooms [40]. Other multispectral spectral bands can be applied to better characterize clouds through bands sensitive to aerosols, water vapour, and clouds to perform atmospheric corrections. Some multispectral bands may also achieve better transmission through very thin and translucent cloud layers than optical bands [41]. Applied to vessel detection, a combination of RGB and NIR measurements better suppresses clutter interference from clouds or large sea waves and improves the ability to correctly identify vessels [27].

Vessel detection using multispectral modalities is a relatively recent development, despite earlier attempts [5]. In the past, the main dissuading factor for their disuse in satellite vessel detection was their lower spatial resolution. However, multispectral cameras have improved and trade-offs between swath width and spatial resolution in satellite sensors have reduced. This means shorter revisit times as wider swaths overlap more frequently while retaining feature resolution, or improvements to feature resolution under same revisit times and swath widths [42]. Compared to RGB imaging, MS imaging better captures vessel features and distinguishes them from natural or artificial ones and removes the requirement to store coastline vector maps [43]. This means the system could operate in near real-time onboard a satellite system if adopted.

2.4 Small Satellites

The ideal satellite-based sensing modality for vessel detection, without considerations for size, cost, and power draw, would be SAR. Its ability to accurately image the Earth's surface through dense cloud cover is unmatched by other modalities. But, as with any engineering problem, other requirements must be considered, and mass is one of the largest cost-driving factors for satellites. Launching heavier-than-air masses into orbit takes a notable amount of fuel and design complexity, which results in increased cost. More power draw from power-hungry systems results in a more complex power management system, and again, increased cost. Increased cost itself can result in further cost increases as the definition of what is considered an acceptable risk narrows as the replacement cost increases. In this manner, to avoid loss of a more expensive satellite, more development time is required which further drives up the cost. A combination of these factors, and the development of small satellite technologies, have truly driven the widespread adoption of small satellites. What constitutes a small satellite is explained next (section 2.4.1), followed by a discussion on how small satellites have increasingly dominated operational satellites by quantity (section 2.4.2).

2.4.1 Satellites Size Classification

Due to the cost of launching a mass into orbit, satellite size is generally classified by their mass, as opposed to their dimensions, i.e., small satellites have lower mass than larger ones. Yet, despite decades of satellite launches, there is still no single classification metric which every organization abides by. For very large (e.g., 6000+ kg James Webb Satellite) or small (e.g., 1 kg 1U CubeSat) satellites there is general agreement. The disagreement exists for what lies in between. The first satellite classification scheme was proposed by Sweeting in 1991 [44] (summarized in Table 2).

Table 2 Sweeting satellite size classification [44].

satellite class	mass
large	>1000 kg
small	500 – 1000 kg
mini	100 – 500 kg
micro	10 – 100 kg
nano	<10 kg

Two more classes were later added to Sweeting’s original classification, pico- and femto-, recognizing advancements in technology which facilitate smaller classes. Additionally, the mass classifications were changed to keep to the logic of mass increases in magnitudes.

Table 3 Revised Sweeting satellite size classification [44] to include two additional classes (pico and femto) and to define mass sizes by orders-of-magnitudes.

satellite class	**mass**
large	>1000 kg
mini	100 – 1000 kg
micro	10 – 100 kg
nano	1 – 10 kg
pico	0.1 – 1 kg
femto	1 – 100 g

While the original or updated mass sizing schemes are the most common mechanism for discussing satellite size, there is no universal size classification. For instance the ESA, European Aeronautic Defence and Space Company (EADS), Centre National D'Etudes Spatiales (CNES) [45], and NASA Small Spacecraft Technology Program [46] all use different criteria for satellite size classification either based on previous criteria or launch and deployment mechanisms. Most of these organizations use a mass criteria of less than 200 kg for the purpose of small satellite classification. For this thesis, and discussion of small satellites, the updated classification of Table 3 will be referenced for specific satellite size classification, with preference given to a source's classification if one is offered.

2.4.2 Small Satellite Development

Small satellites have a long history in space programs, and by modern classification, even out-dating larger satellites. Sputnik, the first artificial satellite launched, weighed 83 kg and would be considered a microsatellite by modern standards. Similarly, most satellites through the early history of space exploration would be considered nano- or micro-satellites. However, this was driven by necessity and not choice [44]. Early launch vehicles could not launch larger satellites, with the early micro-satellites generally being the primary and only satellite payload. Figure 6 shows all satellites launched from 1957, to the end of 2023. This figure was built from the General Catalog of Artificial Space Objects [47].

From this catalog, all standard payloads that were 0 kg mass, had humans aboard, were pressurized without crew, or were suborbital and removed. Starlink and OneWeb satellites were separated to account for their disproportionately large numbers which would bias the masses and numbers launched in recent years. The only object deliberately removed from the remaining satellites was the Buran OK-1K spacecraft which had a mass of 79,400 kg – four times that of the next highest mass and built with a crew module. From this figure, it is apparent that average satellite mass has plummeted since the high of 2005, only increasing with the inclusion of recent Starlink and OneWeb Satellites.

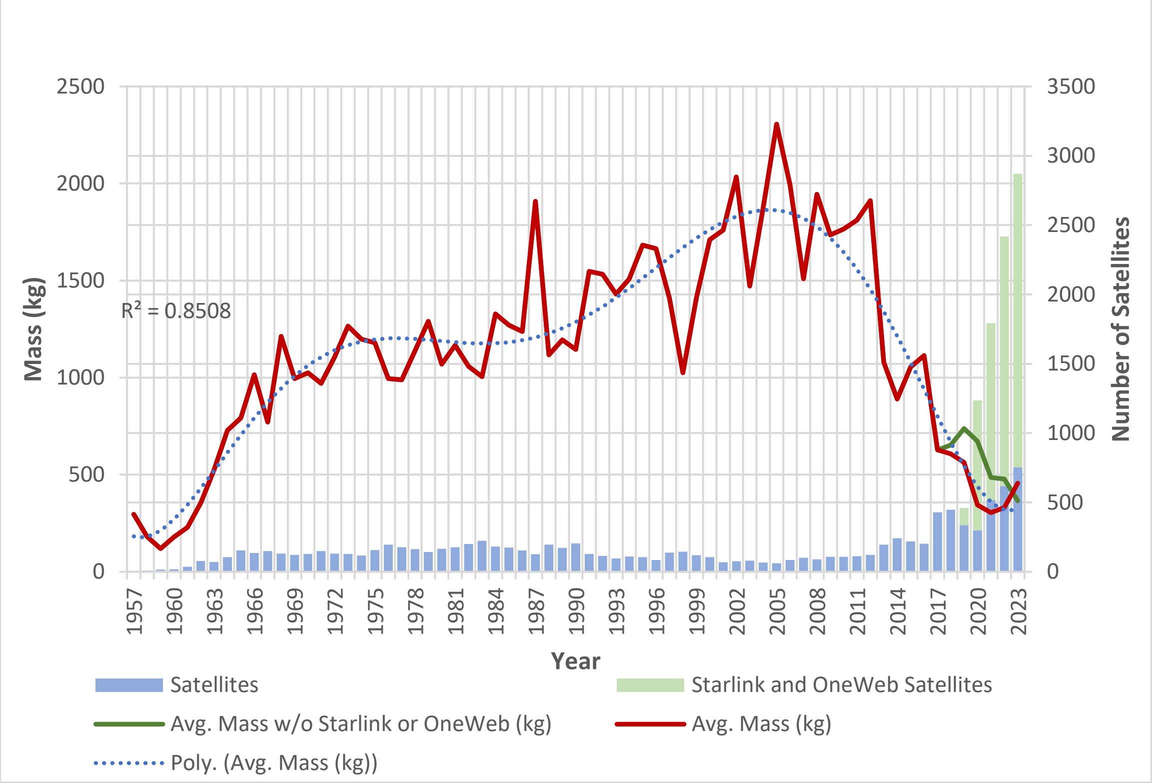


Figure 6 Average satellite mass vs number of satellite launches from 1957 to 2023 based on [47]. Average satellite mass has dropped since the peak in 2005 and has trended downwards since due to widespread use of small satellites.

Applying the constraint of less than 600 kg for what is considered a small satellite – a 2018 Federal Aviation Administration designation – from a period of 2012 to 2021 of the

8,053 spacecraft launched, 87% of launched satellites were small satellites, representing 25% percent of launch mass [48]. A disclaimer to these numbers would be that during this period, 4,163 of the 5,730 commercially launched small satellites were Starlink and OneWeb communication satellites. Additionally, 88% of commercial small satellites launched 2012 - 2021 belonged to SpaceX, Planet, OneWeb, Swarm Technologies and Spire Global, in order of descending market share of total small satellites launched.

The advantages of small satellites compared to their larger counterparts are varied. Small satellites can be launched faster and more economically into orbit than large satellites. As well, they can be in lower orbits (i.e., to achieve better resolution imagery) than larger satellites. They also serve to complement larger satellites constellations through their resiliency or repair of damaged constellations [49]. Small satellites address the satellite remote sensing challenge of coverage rate which is increased for small satellites [50]. They also offer a lower barrier-to-entry and are thus more accessible for scientific and commercial applications and end-user due to the industry's standardized or OTS components and subsequent shorter development times [49] [44]. The following two sections will focus on the temporal resolution (2.4.3) and cost (2.4.4), as these are relevant to the vessel detection within the scope of multispectral imagers.

2.4.3 Revisit Time

Initially, the deployment of a single more capable satellite may be advantageous over a constellation of multiple lesser capable ones. The higher capability of a larger satellite is in its larger imaging swaths, more robust communication systems, higher spatial resolution, or simply the sensing modality that is possible (e.g., SAR). However, it is not a clear decision unless the task demands the capabilities of the individually more capable satellite. To monitor an Earth area for certain features, the value of a constellation of multiple satellites may be significant due to the time it takes for a satellite to revisit an area [44].

For certain applications, such as oil spills or ship control, the revisit time (also known as temporal resolution) is a major challenge and barrier for satellite remote sensing [50]. Revisit time or temporal resolution refers to the time it takes for the satellites in LEO and MEO to revisit a particular position or small area. Only geostationary orbits (GEO) can achieve near-indefinite persistent observation over a position on Earth and therefore a revisit time of the camera's capture rate. But just like decreasing the GSD of a satellite imaging modality only improves a satellite's capabilities within a task to a point, the value of decreasing the revisit time varies for a task. Monitoring geographic features, such as coastal erosion is feasible with a satellite with low temporal resolution as these are slow processes relative to the revisit times of a satellite. However, monitoring of weather or active disasters such as forest fires, or tracking vessels demand much higher temporal resolution [51]. Figure 7 shows the variation in spatial and temporal observation requirements for different types of coastal studies.

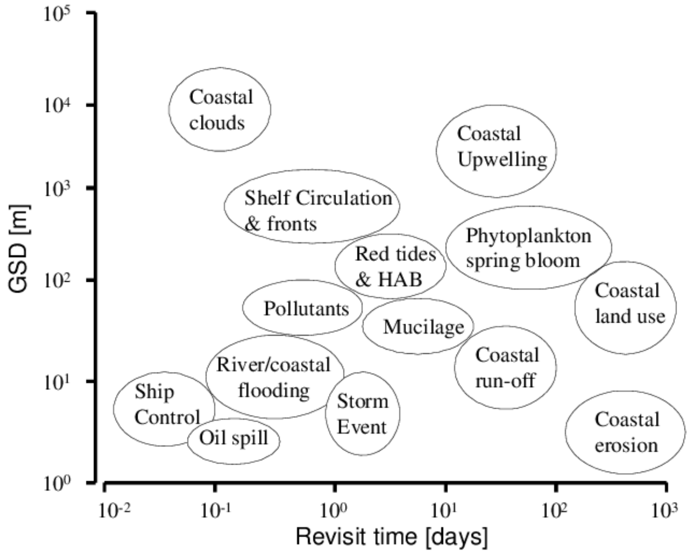


Figure 7 Spatial (GSD) and temporal (revisit time) requirements for coastal studies [51]. By comparison, tracking ship movement with satellite observations has a much shorter required revisit time, and smaller GSD than other areas of study, requiring either very capable imaging sensors with high swath and low GSD, or larger constellations of satellites.

Ideally, a satellite's revisit time is small compared to the characteristic temporal scales of the observed event, and its spatial resolution is high (i.e., low GSD) compared to the physical size of the observed feature. There are trade-offs to consider between the sensor and orbital characteristics [52]. Spatial resolution can compete with revisit time, whether it is the result of the selected satellite orbit or the sensor's swath width (a measure of resolution). One orbit type might revisit specific locations frequently, while another orbit type might rarely, if at all, revisit that same location [53]. For vessel detection, persistence is critical for timely detection, and long satellite revisit intervals are a limiting factor in maritime domain awareness [5]. Regardless of the satellite's temporal resolution or its sensor's swath width, there is a physical limit to the percentage of the planet a satellite system could cover. If the Earth was modelled as a sphere, the observable portion can be described as a spherical cap. Eq. (1) describes the percentage of a sphere or radius r, visible from a point an altitude H from its surface:

$$R_{\%} = 0.5 \times \left(1 - \frac{r}{r + H}\right) \times 100\% \tag{1}$$

such that $R_{\%}$ is the percentage of the sphere which is visible. While the Earth is an oblate spheroid, and not a sphere, due to its rotation, Eq. (1) is still relevant. Solving Eq. (1) for the Earth radius of 6371 km, at altitudes of 400 - 1000 km (LEO), shows the percentage of Earth visible scales nearly linearly with satellite altitude through this region (Figure 8).

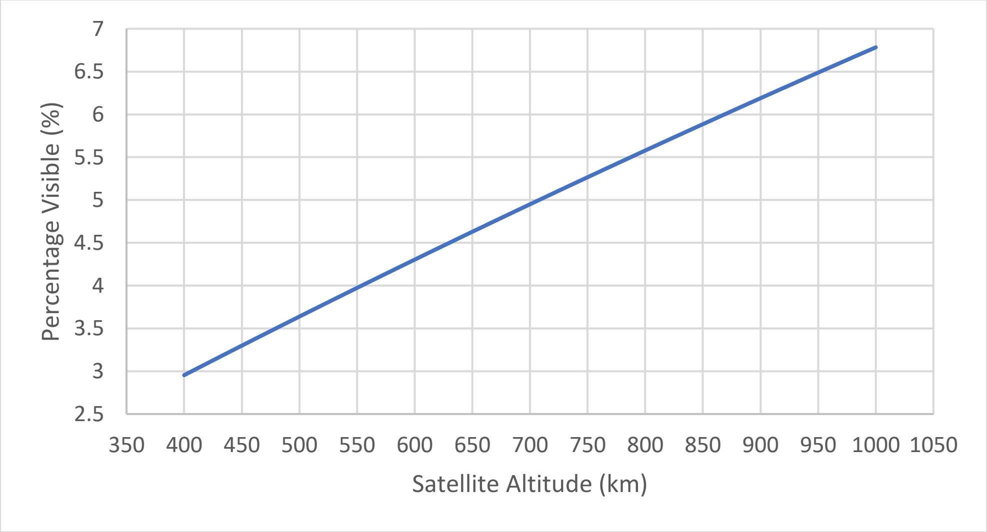


Figure 8 Maximum percentage of Earth sphere observable from orbit. Between 400 km - 1000 km satellite constellations are required to view Earth continuously with idealized imaging sensors.

Between 400 km - 1000 km, the percentage of Earth visible varies from 2.95% to 6.78% for a "perfect" imaging system. No matter the GSD, imaging swath, or satellite sensing modality, a single satellite alone cannot provide simultaneous complete coverage of the Earth, since they operate on line-of-sight imaging and communications. However, reduction of capabilities comes with a reduction in cost, and provided the capabilities of a smaller satellite are sufficient, deploying more satellites within a constellation improves the temporal resolution for the tracking or monitoring ship movement [52].

For marine vessel detection, the capabilities of optical and multispectral imaging systems compared to SAR satellites are quite different. While a direct comparison of the two obviously favours SAR, for a given budget, a single SAR satellite is not competitive against a single optical or MS satellite due to launch mass alone. Instead, the likely scenario is a single SAR satellite would be competitive against a constellation of optical or MS satellites. Temporal resolution is exchanged for the capability to operate through heavy atmospheric obscurants. To that end, the solution to ship detection lies not in a single modality but a fusion of modalities, or until the weaknesses of SAR sensors can be displaced by the strength of optical or multispectral sensors.

2.4.4 Satellite Cost

There are two defining cost considerations for a satellite. The first is the design and construction (assembly) of the satellite system, while the second is its launch. The design and assembly cost varies with satellite mission, though mass still is a driving factor due to its impact on the attitude determination and control system, along with the inclusion of reaction mass and thruster systems for altitude or orbital adjustments. However, this is harder to quantify given the vast array of satellite objectives, and the lack of publicly available component pricing and tools to estimate cost.

The launch cost of a satellite, however, is a more readily known and constant factor. Additionally, launch costs are the greatest limiting factor restricting expansion of missions in both number and scope. Consequentially, the low launch costs of small satellites are one of their most attractive features [54]. As mentioned earlier, more than any physical dimension, the satellite mass is the main cost to launch a satellite. Idealized launch costs – where all cost increasing factors are disregarded and payload for a launch vehicle is maximized on an at-cost basis – are still in the tens of thousands of dollars (USD) per kilogram. Though the cost of launch has improved with time (Figure 9), it is still significant. In fact, with continued technological advancements in satellites systems, components, and manufacturing, such as the availability of relevant OTS components, the launch costs are increasingly significant compared to the total mission budget for small satellites [55].

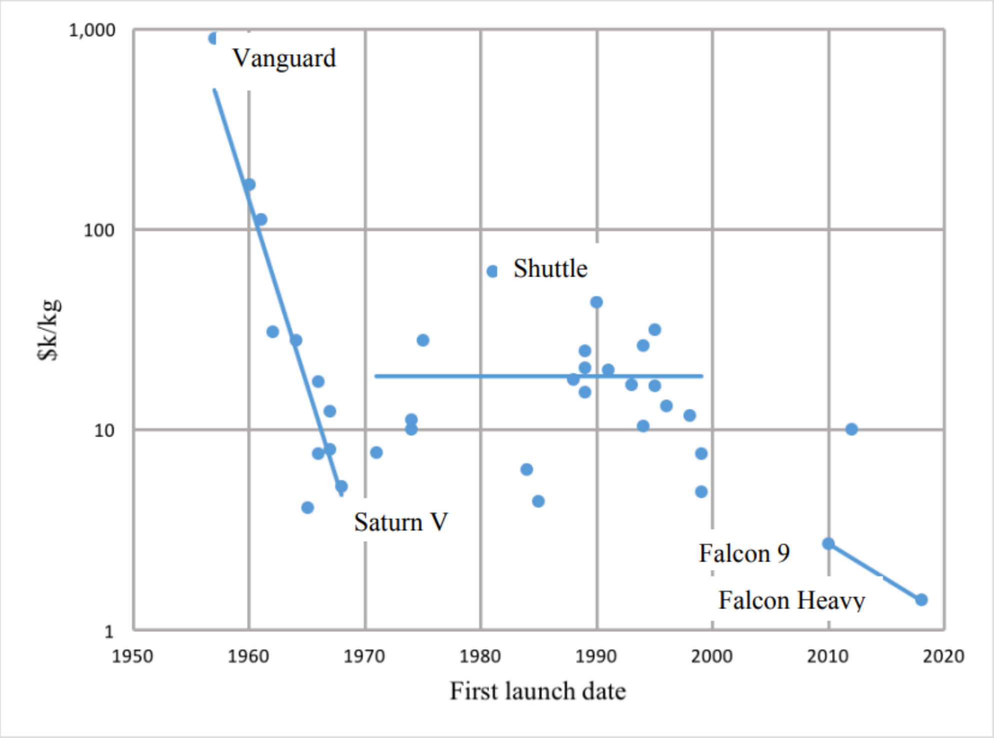


Figure 9 Launch cost (2018 USD) per kilogram to low Earth orbit by first launch date [54]. The launch cost has continually fallen since early launch vehicles and thus improves accessibility to space.

As mentioned, the launch costs in Figure 9 are idealized. The launch costs assume the entire payload capacity is utilized and compared against the total cost of the flight which is often not the case. Limited dimensions for payload, mass and space occupied by items such as deployment systems, or simply the payload being smaller than the launch systems limits can significantly increase prices [54]. While there is only so much that can be done for the first two reasons, a payload that is “too small” is an opportunity to carry a secondary payload like a small satellite. To this end, there are three primary methods to launch small satellites: dedicated launches, rideshare and cluster launches, and piggyback launches selected based on the mission profile, all of which seek to minimize the cost by maximizing launch system payload within the scope of the mission profile [55].

The methods, combined with the lower launch cost per kilogram, results in a lower cost for deploying small satellites like a CubeSat into orbit. Launch providers and deployers such as SpaceX, Rocket Lab, and NanoRacks offer a cost of deploying a small satellite into orbit from ranging from approximately $90,000 USD for a 1U up to $300,000 USD for a 50 kg satellite placed into low sun synchronous orbit (SSO) [56]. When combined with the hardware cost of a 3U satellite [57] (varying with mission objectives and components) a 3U satellite can be assembled and deployed into orbit for under $1M USD. This is extremely low compared to the cost of a modern large satellite constellation which can cost $1B USD [58], such as Canada's RADARSAT constellation that cost more than $1B CAD to develop and deploy 3 satellites [59]. As such, a very large 3U CubeSat constellation could be deployed to achieve significantly shorter revisit times [60] for less.

2.5 Machine Learning

Machine learning "refers to a broad range of algorithms that perform intelligent predictions based on a dataset" [61]. The algorithm produced for the application is referred to as a model, into which the captured data (images or otherwise) is input to return the desired prediction. The output for image-based models can vary from an image level classification ("does this image contain a vessel?") to pixel level segmentation ("what pixels of an image are part of a vessel?"), or a more specific result as deemed by the model designer.

The first machine learning subsection examines some of the roles demonstrated by onboard satellite machine learning, in additional to potential applications (2.5.1). Given the suitability of ML models to process multispectral imaging (2.5.2), the structure of convolutional neural networks (CNNs) – high performance ML models suited for MS imagery – are discussed (2.5.3). Finally, data interoperability between satellite imagers (2.5.4), and existing multispectral vessel datasets are discussed (2.5.5).

2.5.1 Machine Learning in Satellites

Machine learning is maturing in terrestrial applications, but it is only emerging for onboard satellite processing. The first satellite capable of onboard machine learning was the Φ-Sat-1 (Phi-Sat-1) in 2020 which utilized a simple segmentation network structure (Figure 10) to detect and identify clouds present in images to optimize downlink bandwidth [9]. The model was able to achieve over a 95% accuracy for cloud detection during model testing. The practical implications of this for the satellite was a possible 30% reduction of downlink bandwidth required due to filtering out unsuitable images [62].

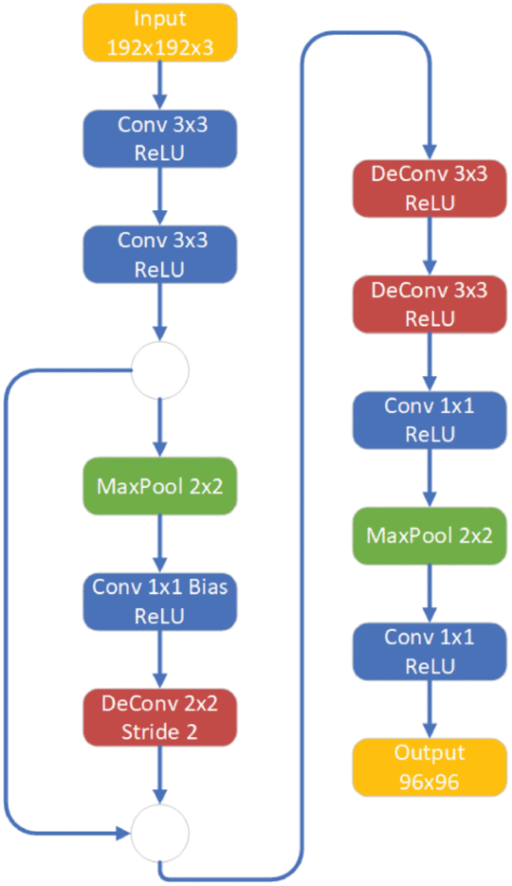


Figure 10 CloudScout segmentation network architecture for Φ-Sat-1, the first deep neural network onboard a satellite for Earth observation [9]. This was used to reduce downlink bandwidth requirements by detecting, and filtering out, clouds present in images.

If only a fraction of captured images is useful, then there is limited value in transmitting all of them to Earth. Additionally, by relying on edge computing, it is possible to reduce steps in the data processing chain. Relying on terrestrial computing means that after the data is downlinked to an Earth-based station, it is transmitted to a data center to process the captured information before delivery to an end-user. If that end-user is another satellite, it is then uplinkeded instead of directly transmitted by inter-satellite hops. Other proposed applications for onboard satellite ML include disaster monitoring, managing satellite communication demands utilizing beam-hopping techniques, fault detection, isolation, and recovery (FDIR), as well as real-time ML for collision and obstacle avoidance tasks [63]. Many of these proposed objectives would utilize deep neural networks, such as convolutional neural networks which are also useful tools for image analysis.

To quantify the benefits of onboard processing the exact implementation of the machine learning model and satellite hardware must be examined. For instance, Φ-Sat-1 [9] was able to reduce downlinked image volume by 30%, meaning that the transmitter required 30% less operation time to downlink the captured imagery. As such, power consumption which ranges from 5 to 15W [64] [65] for a CubeSat transmitter would also reduce by 30%. If transmission speed was the constraint of Φ-Sat-1, 30% more imagery could be captured and downlinked. For vessel detection, the benefits would be only requiring transmission of the immediate image area around a detected vessel, or simply coordinates and size of the detected vessel.

2.5.2 High Dimensional Spectral Imagery

One of the difficulties in extending ML to higher-dimensional spectral imagery is that little attention has been paid to it compared to 3-channel spectral imagery. ML applied to MS satellite imagery may utilize models that accept only 3 of the available spectral channels (e.g., Φ-Sat-1 [9]) to simplify development and might be pre-trained with common RGB datasets like ImageNet to avoid overfitting data. Another issue with multispectral data is the increased computational resources to train a model [66], a consequence of higher spectral channel count compared to RGB imagery. However, there are early efforts towards this on multispectral ML applied to vessel detection [5, 43], however there are fewer resources available on the subject.

2.5.3 Convolutional Neural network.

Convolutional neural networks (CNNs) are mathematical models built of layers of “neurons” like most ML model and utilize convolution layers. ML neurons resemble neurons in animal nervous tissue, in both appearance and function. Figure 11 shows an example of an ML neuron structure. The inputs to a neuron may range from the colour values for a pixel within an image to the output of other neurons.

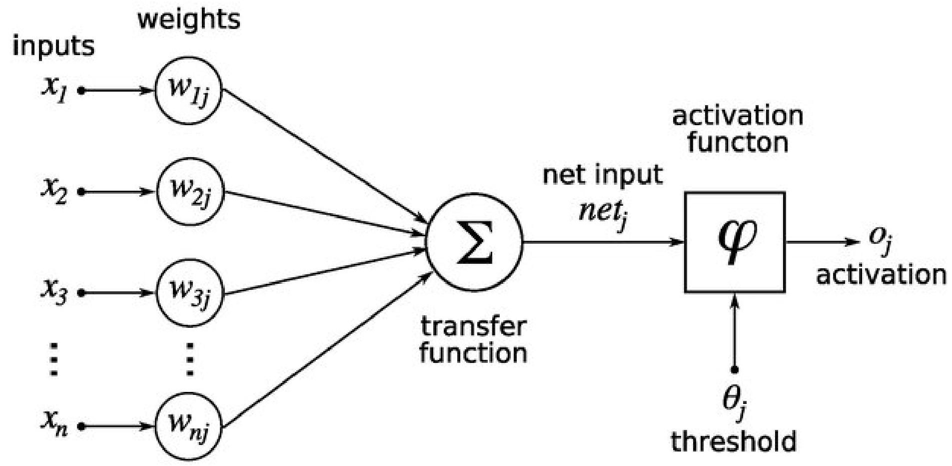


Figure 11 Illustration of neuron structure in a neural network [67] highlighting its operation, trainable parameter (weights), and other functions within the neuron.

Utilized in different configurations, these neurons form the basis of how ML networks are structured and trained, with the weights, w_{nj}, of each neuron adjusted over time. Convolutions, the integral layer from which CNNs received their name, are trained kernels applied to an input array or impact, to produce an activation map. Combined with activation and pooling layers, these form the basis of CNNs (Figure 12). An example of a simple CNN architecture is shown in Figure 13.

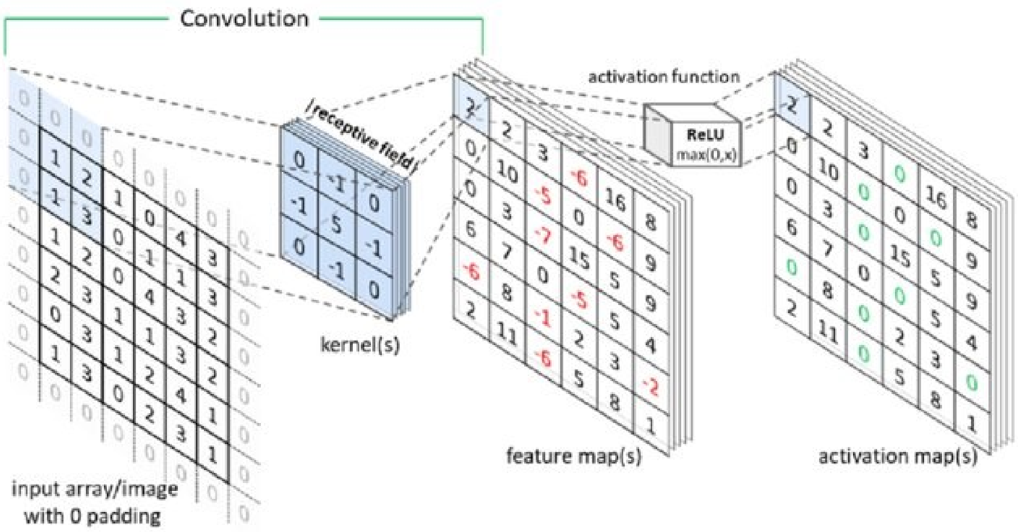


Figure 12 An example of a convolution layer of a CNN [26]

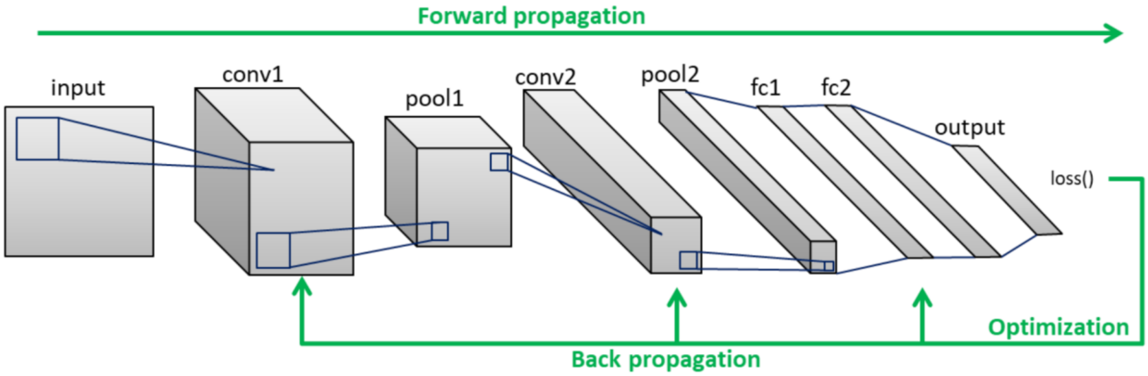


Figure 13 An abstract example of a simple CNN architecture [26]

Training a CNN can be achieved in different ways depending on the model structure. For the application of CNNs to images, however, the most popular method is to utilize supervised learning. In supervised learning a model is trained upon a dataset of input data (e.g., images) with labels ranging from those for classification of an image to coordinates of a feature within the image (such as vessel position). A portion of the dataset is reserved to validate and/or test the model. This ensures the model is generalized – i.e., suitable for cases beyond the dataset – and not overfitting to the dataset used.

2.5.4 Data Interoperability

One challenge to developing MS ML models for satellites is the availability of potential datasets to train the model. MS imagery generally features unique combinations of spectral bands that are not duplicated from sensor to sensor. This can lead to difficulties when trying to implement ML models for satellite systems with new or uncommon sensors; a dataset must have appropriate imagery to train on, but there may not be imagery available without launching a satellite. There are two common methods to avoid this problem. The first is to launch a preliminary satellite or utilize an existing satellite with the selected or similar imaging payload to capture imagery for a dataset and subsequent model to be developed. The second is to utilize images from similar existing imaging sensors.

Φ-Sat-1 utilized HyperScout 2, a hyperspectral imager developed by Cosine to capture its imagery. However, the dataset the model was trained upon was constructed from Sentinel-2 imagery, with the model launched on the satellite trained on the B01, B02 and B8A spectral bands of Sentinel-2 with principal component analysis [9]. These spectral bands, respectively, had central wavelengths of 450 nm, 494 nm and 862 nm, and represented bands similar to those available on the HyperScout 2.

Utilizing data from multiple sources to produce a dataset for a different sensor is not a new concept – referred to as data interoperability [68]. For instance, Sentinel-2 is a constellation of two satellites (Sentinel 2A and 2B) with very similar imaging sensors. While not identical, imagery from both satellites are regularly used interchangeably. Data interoperability may be extended to satellites with more significantly different imagers and orbits. For instance, the data interoperability of Landsat 8's Operational Land Imager (OLI) and Sentinel 2's Multi-Spectral Instrument (MSI) could be optimized [68]. A spectral response comparison of 4 similar bands from Sentinel-2 MSI and Landsat 8 OLI is shown in Figure 14. Each imaging sensor monitors different spectral bands of light, and generally have different capabilities, even though all three are multispectral sensors. The paper also outlines improvements which could be made to improve satellite data interoperability in general.

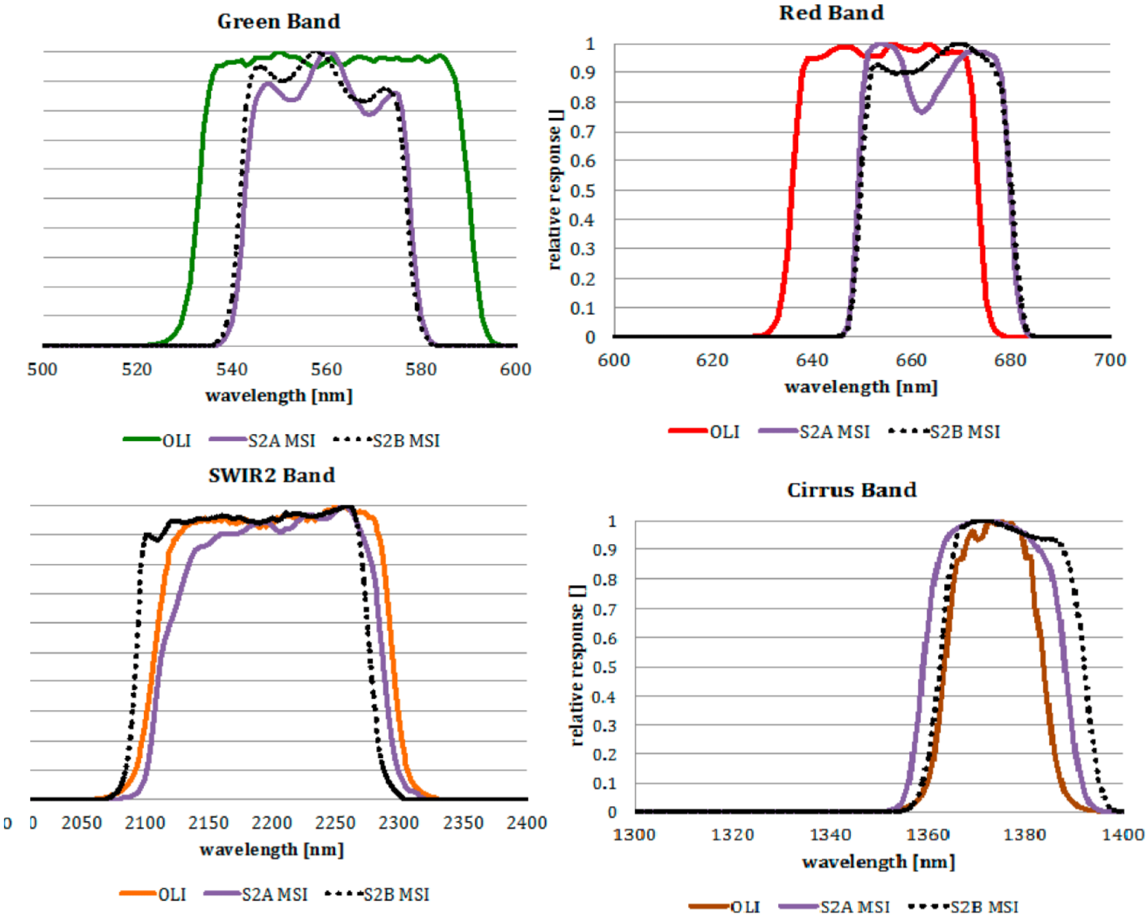


Figure 14 Comparison of the spectral response functions for 4 of Landsat-8 OLI and Sentinel-2 MSI spectral bands [68]. Due to similarities between spectral bands of the satellites (demonstrated in these cases), data can be used within certain applications interchangeably, such as using a dataset created from one satellite imager with another different imager.

2.5.5 Existing Multispectral Vessel Datasets

One of the main problems identified with developing a trained model for vessel detection was the near absence of any existing vessel detection datasets. Only two publicly available datasets for vessel detection were identified, though both had deficiencies limiting use for this thesis. The first, entitled “Contrastive Self-supervised Learning for Ship Detection in Sentinel 2 Images” [43], comprised of 16 large multispectral images with pixel level vessel annotation, and 12 of the 13 spectral bands of Sentinel 2. However, the images were relatively clear of clouds, did not have associated vessel AIS data contained, and were built from only 16 images of harbors containing 1053 ship annotations. The other dataset, entitled “Sentinel-2 dataset for ship detection” [69] was just over 1200 multispectral images of vessels around Denmark, and additional non-vessel images to assist with training. The images measured only 64 by 64 pixels (i.e., too small for better classification), used per-image classification, and only utilized 5 of the 13 possible MS bands. The absence of a suitable training dataset requires the creation of a new one which will be shared with others.

2.6 Thesis Objective

As stated in chapter 1 (Introduction), the goal of this research is to train and demonstrate an ML model that can detect marine vessels within multispectral images, and the benefit of including spectral bands beyond the RGB ones. The model must be suitable to implement on processors representative of CubeSat OBC to run CNN models. Permutations of different spectral band combinations will be trained to examine the effect of spectral content and wavelength on detection performance. Finally, the inferencing rate over several orbital paths relevant to time-over-water and other imaging period constraints are examined to ensure performance of the Nvidia SBCs is suitable for onboard processing. To ensure the steps in creating the dataset, models, and tests are transparent and repeatable, the details of each process are laid out in the methodology (Chapter 3).

Chapter 3 Methodology

In this chapter, the steps and procedure to create and test components and methods needed for the thesis will be detailed. The components can be divided into three subsections: dataset creation (3.1), machine learning model (3.2) and satellite orbital simulation and hardware processing (3.3). The dataset creation section will address the multispectral imagery, AIS message geo-referencing, and methods to create the dataset. The machine learning model section will explain both the CNN model used and the metrics which will be used to analyze the performance of the models. For the final section the satellite orbital simulation portion will detail the orbital simulation under different imaging conditions to understand valid imaging windows. The hardware configuration used to test the inferencing speed of the model, and to produce data to compare the orbit simulation against, is also discussed.

3.1 Dataset Creation

Due to the lack of adequate multispectral vessel datasets with all spectral bands, sufficient annotation, and sufficient class (vessel) support for this thesis, it was necessary to create a new dataset. To create the dataset, AIS data was used to geo-reference vessel locations within Sentinel-2 MS imagery. The intent was to create a dataset featuring any identified ship within the limitations of the satellite imagery (primarily GSD) and available AIS messages. The following sections will detail the process of collecting and handling the satellite imagery (3.1.1) and AIS data (3.1.2).

3.1.1 Satellite Imagery

The multispectral satellite imagery used in the dataset was collected from the two European Space Agency's (ESA) Sentinel-2 Satellites through the Copernicus Programme. Sentinel-2 imagery has the highest spectral band count and spatial resolution multispectral imagery that could be acquired with the resources available. Each multispectral image tile from a region is measured at 110 km by 110 km with a mixed GSD of 10 m to 60 m depending on the specific band. The bandwidths covered range from 424.7nm to 2289.4nm for the Sentinel 2A satellite, covering visible light to short-wave infrared. The bandwidths covered, GSD, and central band frequencies are summarized in Table 4. The Sentinel 2B satellite monitors offer marginally different wavelengths, but generally ranges within 1-10nm wavelengths when compared to Sentinel 2A and are thus used interchangeably.

Table 4 Sentinel 2A spectral resolution and ground sample distance [41]

band	**centre wavelength (nm)**	**bandwidth (nm)**	**gsd (m)**	**description**
1	492.7	65	60	Aerosol
2	559.8	35	10	Blue
3	664.6	30	10	Green
4	664.6	30	10	Red
5	704.1	14	20	Red Edge
6	740.5	14	20	Red Edge
7	782.8	19	20	Red Edge
8	832.8	105	10	NIR
8a	864.7	21	20	Narrow NIR
9	945.1	19	60	NIR
10	1373.5	29	60	SWIR
11	1613.7	90	20	SWIR
12	2202.4	174	20	SWIR

The ESA also offers the spectral response functions for the visible and near infrared (VNIR), and SWIR imaging sensors used in Sentinel-2A and 2B. These spectral response functions are also represented as plots in available documentation, shown in Figure 15 and Figure 16.

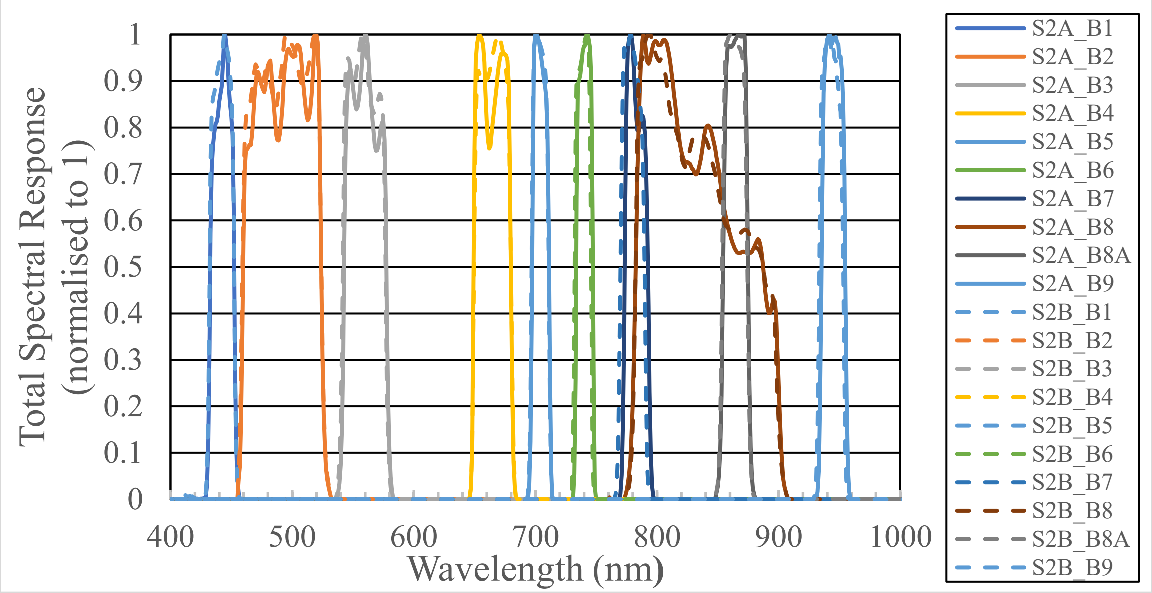


Figure 15 S2A & S2B MSI spectral response average – VNIR [41]. This figure shows spectral bandwidth and location of Sentinel-2 VNIR, as well as the overlap that occurs between B07, B08, and B08A.

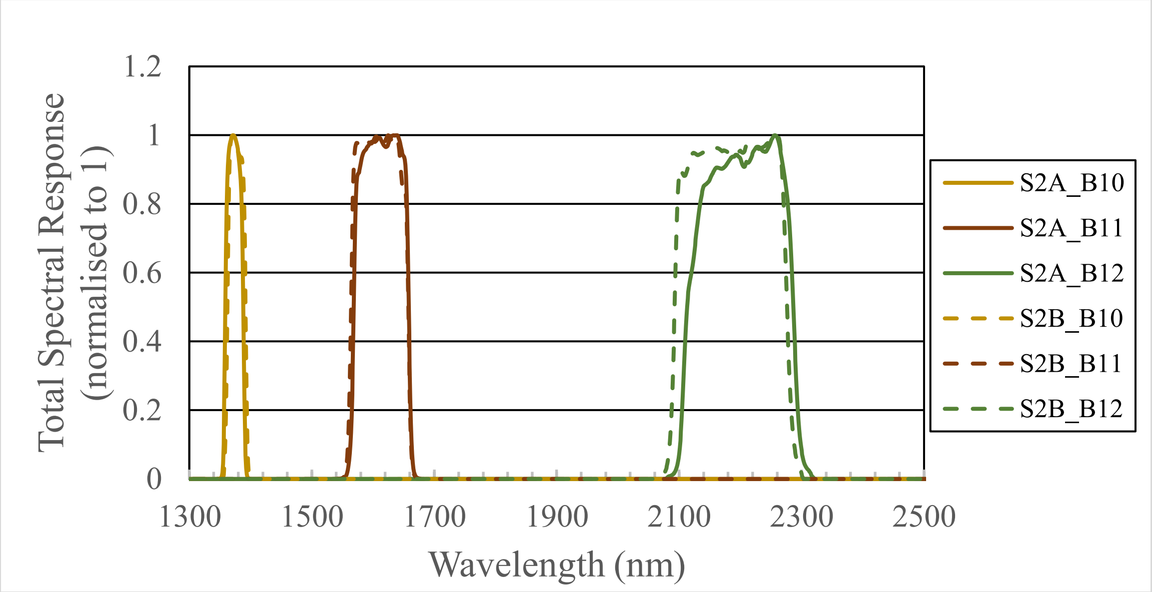


Figure 16 S2A & S2B MSI spectral response average – SWIR [41]. This figure shows spectral bandwidth and location of Sentinel-2 SWIR bands.

Satellite imagery is traditionally processed through multiple levels prior to delivery. For satellite imagery, the processing is represented as a Level between 0 and 4. Level-0 imagery is the raw captured sensor data before it is processed, time-stamped, or geo-referenced. At the other end, Level-4 data has the greatest amount of processing applied and may include modelled measurements compiled from several satellites.

Sentinel-2 tiles could be collected as Level-1C and Level-2A images. Level-1B imagery is only available to users through special request, and the imagery is not ortho-rectified or spatially registered. A summary of the differences between the Sentinel-2 imagery Levels can be found in Table 5. With the necessity of geographically correct imagery for geo-localizing AIS messages within imagery, Level-1B was removed from consideration. Given the goal of replicating the environment of onboard processing, Level-1C was selected as it represents the most limited level of processed imagery suitable for the task. Further processing, in the context of onboard processing, would require further utilization of onboard resources. It is possible to replicate the data processing as ESA offers the algorithms and code to convert Sentinel 2 Level-1C imagery to Level-2A for instance. However, this is outside the scope of the thesis.

Table 5 Sentinel-2 product processing level description [41]

product level	**incremental improvements from previous level**
Level 1B	▪ radiometric corrections ▪ geometric viewing model refinement
Level 1C	▪ resampling (geometric interpolation grid computation) ▪ conversion to reflectances ▪ preview images and mask generations
Level 2A	▪ scene classification ▪ atmospheric corrections (S2AC)

Based on the availability of AIS data (discussed further in the following section), imagery collection for vessel imagery was based around the high-traffic marine areas of Denmark and the USA. Altogether, 17 110 km by 110 km image tiles were collected over Denmark, along with 84 from the USA. Beyond the creation of vessel imagery for the dataset, it was also good practice to train the dataset on non-vessel imagery so it can better recognize non-vessel features within the image. To that end, in addition to utilizing non-vessel areas of the Denmark and USA imagery, the imagery collected was supplemented by an additional 35 images from randomly selected areas around the world for non-ship marine imagery to prevent the ML model from over-fitting to the selected datasets.

Given the mixed GSD of Sentinel-2 imagery, images from each spectral band were interpolated to the same number of pixels as the 10 m GSD channel (10980). The number of pixels in channels with 20 m or 60 m GSD was increased using the Scikit-image rescale function (Python library) to match the 10 m GSD channel. This Scikit-image function increased the number of pixels for each channel and interpolates the values using nearest-neighbor interpolation. Each of the 136 Sentinel-2 tiles were collected, converted, and saved as an unsigned 8-bit integer NumPy array of size $13 \times 10980 \times 10980$ from the original 12-bit radiometric resolution. During the conversion process, the 1% lows and the 99% highs were mapped to 0 and 255, respectively, within the image. While the imagery could have been saved as a 16-bit float array to preserve the radiometric-resolution, it was infeasible to store the imagery without removing data from intermediate steps given its volume and hardware limitations of computer memory and storage for processing.

To geo-reference each vessel within their respective image, each image required an associated geographic grid to map each pixel to each AIS message. Using the metadata accompanying the Sentinel-2 imagery and processing it with the rasterio geographic image python library [70], a $2 \times 10980 \times 10980$ NumPy array was created containing the latitude and longitude of each point.

3.1.2 AIS Data Collection

The ships within the MS imagery are identified using date-matched historic AIS data. Both Denmark and USA offer such AIS data, originally collected through respective government organizations, at no cost to the thesis. Other countries offer limited historic AIS data but had deficiencies for the purpose of geo-referencing ships within multispectral images. Third-party historic AIS providers were considered and contacted but were infeasible for inclusion into the dataset due to cost or lack of response. As a result, only images from Denmark and USA were used to construct the dataset. Danish AIS data was collected from the Danish Maritime Authority [71], and USA AIS data was collected from the MarineCadastre.gov [14], a collaborative effort between the Bureau of Ocean Energy Management and the National Oceanic and Atmospheric Administration

The AIS data that was collected corresponded to the date each Sentinel-2 tile was captured. Each AIS data file was supplied in the form of a comma separated (.csv) file and included all AIS messages in the country's immediate area over a 24-period. Due to the volume of data present in each file – e.g., over 100 million AIS messages in one such file – the volume had to be significantly reduced to feasibly work with. The category names in both Denmark and USA data were adjusted to match, and incomplete or irrelevant data unique to either countries' data format were removed. AIS messages of AIS transceiver class B were discarded, as these vessel classes are not mandated to carry AIS transponders or to populate all fields in their messages. AIS messages were initially eliminated to within ± 6 minutes of the time the multispectral image was captured, ± twice the mandated minimum broadcast interval for when class A AIS transponder ships are anchored. Finally, the message occurring nearest to time of capture of the multispectral image was used to generate geo-referenced ship positions. Speed-over-ground and heading were used to calculate any vessel displacement if the MS image capture time differed from AIS message time. An example of these positions generated can be seen in Figure 17.

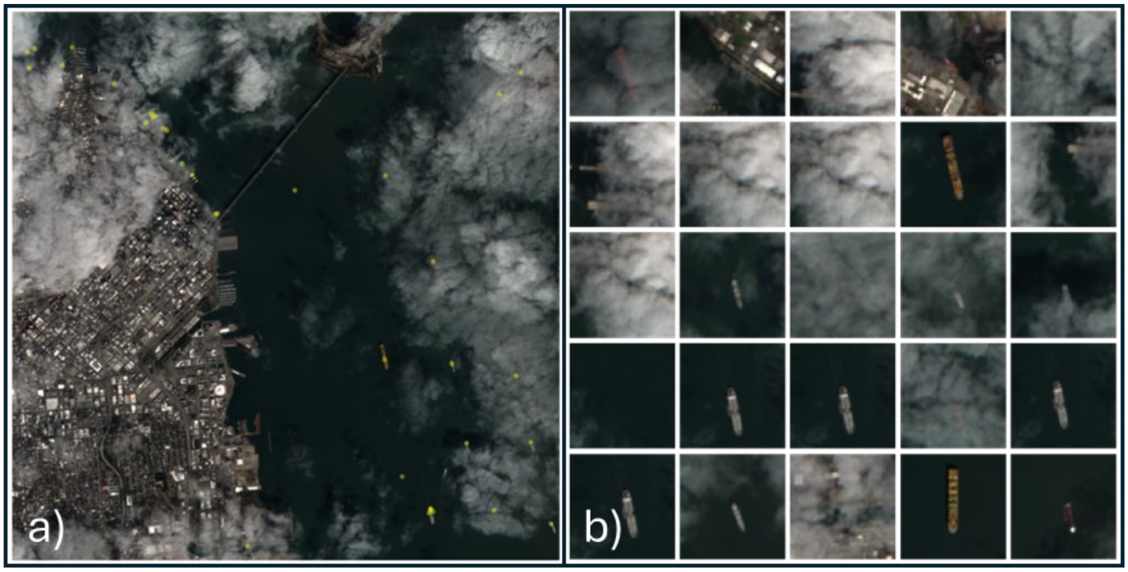


Figure 17 (a) RGB composite image from Sentinel 2 with geo-referenced ship positions (yellow dots) and b) immediate area (64 × 64 pixels) around 25 of the georeferenced points within the image. Comparison of imagery with geo-referenced positions.

The dataset image size was selected as 128 × 128 pixel (10 m GSD) regions, or 1.28 × 1.28 km when GSD is considered. This size was selected to restrain the end model size to just what was needed and to reduce occurrences of multiple ships being present in the same image, as each image only has a single AIS reading associated with it. However, despite the process of georeferencing and cropping images from the larger tiles producing many suitable images, a systematic problem emerged. Whether a result of an inaccurate time stamp on the image tile, AIS message inaccuracies, or an undetermined problem with the data or process, vessels with a SOG much greater than 0 would be inconsistently out of place. The issue was not uniform as only some vessels were affected and to varying degrees. This problem was resistant to straightforward attempts to fix it. While this could have been investigated further, it was determined beyond the scope of the thesis.

Therefore, to compensate for this, each image with a vessel was instead first cropped into a larger 768 × 768 pixel image, and the ship in each position manually identified. The image is then cropped again into a smaller 128 × 128 pixel image. Figure 18 shows an example of this cropping process. Vessels of less than 15 m length overall and 6 m beam were removed from consideration due to difficulties labelling the data manually. While this does offer a level of uncertainty from human error between AIS message and the ship within the image, all efforts were made to minimize this. The one critical issue, however, was vessels that were partially obfuscated behind thin to light cloud layers were identified according to only their RGB composite image instead of their full multispectral combinations during the cropping and labelling steps. While stationary vessels were unaffected during the cropping step, moving vessels were affected due to the mismatch between their AIS message and MS imagery. Image labelling did not utilize additional spectral bands due to limited room for supporting information – primarily the displaying the parent image for scene context, the AIS message, and heading. This likely lowered the quantity of lightly and heavily obscured ships within the final dataset, but there is no way to verify this claim.

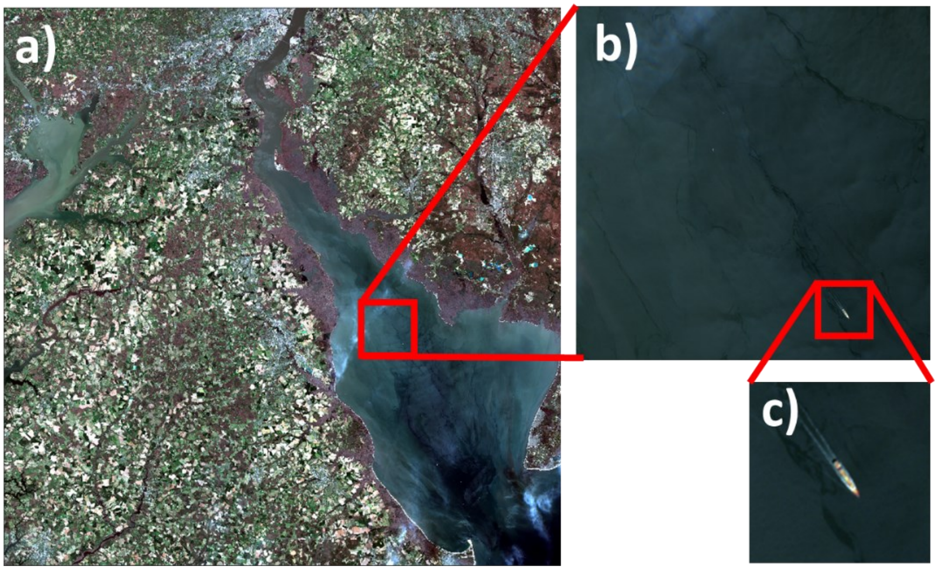


Figure 18 Representation of the Sentinel-2 imagery processing steps with a) RGB composite image of a Level-1C Denmark Tile; b) automatic crop to vessels reported position, and c) manual crop to ship's actual position. The images display the difference between reported and actual position within the image, and the steps to correct them.

When an image was cropped, the resultant image was matched with the accompanying AIS data. An example of a paired image and AIS message for a vessel can be seen in Figure 19 (image) and Table 6 (AIS message).

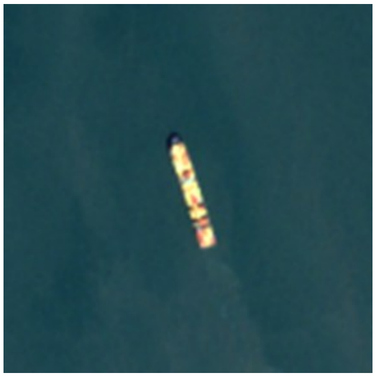


Figure 19 RGB composite of cropped vessel image in the dataset that is correlated with the AIS message in Table 6.

Table 6 AIS message of vessel in Figure 19.

description	value	description	value
image name	Image_398.npy	vessel type	70 (Cargo)
vessel name	MSC VEGA	length	366m
MMSI	636015506	width	48m
latitude	37.76904°	transceiver class	A
longitude	-122.35794°	date	2022-03-20
speed-over-ground	0	time	18:51:06
heading	347°		

The final step of creating the dataset was to collect a range of additional marine but non-ship images from around the world. These include inland, coastline, and mid-ocean areas. Each image was manually checked to ensure no visible vessel was present. Each image was then labelled for level of cloud coverage, whether it was a coastal region, and if a ship was visible within the image. These labels were primarily intended to help determine suitability of the image during test inferencing and to eliminate unsuitable ship images.

To complete the assessment of the dataset’s suitability for vessel object detection, the objects within the images to be detected – marine vessels – must be labelled. The options considered were semantic segmentation (label each pixel with a vessel belonging to the desired vessel class) and bounding box labelling (draw a box around the vessels within an image). With over a thousand images to label, semantic segmentation was ruled out due to time limitations. During the labelling, some images were removed due to redundancy. This generally occurred when a vessel was docked near other vessels within an image where every vessel was broadcasting a valid AIS message. An example of this is in Figure 20 where two ships are docked alongside beside each other. The concern is that as multiple images are derived from the original images, the model's performance could be measured by images which contained a majority of the same pixels. In effect, this would increase the likelihood of overfitting and, critically, risk testing performance on what is virtually training data if the "same" image was present in both the training and testing split.

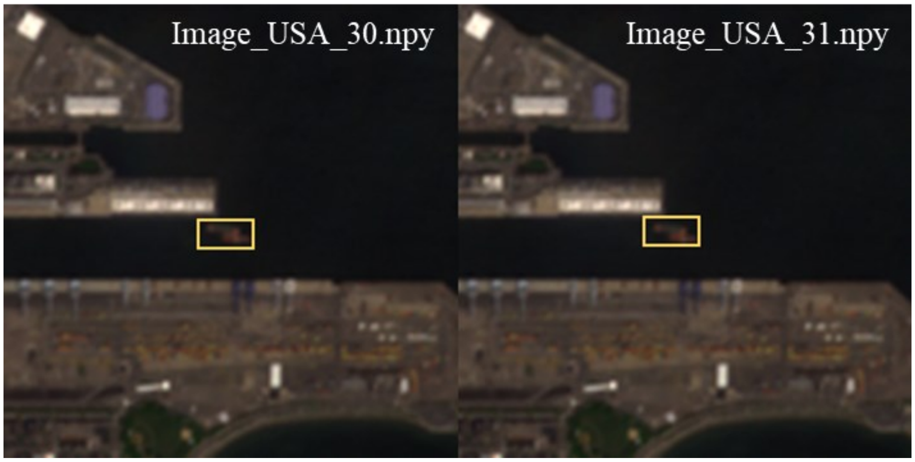


Figure 20 Visually similar repeated images caused by two docked vessels highlighted by yellow squares each with a unique AIS message. The two produced images are cropped from the same parent image. One image is retained, the other is removed from the dataset to reduce chance of overfitting and testing on (effectively) training data.

To summarize the results of each step of the process, 1736 vessel images and 1260 non-vessel images were created from the USA MS imagery. Another 520 vessel images and 510 non-vessel images were created from the Denmark MS imagery. Supplementary additional MS imagery yielded another 1050 non-vessel imagery, all of which was manually checked and labelled. After filtering out unsuitable imagery, such as absent vessels in composite RGB imagery due to heavy clouds or near-repeated imagery, 635 vessel images from the USA MS imagery were labelled utilizing bounding boxes, along with another 458 vessel images from Denmark. A multispectral vessel image used in the final dataset is shown in Figure 21, with each channel normalized independently to highlight the differences for each pixel within the image. Due to consistent values in channel B10, the image for B10 appears black and white as the maximum and minimum values were 244 and 243, which were normalized to 255 and 0. The reason for these consistent values is the presence of water vapor for which B10 is heavily reflected by (see Figure 22). Due to this interaction, B10 only observes high altitude features such as cirrus clouds or ground higher than 2000 m.

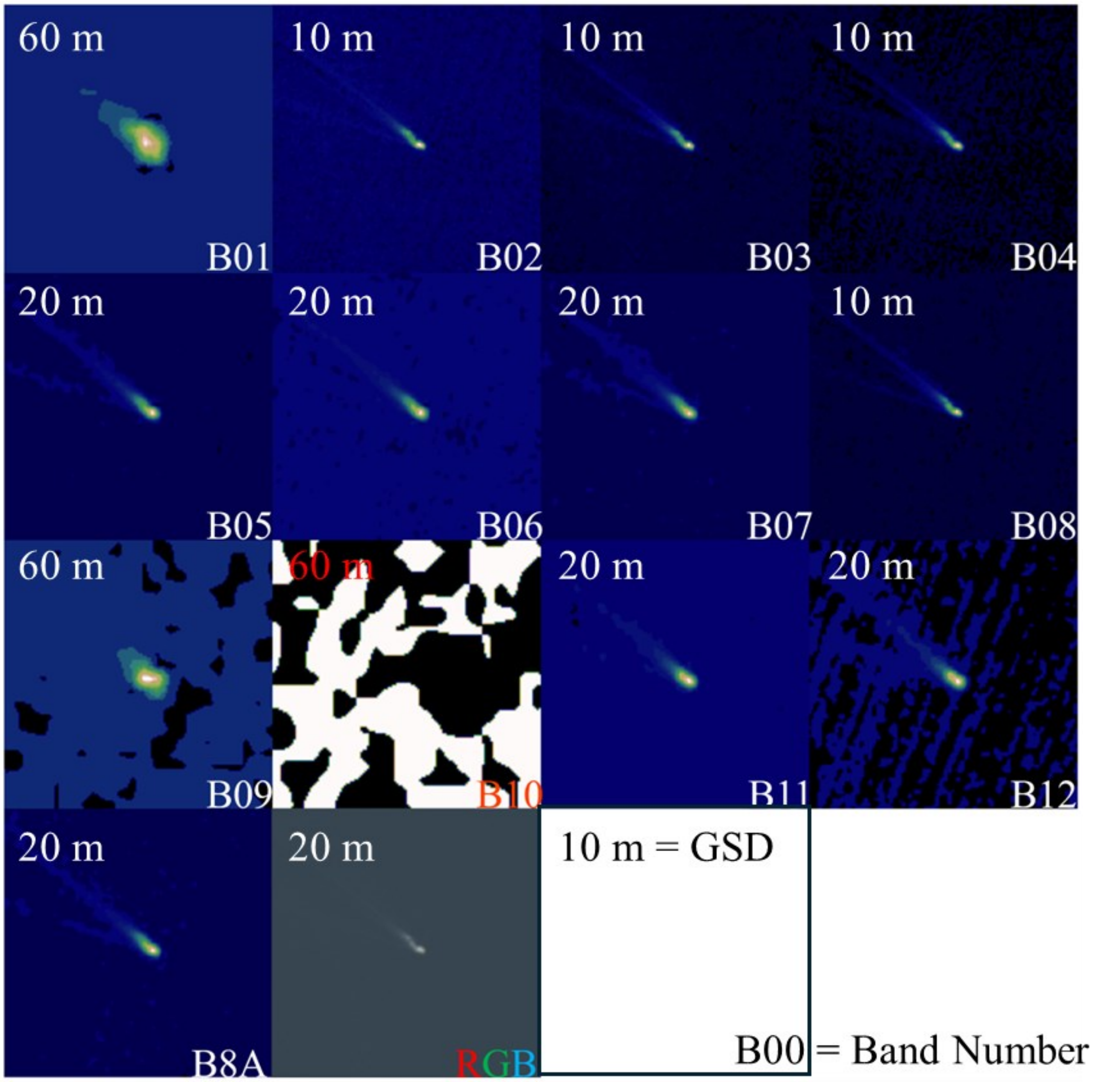


Figure 21 Breakdown of multispectral image by channel. Normalized values per band were plotted to highlight pixel difference along with non-normalized RGB

composite image for Denmark vessel image #250. Additional spectral bands increase available information.

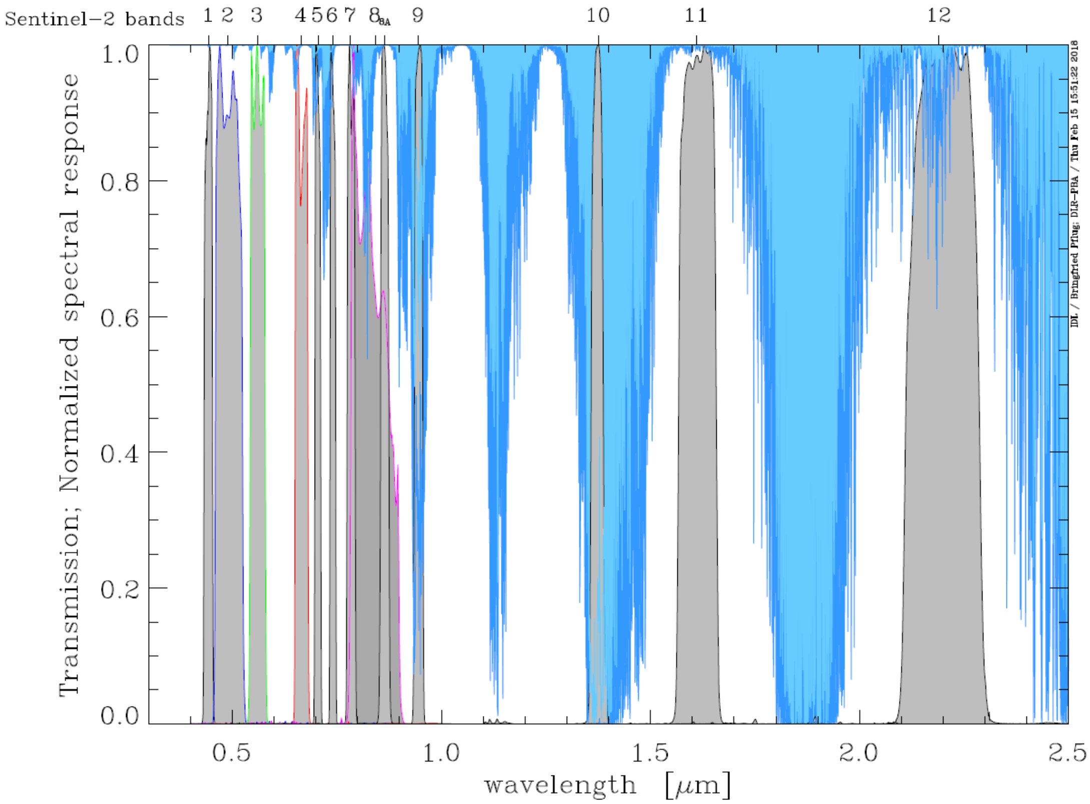


Figure 22 Spectral response functions and transmission due to vertical column water vapour absorption [41]. B10 was selected for Sentinel-2 for its water vapour interaction and is incapable of detecting features at sea level.

3.2 Machine Learning Model

For the purposes of coding, constructing, and training the ML model, the scripting language, Python, was selected. The reason was that Python has a wide range of ML libraries, including PyTorch which was used for this thesis. PyTorch [72] is a ML library which is an end-to-end machine learning model framework used to construct and train the models considered. There were several CNN model structures considered for vessel detection and the two that were initially selected: Faster R-CNN with a ResNet-50 feature pyramid network (FPN), and Ultralyics' YOLOv8 model. The model for Faster R-CNN model [73] is the "fasterrccn_resnet50_fpn_v2" model in PyTorch [72], with the only changes being modifications to the input layer to accept high channel count imagery and operating without initialized weights. Pre-processing, image augmentation, data loading, and other training codes had to be created. All coding was completed using the JetBrains PyCharm Integrated Development Environment (IDE).

Compared to the faster R-CNN, YOLOv8 [74] is a complete training pipeline designed to be fully functional with only a set of input data (RGB or monochrome) and associated bounding box labels required to train a model. The underlying model itself supports multispectral imagery, but supporting framework and functions– from augmentation, memory checks, and simply data input – rely on the Pillow [75] and OpenCV Python library [76]. Many functions within these code libraries support 4 or fewer channels and must be removed and/or have their functionality replicated to support up to the required 13 spectral channels for this thesis. As a result, YOLOv8 was modified so it would accept multi-channel NumPy arrays or multi-channel TIFF images. The structure for YOLOv8 is shown in Figure 23 for a 3-channel input (such as RGB).

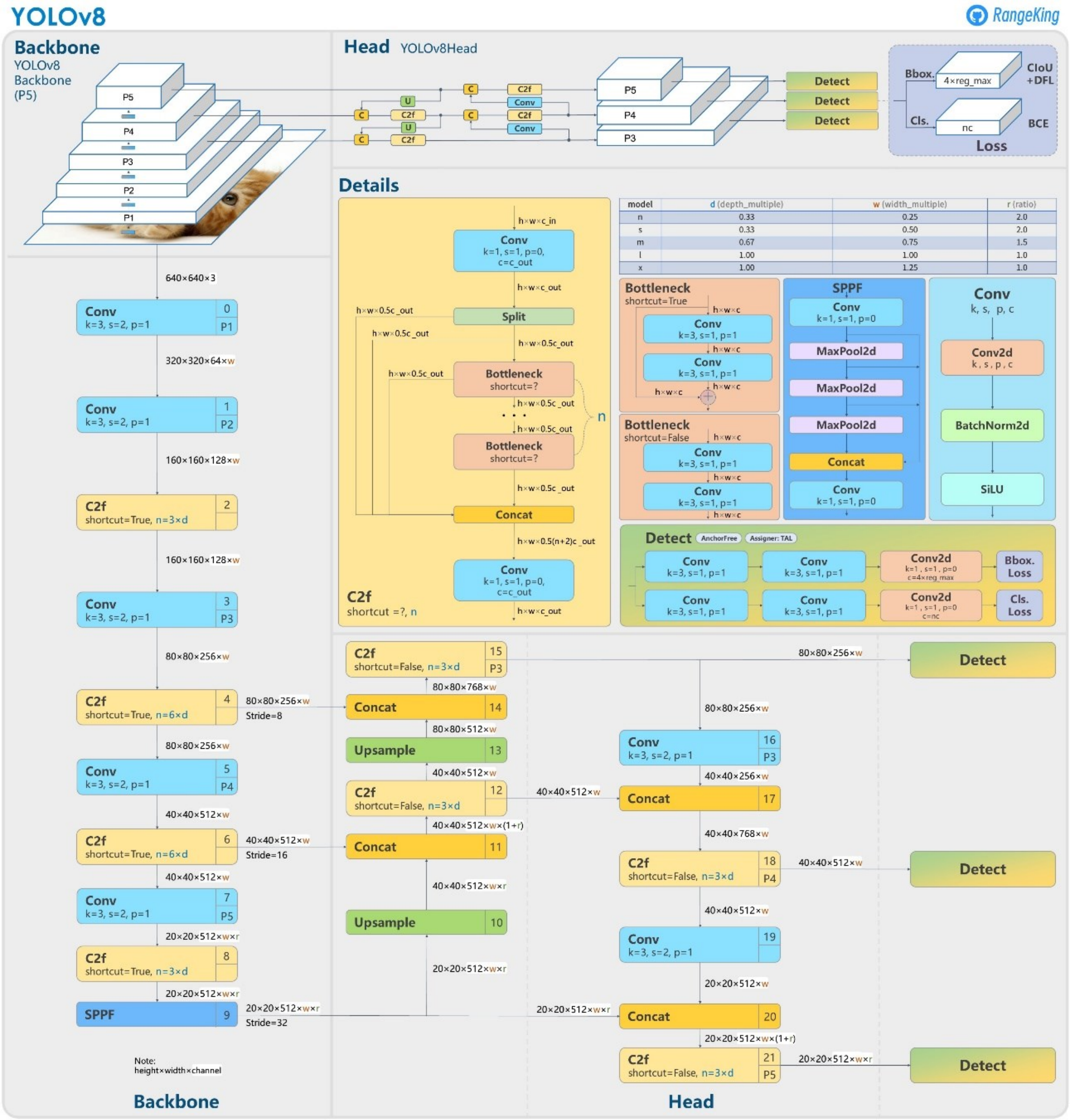


Figure 23 YOLOv8 model structure for 3-channel imagery [77], unmodified and suitable for RGB images. YOLOv8 was modified to accept a variable quantity of spectral channels compared to the original 3.

Due to the difficulties obtaining a performant trained model, the Resnet-50-FPN object detection model was abandoned. This will be discussed further in the future work (section 5.1), but the model experienced significant problems with the quality of the solution. Therefore, performance of the models will only be discussed for the YOLOv8 models.

3.2.1 Model Training Configuration

The modified YOLOv8 model training – and the failed Faster R-CNN training – was implemented on a desktop computer with an NVIDIA RTX 3090 GPU, 64GB of memory, and an Intel i9-10900 10-core processor. The dataset was split into train, test, and validation datasets with a 75%, 12.5%, 12.5% split, respectively. The same split of images was used when resampling the dataset for different band count imagery. Each model was trained for 4000 epochs, with the model weights from best performing and final epoch saved. The band combination permutations are outlined in green in Figure 7.

Table 7 Trained spectral band combinations. Green cells in a row indicate the bands (column) used with the specific band combination (row).

			Sentinel-2 Spectral Bands												
Name	Size	Abbrev.	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B09	B10	B11	B12
All_Band	Nano	ABn	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All_Band	Small	ABs	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All_Band	Med.	ABm	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All_Band	Large	ABl	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
RGB	Nano	RGBn		🟩	🟩	🟩									
RGB	Small	RGBs		🟩	🟩	🟩									
RGB	Med.	RGBm		🟩	🟩	🟩									
RGB	Large	RGBl		🟩	🟩	🟩									
RGB + B08 / 10m	Med.	R08		🟩	🟩	🟩				🟩					
20m GSD	Med.	G2					🟩	🟩	🟩		🟩			🟩	🟩
60m GSD	Med.	G6	🟩									🟩	🟩		
10m + 20m	Nano	G12n		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩
10m + 20m	Small	G12s		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩
10m + 20m	Med.	G12m		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩
10m + 20m	Large	G12l		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩			🟩	🟩
10m + 60m	Med.	G16	🟩	🟩	🟩	🟩				🟩		🟩	🟩		
20m + 60m	Med.	G26	🟩				🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩
RGB + B01	Med.	R01	🟩	🟩	🟩	🟩									
RGB + B05	Med.	R05		🟩	🟩	🟩	🟩								
RGB + B06	Med.	R06		🟩	🟩	🟩		🟩							
RGB + B07	Med.	R07		🟩	🟩	🟩			🟩						
RGB + B8A	Med.	R8A		🟩	🟩	🟩					🟩				
RGB + B09	Med.	R09		🟩	🟩	🟩						🟩			
RGB + B10	Med.	R10		🟩	🟩	🟩							🟩		
RGB + B11	Med.	R11		🟩	🟩	🟩								🟩	
RGB + B12	Med.	R12		🟩	🟩	🟩									🟩
All Band - B01	Med.	A01		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B02	Med.	A02	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B03	Med.	A03	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B04	Med.	A04	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B05	Med.	A05	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B06	Med.	A06	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩	🟩
All Band - B07	Med.	A07	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩
All Band - B08	Med.	A08	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩
All Band - B8A	Med.	A8A	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩
All Band - B09	Med.	A09	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩
All Band - B10	Med.	A10	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩
All Band - B11	Med.	A11	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		🟩
All Band - B12	Med.	A12	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	

In addition to the different permutations of spectral band combinations, the size of YOLOv8 can be altered. This is important because as the YOLOv8 model size is increased, the number of parameters in the model changes, more layers are added to the CNN, or the YOLOv8 size increased. YOLOv8 has 5 default model sizes to select from: nano, small, medium, large, and extra-large. Based on a simple preliminary analysis of the *All Band* and RGB models, the medium model size was selected as the default for all spectral band combinations. In addition to the medium models for the RGBm, G12, and ABm spectral combinations, each of these three had nano, small, and large YOLOv8 models trained for them. These would be used to discuss model size impact on accuracy and inferencing speed.

Table 8 Modified YOLOv8 model parameter accounts for 13 channel models of different default sizes. The total number of parameters, and total model size, increases by 2 to 4 times for each size increase.

model size	**parameter count**
large	87,298,515
medium	39,898,195
small	13,810,643
nano	3,693,011

Inferencing the test dataset to evaluate performance metrics was completed on a different computer than that used for model training. This was due to the training computer being a shared research lab computer, in addition to lesser hardware being capable of testing (not training) the model. The testing computer was desktop computer a GTX 1050ti, 48GB of memory and an AMD R7 3700X 8-core processor.

3.2.2 Analysis and Discussion

The analysis of the ML model performance was achieved through calculating the standard precision, recall and F1-score against the model's confidence in its predictions. Precision represents the percentage of vessel predictions which are correct out of the validation dataset, as in Eq. (2):

$$precision = \frac{true\ positive}{true\ positive + false\ positive}. \quad (2)$$

Recall is the percentage of vessels present in the validation dataset which are correctly predicted as vessels as in Eq. (3):

$$recall = \frac{true\ positive}{true\ positive + false\ negative}. \quad (3)$$

Neither metric completely captures the performance. For instance, if every pixel within an image was predicted as a vessel, recall is unnaturally high as there are no vessels incorrectly predicted as part of the background (false negative). Yet, the precision, and critically the utility of the model, would be poor to detect vessels. On the other hand, a model which only predicts vessels that it has 100% confidence in, but predicts most actual vessels as part of the background, may have perfect precision but poor recall and again would be of limited use. The F1-score (Eq. (4)) is a commonly used accuracy metric. It combines precision and recall as a harmonic mean, which can be written as:

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

The combination of the two – and avoidance of problems native to each of them individually – is the main reason it will be used as the primary performance metric for the resultant models, though the others will be discussed in select cases. Precision, recall, and F1-score will be plotted against several YOLOv8 confidence minimums. Confidence in YOLOv8 is an internal measure of the model's certainty in its prediction, a product of the class certainty and objectness, i.e., the model's certainty a predicted box contains the object. In a trained model, increasing the minimum accepted confidence (MAC) increases precision overall but could also decrease recall. What is used for MAC depends on the application and what precision is required. For instance, sending a vessel to intercept a detected dark vessel far from shore demands a higher confidence (and consequently precision) than spotting a detected vessel that can be identified in a remote accessible land-based camera. To ensure that performance is considered at multiple MACs, recall, precision, and F1-score for each model will be compared at a MAC of 0.25 and at 0.4 to

0.8 with 0.1 step intervals (6 MAC values total). These values were selected based on default YOLOv8 minimum confidence threshold for prediction (0.25) and the examination of a continuous F1-score vs confidence curve for a trained model. 0.4 to 0.8 at 0.1 step intervals were identified as suitable for representing the shape of the curve as the threshold increases. The area under curve (AUC) for the F1-score vs MAC curve was used to produce a single easily comparable performance value for each model. Fine tuning for this parameter was completed on the results of the first set of trained models.

While there are other metrics, like comparing the overlap of the inferenced bounding box with the labelled bounding box (intersection of union), this relies on the quality of the label box size that were manually bound to 10m increments. Ultimately, the most important factor in performance is whether it detects all the vessels present, and accurately, in the image. Given the models will not be tuned with adjusted training parameters, meaningful changes will be coarser to account for possible. A change of 0.05 at one point (MAC) for recall (5% more vessels identified), precision (5% fewer background patches mislabeled as vessels), and F1-score was considered significant and warrant closer inspection. Performance was also considered meaningfully better if one model outperforms another at all recorded MACs, which also resulted into a better AUC value for the model.

3.3 Satellite Path and Hardware Processing

While the accuracy of an ML model is important, so is its deployment. One of the reasons that the model was developed from small images was to ensure the developed models could be deployed to the emulated hardware to inference images. The satellite orbit, orbital velocity, GSD, and swath directly impact the feasibility of onboard processing. Real-time processing can only be achieved if the hardware can process and inference images as quickly as it can capture them. If the system is incapable of achieving real-time process, the next question is whether it is capable of near real-time processing suitable for the application and platform. Given the orbital speed of a satellite, along with the sensor GSD and swath size, the volume of data can rapidly become unmanageable so the possibility of both real-time and near real-time processing must be explored.

As a simplified illustrative example, the approximate orbital velocity, minimum GSD, and swath of the Sentinel 2 satellites are 7.46 km/s, 10 m, and 290 km, respectively. If the curvature of the Earth is ignored, this means that for each second, Sentinel-2 produces 2163 km^2 of imagery, or a 21.6-megapixel image with no overlapping sections. Therefore, for real-time onboard edge processing, the system must be capable of passing, handling, and inferencing a 21.6-megapixel image – approximately 280 MB for unsigned 8-bit data with 13 color channels without compression – every second for real-time processing when all channels are resized to 10m GSD. With terrestrial computing resources, this is a manageable task, but with edge computing, it is not.

For meaningful edge compute, processing capabilities must be relevant to the application. For vessel detection, if delivery of satellite onboard processing (edge compute) results takes half a day, the edge compute becomes a tool to detect where vessels were not where they are. With the modifications made to the YOLOv8 model to permit multispectral imagery, and the absence of benchmarks for the modified model, it becomes pertinent to test suitable hardware configurations for the task and relate them to orbital conditions. The following sections explain the hardware and methods used to reference inferencing performance for the model, with hardware configurations (section 3.3.1) and orbital simulations (section 3.3.2).

3.3.1 Hardware Emulation Configuration

To emulate different hardware configurations of a satellite OBC, two NVIDIA SBCs will be used. Current and older versions of these boards are utilized in space products offered by several companies – including Galaxia Mission Systems – and due to NVIDIA's advanced work on CUDA cores, they represent a suitable platform for machine vision. The different hardware and power configurations are summarized in Table 9 from NVIDIA documentation [78] for information on maximum CPU or GPU frequencies. Both boards utilized Jetpack 5.1.2, the latest version of software available for both boards at the start of testing. Jetpack includes the Linux OS images, and hardware specific libraries, APIs, and developer tools for the Nvidia SBCs. To evaluate inference rate, the images will be loaded aboard a microSD card capable of up to 130 MB/s read and 30 MB/s write.

Table 9 Target hardware and power configurations tested.

NVIDIA SBC	power configuration	power mode ID
Orin Nano 8 GB [78]	15W	0
Orin Nano 8 GB	7.5W	1
Xavier NX 8 GB [78]	20W – 6 Core	8
Xavier NX 8 GB	10W – 4 Core	4

3.3.2 Orbital Simulations

To simulate the orbital characteristics and position of several satellite orbits relative to Earth, in the north-east-down (NED) reference frame, MathWorks MATLAB and its Aerospace Toolbox were used. For each orbit, the simulation plots the satellite latitude, longitude, and velocity at 1 second time steps for 30 days. Due to computational limits posed by system memory, the status of whether a satellite was in eclipse by the Earth or Moon at any point in its orbit was simulated at 4 second time steps and linearly interpolated to 1 second intervals. The eccentricity used was set as 0.001, to approximately match the non-zero eccentricity of the spacecraft. Alone, this information is of limited value to determine OBC processing limitations, as orbital velocity was nearly constant throughout. However, with the latitude and longitude of the satellite, and treating each as a nadir-pointing satellite, cross-referencing it with geological locations offer more meaningful information. For instance, as the trained model detects marine vessels, only ocean regions were considered.

General Bathymetric Chart of the Oceans (GEBCO) offers global gridded bathymetric data [79], that has elevational data in meters on a 15 arc-second interval grind represented as an 86400×43200 array of latitude, longitude, and elevations. At the equator this represents a grid approximately 463×463 m while at 45° north or south this represents a grid of 327×463 m as the distance between lines of longitude decrease closer to the poles. Every point along the 30-day satellite orbit (2,592,000 points) was compared against the GEBCO data – extrapolating linearly from the nearest longitude and latitude of the GEBCO data – to produce an elevation of the point nadir beneath the satellite. With these elevation measurements, it is possible to determine when the satellite was over areas that are below sea level. While there is land below sea level, there was no straightforward method to determine dry land below sea level from “wet” land below sea level.

With the bathymetric data brought into the simulation, every second of a nadir-pointed satellite’s orbit has a longitude, latitude, ground elevation, velocity, and eclipse state associated with it. The bathymetric data provides context to situations that can be broken down by the elevation (or depth) of an area to the inferencing rate of the different Nvidia SBCs configurations. For instance, a satellite deployed to exclusively monitor vessels does not need to capture images over land, during which it can process images should the inferencing rate fall short of real-time processing. With the bathymetric data, periods of time over areas below sea level can be discounted. An example of this is shown for a 705 km orbit with 98.2-degree inclination for a 1-day period in Figure 24 and for a 30-day period in Figure 25. It may be reasonable to remove portions of the capture area from consideration for data processor, such as when a satellite is in eclipse. Alternatively, a satellite may simply be deployed to monitor several discrete areas instead of the entire globe. The combinations of orbits and conditions are described in Table 10 and Table 11.

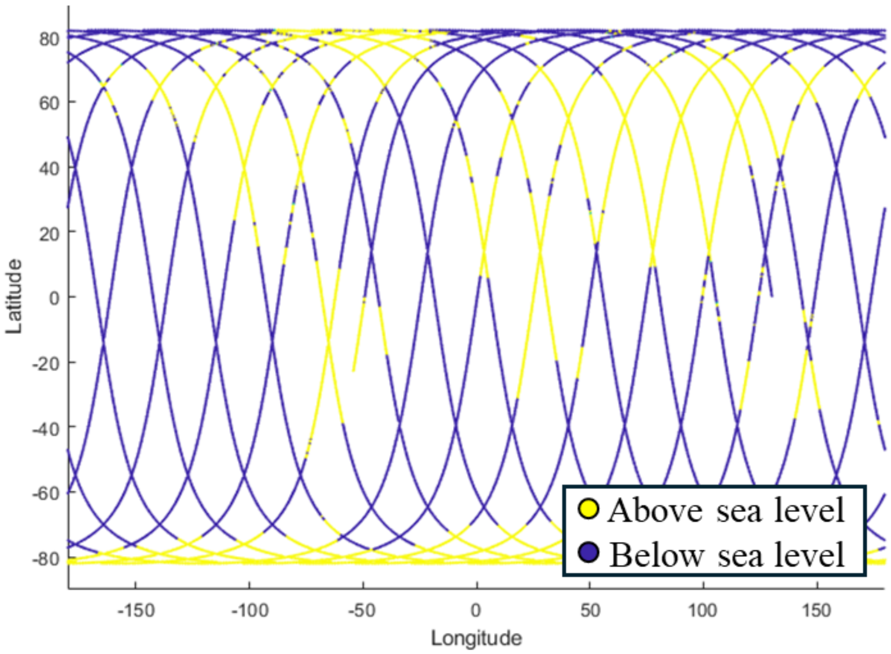


Figure 24 1-day simulation of nadir-pointed satellite, with points split into above or below sea level based on bathymetric data. The orbit is simulated at an altitude of 705 km with a 98.2-degree inclination. Distinguishing by ground elevation means the simulated satellite need not capture imagery over land.

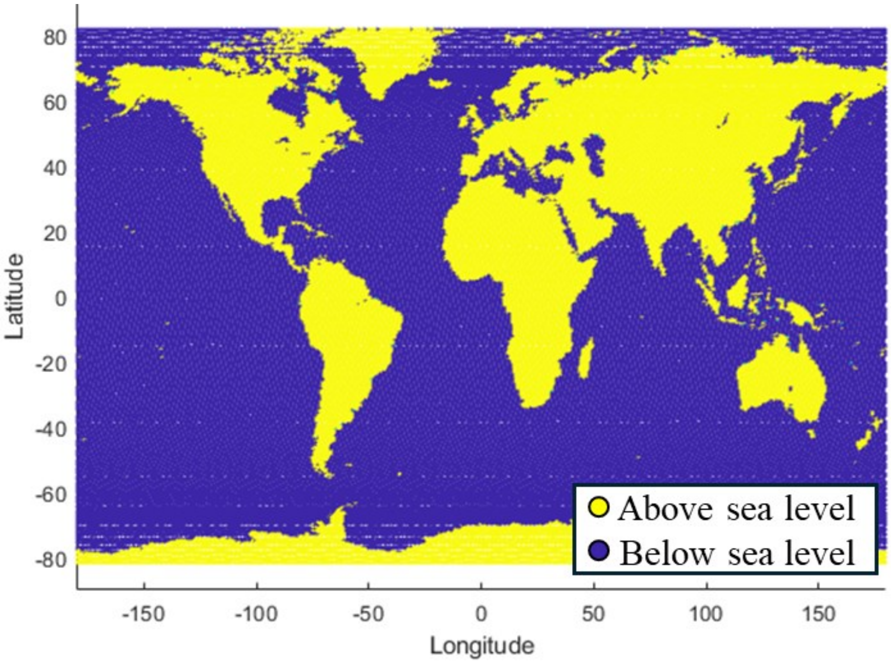


Figure 25 30-day simulation of nadir pointed satellite with points split into above or below sea level based on bathymetric data. This is an extended simulation from that of Figure 24. The orbit is simulated at an altitude of 705 km with a 98.2-degree inclination. Distinguishing by ground elevation means the simulated satellite need not capture imagery over land.

Table 10 Parameters for orbits simulated to determine valid imaging criteria when restrictions (Table 11) are applied.

orbit	**mean altitude (km)**	**eccentricity**	**inclination** (°)
Sentinel-2	786	0.001	98.2
ISS	420	0.001	51.6
low SSO	550	0.001	97.6
Sentinel-6A	1336	0.001	66.0

Table 11 Conditions applied to orbits to determine if satellite position meets valid imaging criteria. Conditions applied are cumulative from the top of table to the bottom.

description	**constraints**
eclipse	satellite obscured by Earth or Moon shadow
below sea level	less than 0 m
surface to deep sea fishing depths	0 m to - 1500 m
restricted latitude	70° N to 70° S
continental shelf depths	0 m to -250 m
region (Nova Scotia)	40° to 50° N, 53° to 73° W

The conditions in Table 11 were selected due to their relevance to the objective of vessel detection. Eclipse was selected due to the Earth not being directly lit during that period and the satellite not charging its solar cells. Below sea level was selected as ocean-going dark vessels cannot operate in areas with elevation above sea level (within the tolerance of the bathymetric elevations), and the range of deep-sea fishing depths due to these being areas where dark fishing vessels may operate. Continental shelf depth of 250 m was selected due to a variety of locations citing the depth between 200 m and 300 m, and the seabed rights of countries over them. Finally, the region of Nova Scotia was selected as an example of a region for monitoring as it is a peninsula surrounding by the ocean.

The swath and GSD of the nadir-pointing satellite will be considered constant for the purposes of providing context to the interference rate. Retaining the same GSD and swath between different satellite altitudes would require different camera focal lengths (increasing as altitude increases) but this will be ignored as it is secondary to the discussion objectives. Any considerations for atmospheric or illumination effects from capturing images at different altitudes are outside the scope of this research and not examined.

While real-time processing can be simply described as being capable of processing images as fast as they are captured, near-real time requires definition for the task. At some point of continually increasing the time to deliver actionable information, the information will begin to lose significance. For near-real time vessel detection, this would stem from a vessel moving too far to be intercepted from where an image was captured. A reasonable distance to consider would be distance to horizon, and the maximum viewable distance to an object over the horizon. As the height above the water from which observations are made on a vessel vary with a vessel's mass (affecting draught), and from where the observation is made (e.g. mast vs wheelhouse) several values are shown in Table 12. Comparing these distances against the speed of a vessel, and the time a detection takes to make can determine how useful a near-real time measurement is.

Table 12 Maximum observation distances for horizon and vessels over the horizon at different observation heights and air drafts. A vessel that moves less than 8 km from the time a detection is made, will be visible from the originally detected point.

observation height above sea level (m)	**distance to horizon (km)**	**maximum viewable distance for vessel of specified air draft over the horizon (km)**		
		10 m	**20 m**	**30 m**
5	7.98	19.27	23.95	27.53
10	11.29	22.58	27.25	30.84
20	15.96	27.25	31.92	35.52

Chapter 4 Presentation of Results

The results of the work described in the methodology (Chapter 3) will be presented in this chapter. This section will be broken up into two parts: model performance of YOLOv8 models trained using the constructed MS dataset (4.1), and the model inferencing speed and orbital performance (4.2). The model performance section presents the results of using different spectral bands, and model sizes, on model performance (F1-score, precision, and recall). This is accomplished by analyzing the results of different spectral bands permutations and noting consistent patterns in the results. Model inferencing speed and orbital performance examines the inferencing speed of the produced models onboard the simulated OBC configurations and compares it to the results of a satellite simulation to gauge real-time or near real-time onboard processing feasibility.

4.1 Model Performance

Due to the number of spectral band combinations (39 models), the results will be presented in groups. For instance, showing a relationship exists between quantity of spectral bands and model performance, does not necessitate including model performance pertaining to specific spectral bands, incrementally beyond RGB, unless it was relevant. The intent of the model grouping evaluated in this chapter serves to highlight incremental performance improvements of multispectral over RGB wavelengths, effect of model size on performance, value of specific spectral bands when added to the RGB colour channels, removal of specific bands from the Sentinel-2's 13 spectral bands on the model performance, and to compare bands based on different GSD combinations. An example of model output, with respect to inferencing images, is shown between Figure 26 (labels) and Figure 27 (predictions) for the large RGB model.

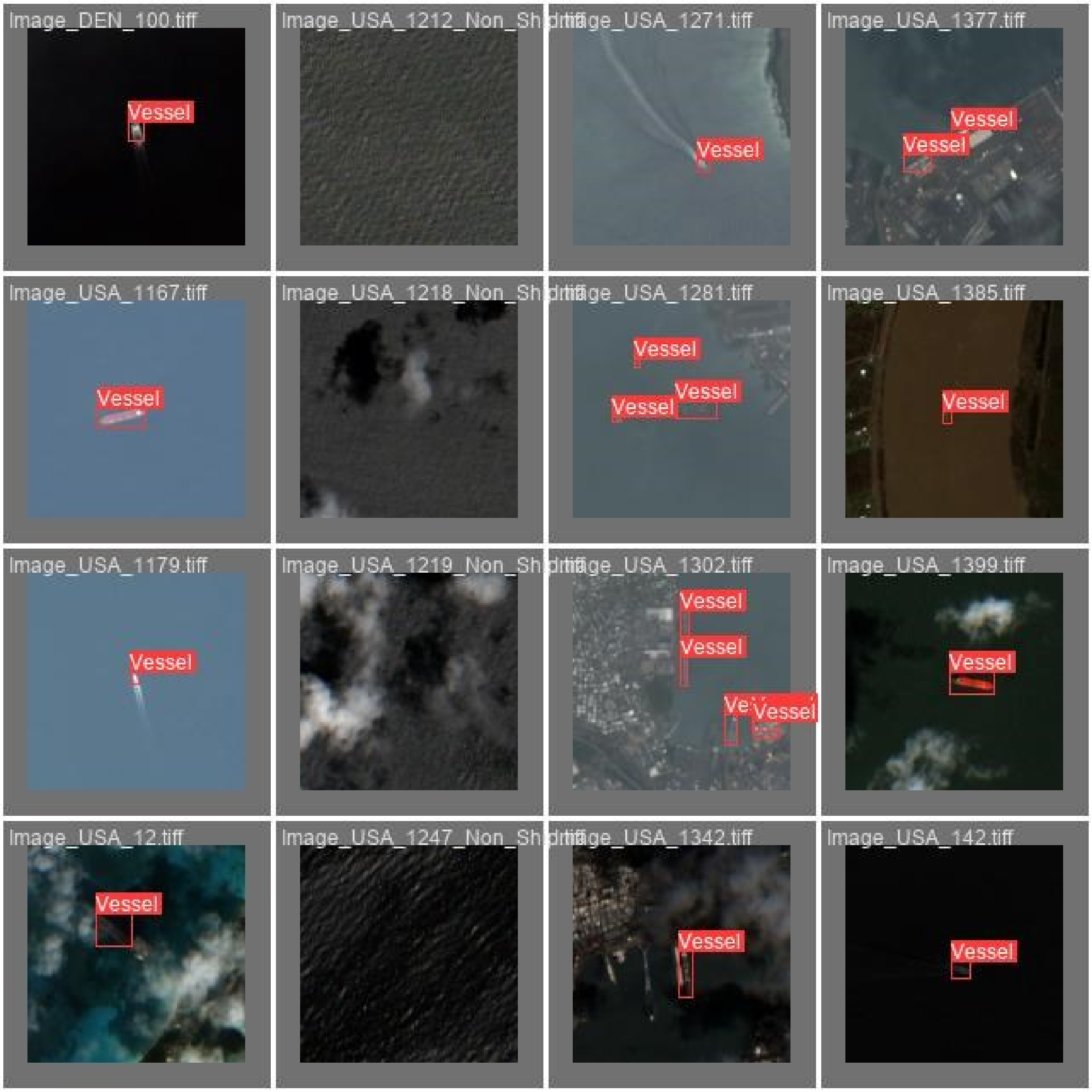


Figure 26 Example of 16 RGB composite images from the created dataset with labelled vessels locations shown. These images are the ground truth for the predictions shown in Figure 27.

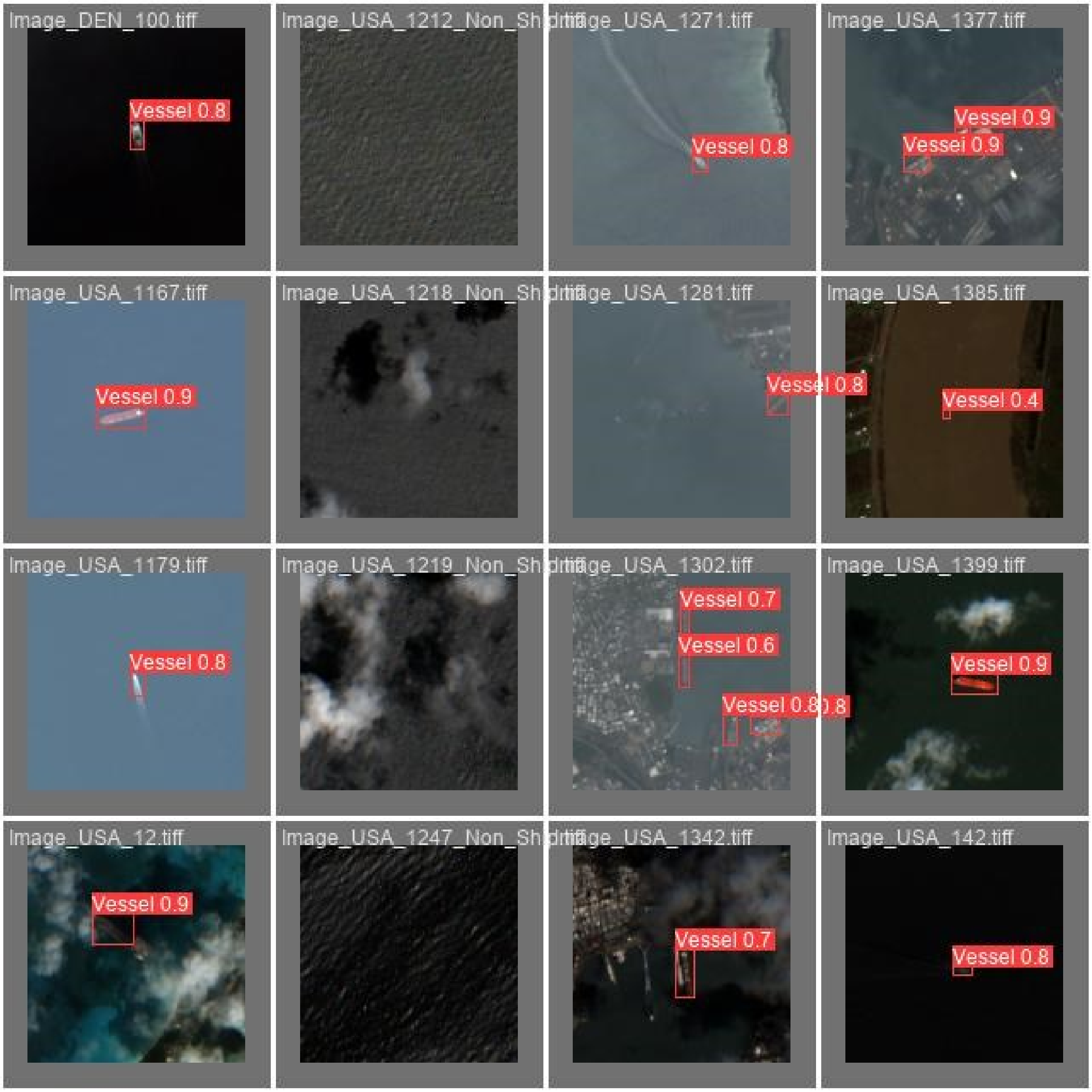


Figure 27 Example of 16 RGB composite images from the created dataset with inferenced vessels locations and prediction confidence shown (rounded to nearest tenth). The predictions were made using a fully trained RGB YOLOv8 model.

4.1.1 Multispectral vs RGB Impact on Vessel Detection

The first, and arguably most important question regarding multispectral vessel detection is if multispectral, as a sensing modality, offers any advantage over visible light imaging modalities. If no advantage can be offered, then there is little reason to pursue multispectral over conventional optical sensors. Figure 28 shows the F1 score of the AB models (Table 7) and the RGB models, at 5 different minimum confidence levels. Figure

29 shows the AUC of the F1-score curve for all five possible intervals starting from MAC of 0.25 (i.e. 0.25 to 0.6). There are four different versions of the AB and RGB models corresponding to the model size used. These models are noted by their lower case “l”, “m”, “s”, and “n” at the end of their respective abbreviation and correspond to “large”, “medium, “small”, and “nano” YOLOv8 model sizes.

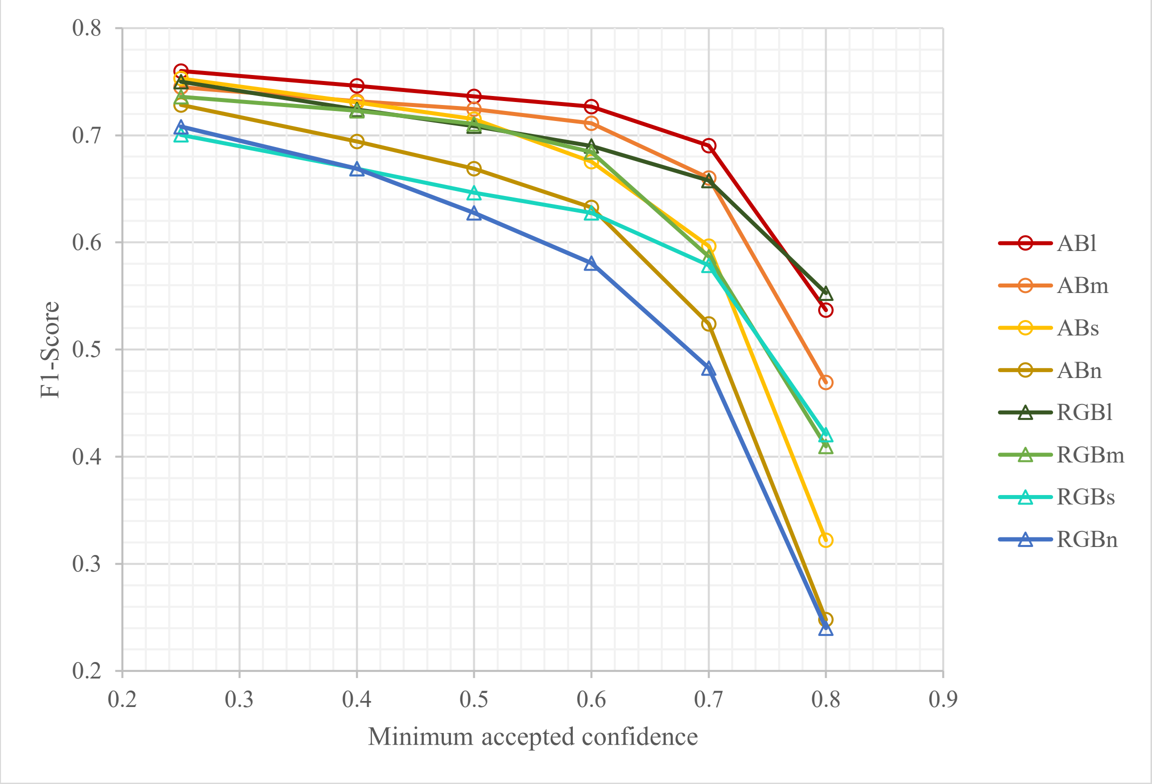


Figure 28 Comparison of F1-score at minimum accepted confidences for RGB vs AB vessel detection models of different sizes. AB models consistently outperform

RGB models at every model size, indicating there is value from the additional spectral information for ML vessel detection.

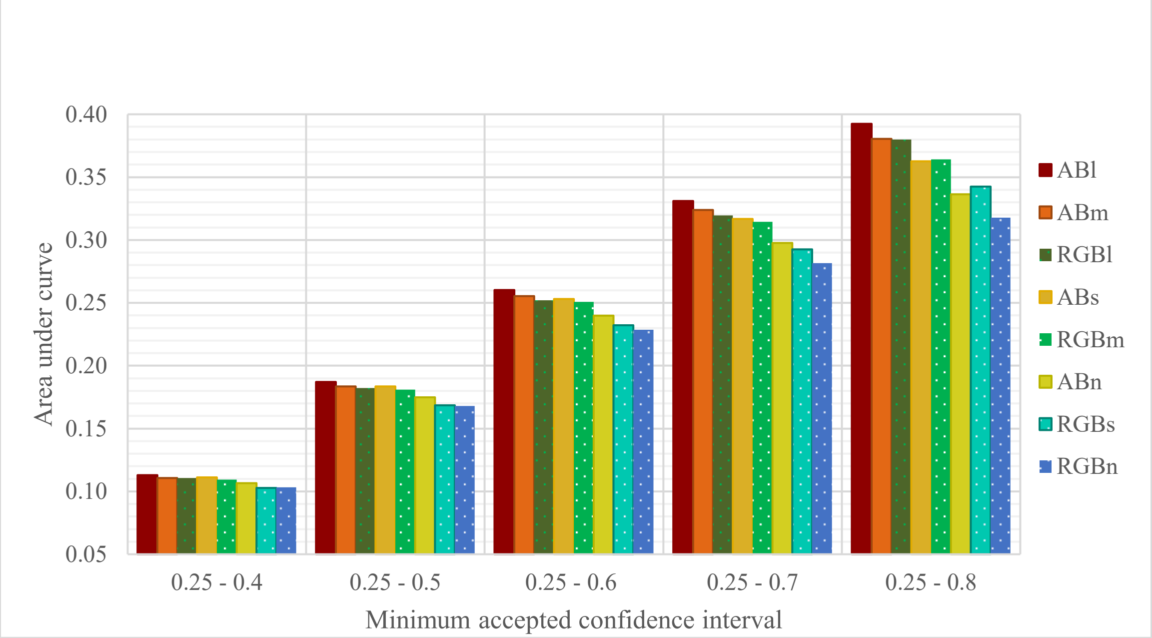


Figure 29 Comparison of AUC at MAC intervals for RGB and AB models of different sizes. Results match those in Figure 28, with the AB models consistently outperform RGB models at almost every model size.

For each model size (nano, small, medium, large), the AB models generally outperforms the RGB models of the same model size, with similar or higher F1-scores through most of the MAC intervals. For 0.7 MAC, or lower, every AB model outperforms the RGB model which is one size larger (e.g., the ABm medium model outperforms the RGBl model). It is only moving from 0.7 to 0.8 MAC that RGBl model overtake the smaller AB models, but this also corresponds with significant F1-score drops. This will be further examined with the precision and recall performance metrics (Figure 30 and Figure 31) to determine the cause of the drop.

- *It is concluded that multispectral imaging offers greater detection capabilities to sea level vessel detection beyond RGB wavelengths as implemented in an ML model.*

To investigate this result further, the recall and precision model metrics can be similarly compared across the ML models. Figure 30 shows the results of the precision versus

MAC for the trained RGB vs AB models, while Figure 31 shows the results for recall versus MAC.

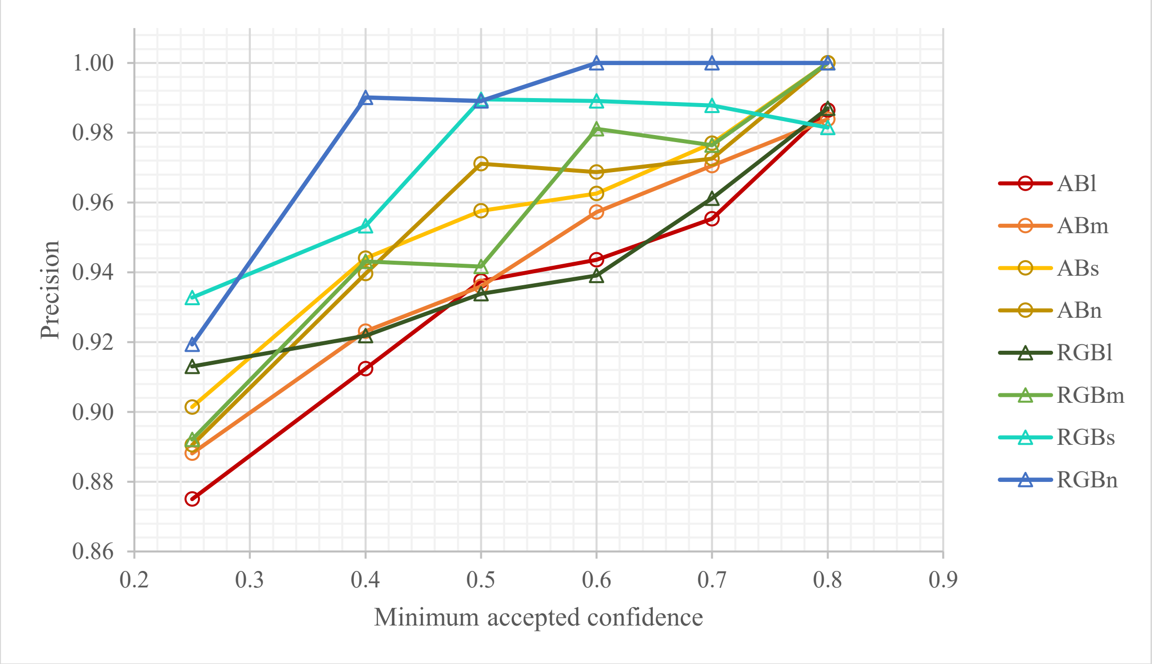


Figure 30 Comparison of precision values at minimum accepted confidences for RGB vs All-band vessel detection models of different sizes. Model size increases correlate with a smaller performance difference (improvement) between RGB and AB models. Precision reduces with additional spectral information until a certain model size is reached.

The precision metric, across models, (Figure 30) is less conclusive than the F1-score (Figure 28). Only the RGBn and RGBs models consistently outperform their respective counterparts (ABn and ABs). The RGBl and RGBm models perform slightly better than the ABl and ABm models in spots, (e.g. at 0.4 and 0.5 minimum confidence) otherwise they do not greatly exceed them. This suggests that for the YOLOv8 model, the additional spectral information may cause a drop in precision until a certain model size is reached. With no notable improvement to precision – if not the opposite – the benefits for a multispectral treatment must come from recall given the F1-score (Eq. (4)) is the geometric mean of precision and recall. The recall of the models at different MACs are shown in Figure 31.

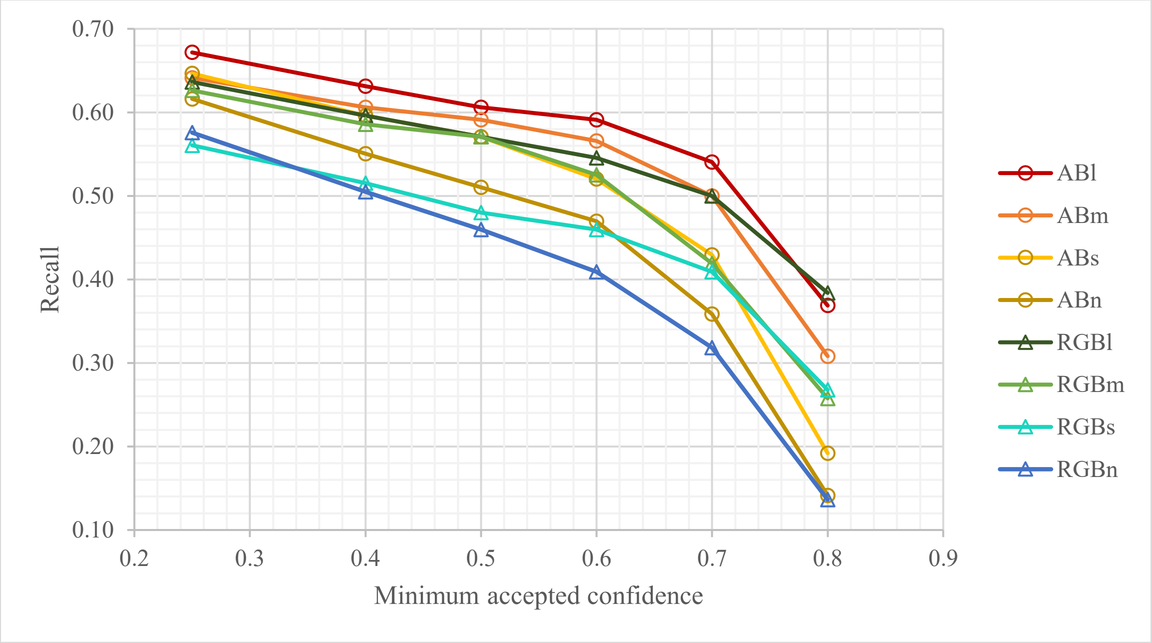


Figure 31 Comparison of recall values at minimum accepted confidences for RGB vs All-band vessel detection models of different sizes. Recall for AB models consistently outperforms RGB models, indicating additional spectral information improves capability of detecting vessels which are present in images.

Recall offers a clearer picture of the benefit of MS imaging for ML vessel detection. Each AB model outperforms the same-sized RGB, detecting anywhere from 5% to 10% more vessels present in the image depending on the targeted minimum accepted confidence. Combined with the results of recall-confidence curve Figure 30, the drop between this means that for YOLOv8 models, the strength of multispectral imaging for vessel detection appears to stem from the ability of multispectral models to more accurately detect vessels present within the image and not the reduction of false positives (i.e. when the background is predicted as vessels) within the inferenced images. More succinctly,

- *Using additional spectral information during improves the capability of ML models to detect vessels present within an image.*

The advantage of multispectral over RGB imaging is the additional spectral information captured for each pixel, resulting in the models more accurately recognizing patterns between spectral channels and physical features. For instance, red-edge bands are frequently used to detection vegetation due to chlorophyll reflected in these bands.

Inclusion of these spectral bands facilitates the model to better train on relationships between red-edge or NIR light and vegetation. Vegetation would only be co-incident with vessels not at port if there were heavy algae blooms present. An example of this can be observed in Figure 32. Similarly, other relations between different spectral bands and viewed objects may help the model characterize other features unique to vessels and the surrounding non-vessel backgrounds.

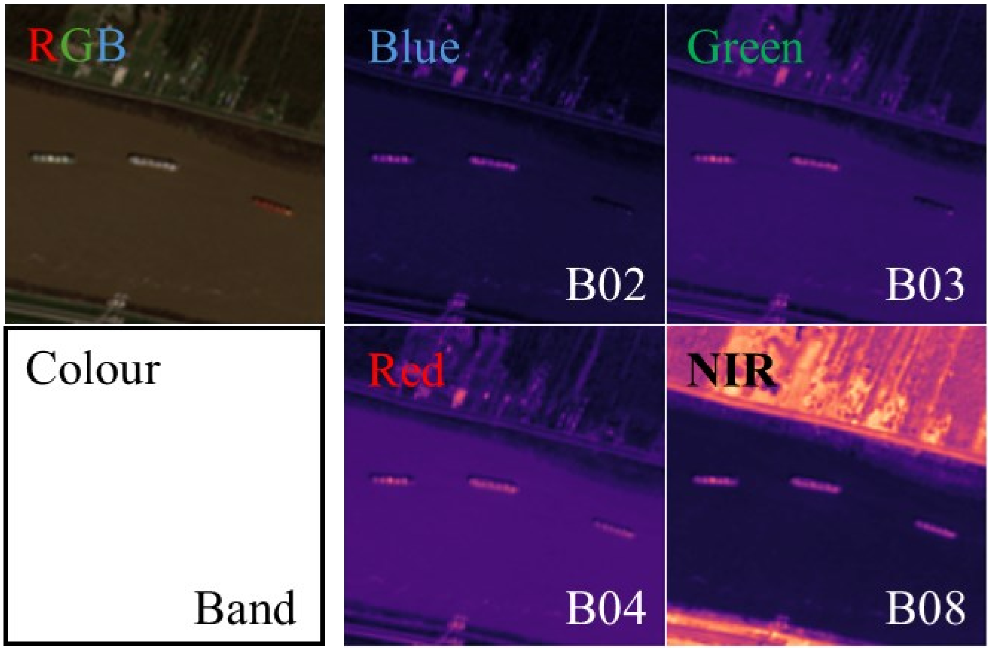


Figure 32 RGB composite image compared against non-normalized RGB bands and B08: a NIR band. Due to vegetation reflecting NIR light, vegetation on the shore has a high response. Additionally, due to water's high absorption of NIR light, the vessels appear in higher contrast in B08 compared to the RGB channels.

The results in Figure 30 and Figure 31 indicate that the cause of the performance drop when moving from 0.7 to 0.8 MAC stems from a low number of high confidence predictions (at or above 0.8 confidence). Given that precision also increases during this period, this is expected behavior for the YOLOv8 model. Evaluating the performance at different MACs better characterizes the nature of the predictions made. For instance, a spectral band combination that only results in improved performance at high confidence likely means that existing predictions are made with improved confidence, whereas only performance improvements at low confidence likely means that more predictions in total are being made. However, unless precision is the only metric to be considered, the drop in F1-score from 0.7 to 0.8 MAC also indicates that the models' performance for general use begins to rapidly fall off. The difference in F1-score between true positive vessel detections at 0.7 and 0.8 MAC is similar from 0.25 to 0.7 MAC for RGB and AB models. To simplify analysis, and produce a singular value for each curve, results from 0.25 to 0.7 MAC are used to produce the AUC, not all combinations of MAC intervals such as in Figure 29.

4.1.2 Model Size Impact on Performance

Model size, which increases with model complexity and number of model parameters (feature vector), impacts the model performance and inferencing speed. As shown in section 4.1.1 both the RGB and AB model have large, medium, small and nano models, but there was an additional model trained with four sizes: G12 which are models which utilize all Sentinel-2 bands with GSDs of 10 m and 20 m but exclude the three bands with 60 m GSD. The F1-score of the D12m model is shown in Figure 33 and the AUC in Figure 34, which – along with Figure 28 – show a correlation between F1-score and model size. As a point of comparison, the AB model results are included in the image. The RGB models were not included due to them affecting the clarity of Figure 33, but can be observed in Figure 28.

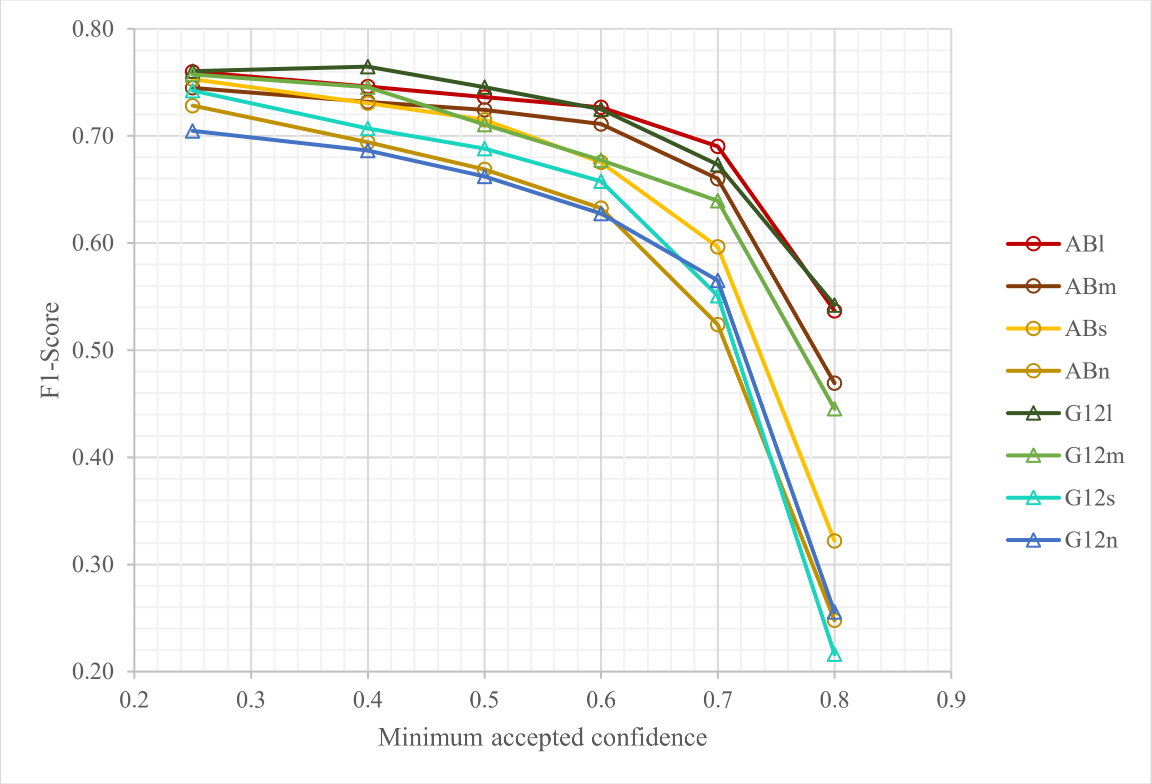


Figure 33 Comparison of F1-scores at minimum accepted confidences for AB and G12 models of different sizes. Differences between same size models are less pronounced than those in Figure 28, but still indicate a general positive correlation between number of bands.

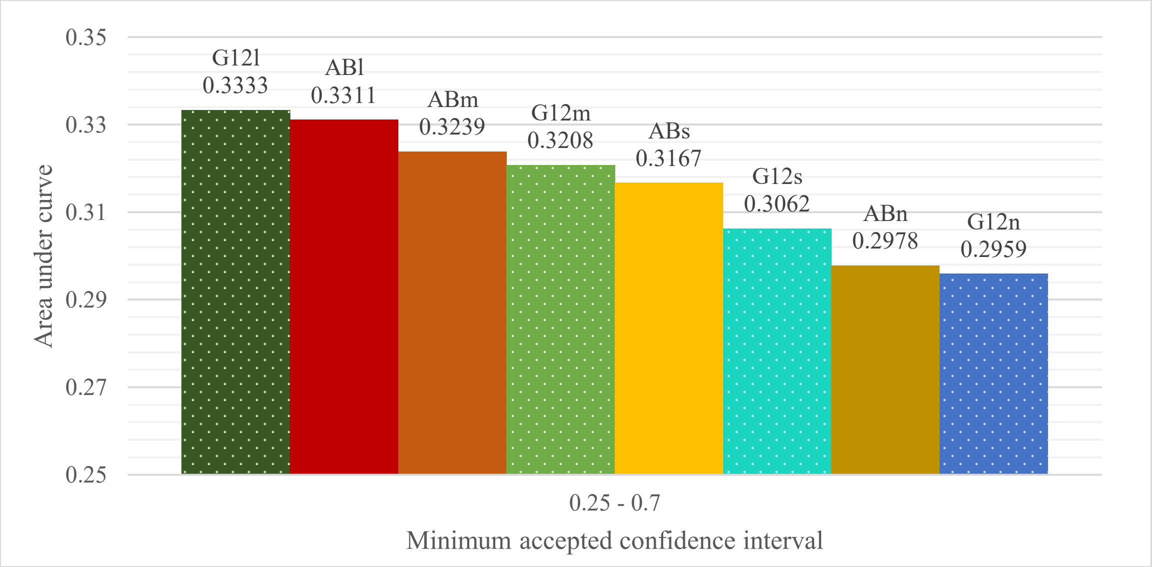


Figure 34 Comparison of AUC at MAC intervals for AB and G12 models of different sizes. Differences between AUC for each model are smaller than those in

Figure 29 for AB and RGB, indicating a general positive correlation between the number of bands.

Since the information offered by comparing model performance in both Figure 28 and Figure 33, along with the AUC in Figure 29 and Figure 34, only reaffirms the initial conclusion, it is worth analyzing model size performance in a different way. Figure 35 displays the difference in F1-score between different model sizes as a stacked group bar chart. From this, it is apparent that as the MAC increases so does the impact of the model size. Additionally, the largest improvement occurs in the progression from the nano to the small models. The improvements from small to medium, and medium to large model sizes are each less pronounced.

From Figure 35, the decision was made to use the medium-sized models to analyze the remainder of the spectral band combinations. Medium model sizes offer a notable improvement over smaller model through their lower MAC thresholds, while large models through this region only offer minor improvements. It was only at high MACs of 0.7 or higher that the large model offered notable improvements over the medium model, but at those points the models' F1-scores drops off.

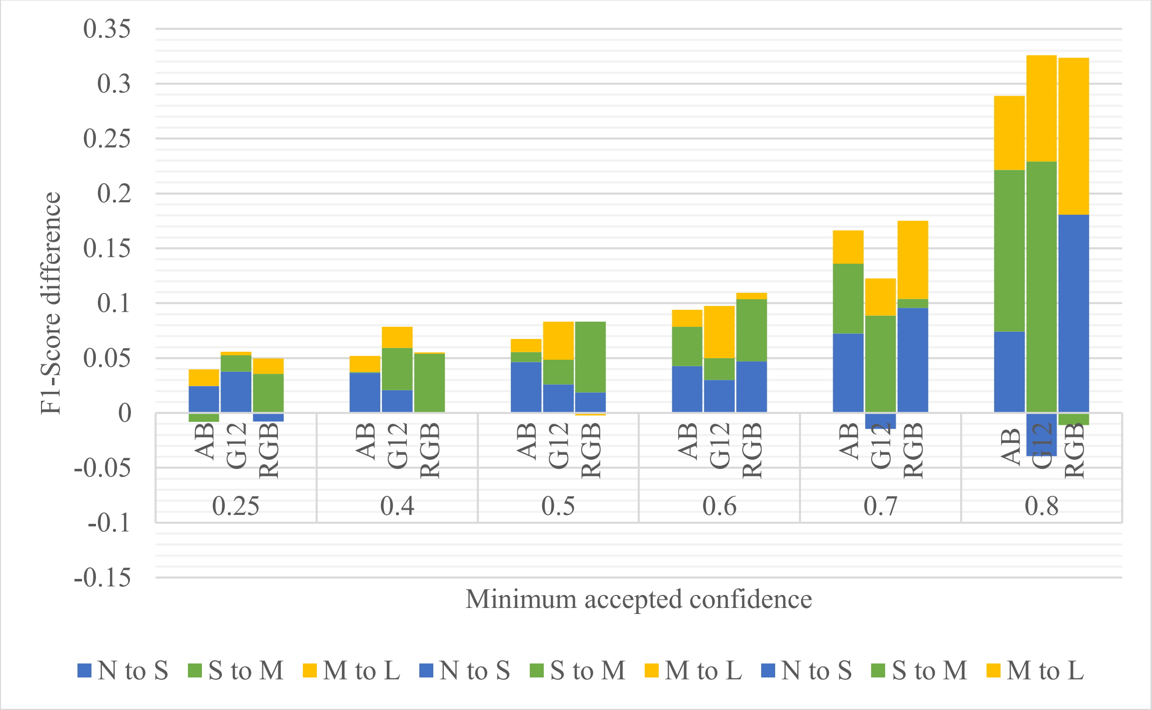


Figure 35 F1-Score difference between different model sizes for RGB, AB, and G12. Models experience diminishing returns as model size increased, and model size has a larger impact on F1-score at higher minimum accepted confidence.

The F1-score improvements from increasing model size are around 0.05 - 0.3 when model size is increased from nano to large at different MAC. Similarly, the effect of selecting a model has a larger impact on model performance with larger improvements.

The difference between AUC for RGB, AB, and G12 models also helps indicate what a significant change of AUC is. Given the parameter difference between model sizes (Table 8), a same size model's performance difference can be considered significant if the AUC difference matches those of a model size increase or decrease. From the values shown in Figure 36, changing model size resulted in an AUC change of 0.008 to 0.014. An increase of AUC within or beyond this range can be considered equivalent to the performance improvement gained by increasing model size for vessel detection.

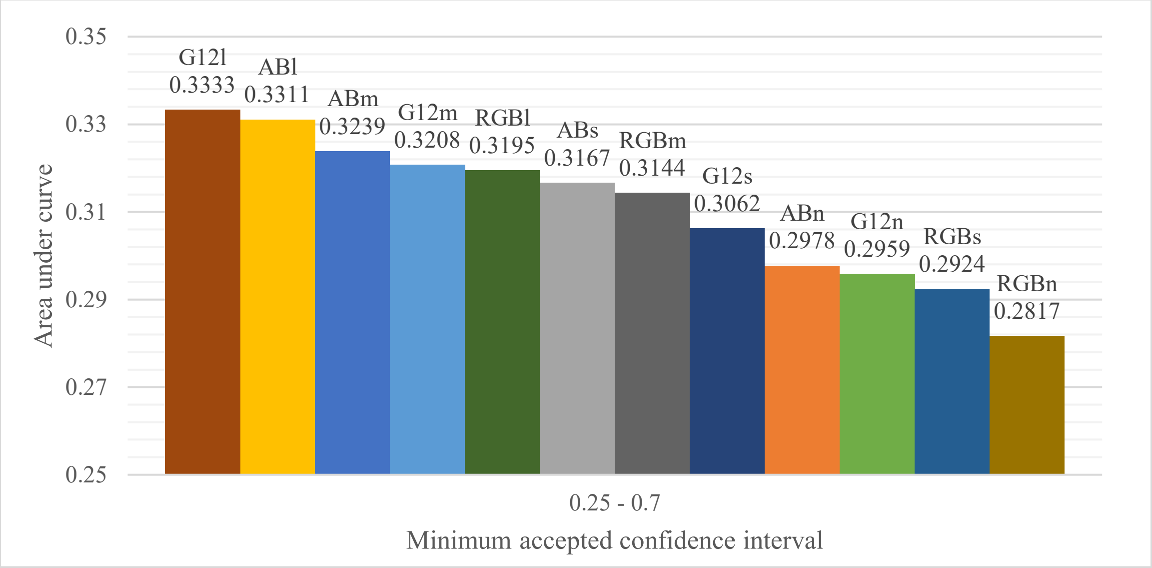


Figure 36 Comparison of AUC at MAC intervals for AB, G12, and RGB models of different sizes. Changes in model sized, averaged across each of the three band combinations, results in a 0.008 to 0.014 change in AUC, which can be taken as a metric for significant improvement in a model.

4.1.3 Value of Additional Spectral Band to RGB Model

A natural question is whether there were spectral bands that when incrementally added to an RGB optical system would significantly improve vessel detection performance. This is relevant to size and for insight into available imaging systems for small satellites. There are MS imaging systems that feature RGB plus a single NIR spectral band. Of the 13 Sentinel-2 bands, 3 are RGB. The addition of a single additional spectral band yields 9 possible combinations to trial. For legibility, the resulting model results are divided into two plots: the first (Figure 37) features RGB plus a 20 m GSD spectral band or the single remaining 10 m GSD spectral band (Figure 39) which features RGB plus a single 60 m GSD spectral band.

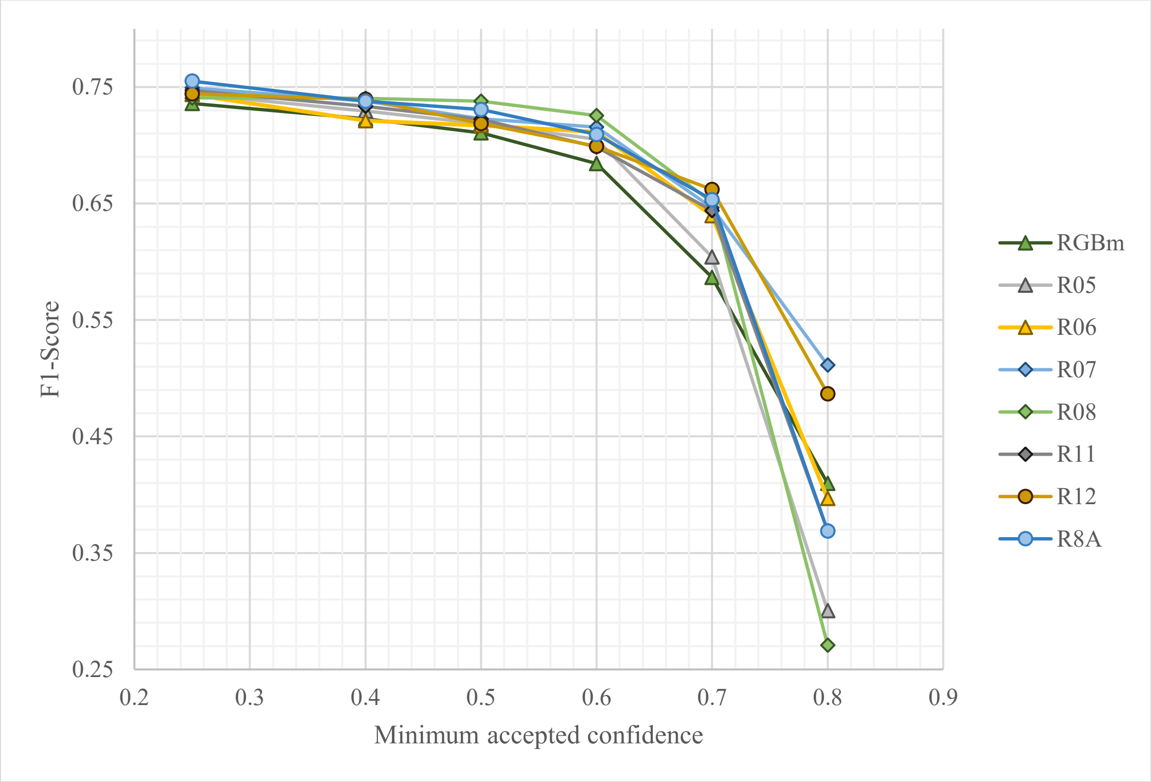


Figure 37 Comparison of F1-scores at minimum accepted confidences for models trained on RGB spectral bands and one additional 10 m or 20 m spectral band. The addition improves the ML model performance.

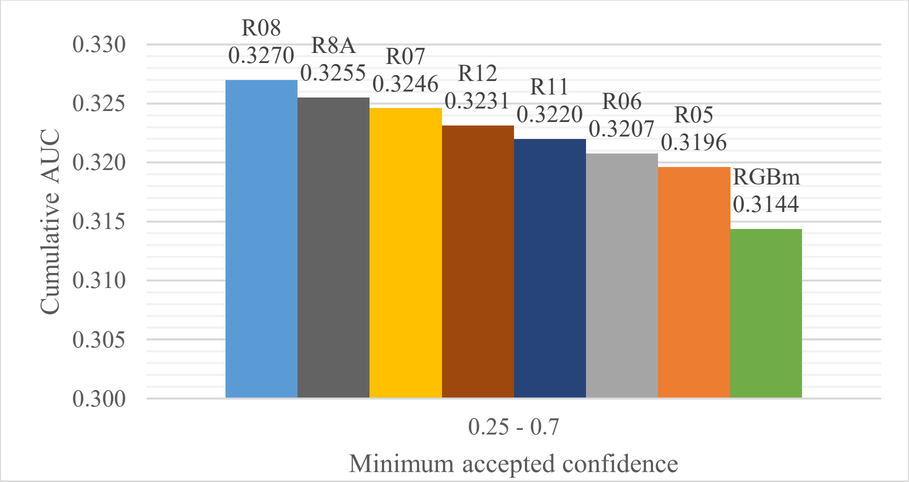


Figure 38 Comparison of AUC at MAC intervals for models trained on RGB spectral bands and one additional 10 m or 20 m spectral band. All models with added spectral bands outperform RGB.

Of the band combinations presented in Figure 37, at most points adding a 10 m or 20 m GSD spectral band improves model performance. The exception for this occurs at 0.8 minimum confidence, where most combinations fall below the RGB value, except for the models R07, and R12. The improvement to F1-score ranges from 0 to 0.08 from 0.25 to 0.7 minimum confidence, and -0.14 to 0.10 at 0.8 minimum confidence. Considering Figure 38, all new models outperformed RGBm. Only R06 and R05 failed to exceed an AUC improvement of 0.008 over RGBm.

- *The inclusion of any 10 or 20 m GSD spectral band with RGB bands has a positive impact on the ML model performance.*

Of the models trained, R07, R08, and R12 are notable for having the largest performance improvements or regressions over the RGBm model at specific MACs. R08, which is RGB plus B08 – broad NIR – is likely performant due to R08 being a 10 m GSD band instead of 20 m. Smaller and finer features are capable of being captured, but B08 has the widest spectral width of all non-SWIR spectral bands on Sentinel-2 at 105 nm. By AUC it has the best performance but at 0.8 MAC specifically, it has the worst F1-score. While the lower GSD of 10 m ensures that more detail is captured, it is possible that it inserts uncertainty into predictions due to additional features that are not visible in other bands due to its bandwidth and GSD. Theoretically the model should account for this, but it is possible that it is identifying patterns and features that do not correlate and is having difficulty rejecting them confidently from the training process. This is not improbable since at lower minimum confidences, R08 is the highest performance model, while the reverse is true at 0.8 MAC.

There are less obvious indications of why R07 and R12 performs as they do; both are two of six 20 m spectral bands, that observe different bandwidths at 20 nm and 180 nm bandwidths respectively. The spectral function document for the Sentinel-2 MSI imager notes the purpose of these bands. B07 is noted as being used for leaf area index, a dimensionless estimating leaf area over land, and B12 the assessment of mediterranean vegetation conditions, distinction of clay soils for the monitoring of soil erosion and burn scar mapping. It is possible these relationships, along with others uncovered through research, trained the model to better discern vessels from its background with positive and negative correlations.

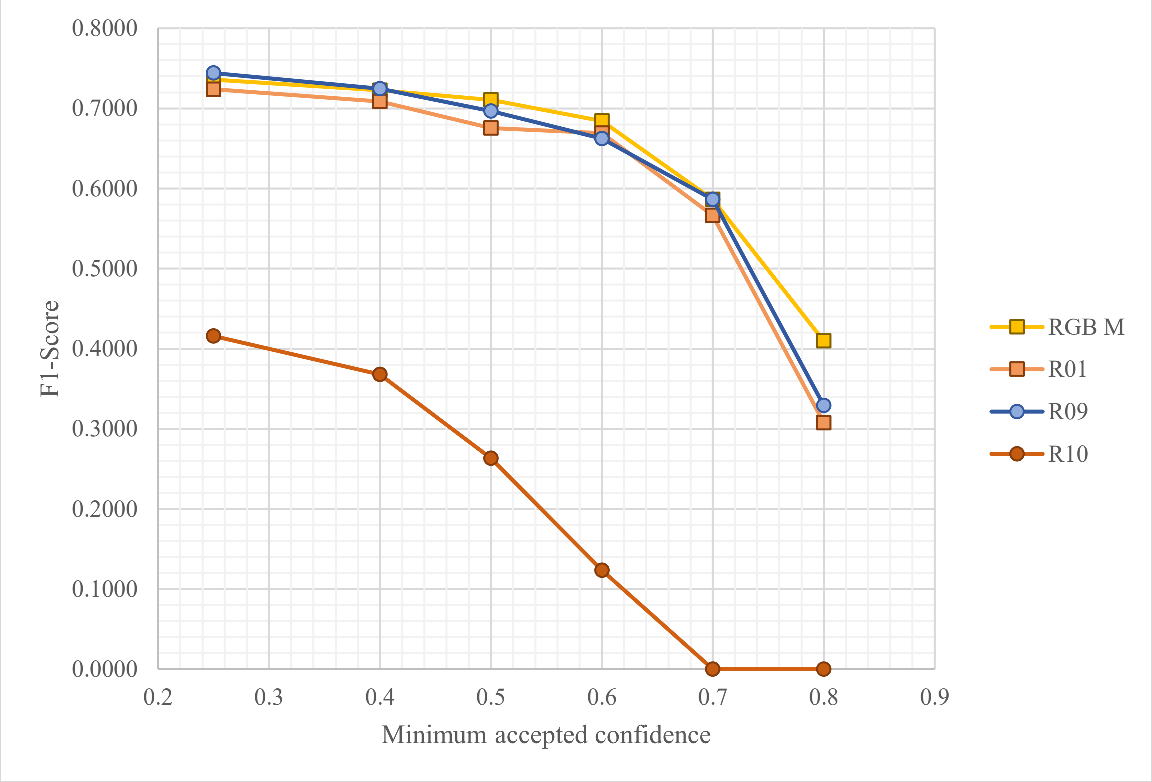


Figure 39 F1-score vs minimum accepted confidence for RGB spectral bands and one additional 60m spectral band. The inclusion of 60 m spectral bands made no

improvement on the F1-score of the trained ML models and the R10 model experiences the worst performance with B10.

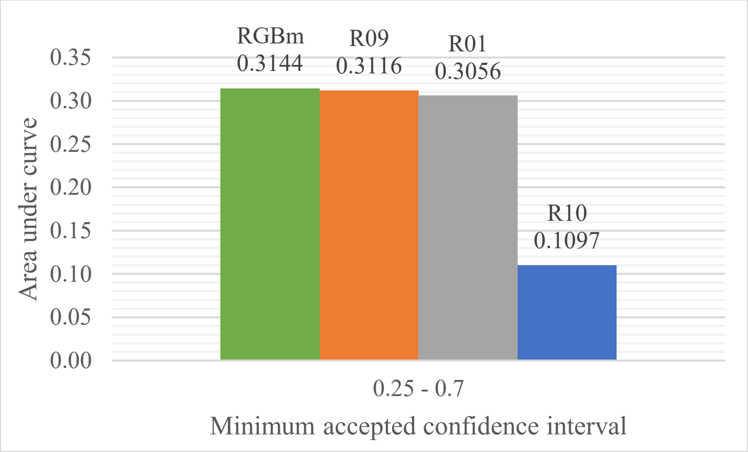


Figure 40 Comparison of AUC at MAC intervals for models trained on RGB spectral bands and one additional 60 m spectral band. The resultant models see minor to significant performance loss compared to RGBm.

Unlike adding a 10 m or 20 m spectral bands to the RGB spectral model for training, utilizing the 60 m spectral bands has at best a negligible to negative effect on the results. In Figure 39 and Figure 40, models R01 and R09 have slightly worse performance than RGBm, but model R10 has dismal performance: the worst of all models tested in this thesis. R10 was also trained an additional two times to ensure there was not an error in this process. R10 was trained on RGB and B10, the later of which is a SWIR band that was selected for Sentinel-2 due to its near complete reflectance with water vapour – particularly that of cirrus clouds – within the atmosphere (see Figure 22). B10 cannot meaningfully image features below 2000m, limiting its use to detecting high altitude features such as clouds. A result of this can be observed in Figure 21, where B10 is dramatically different than any other band. This means the poor performance of model R10 was due to it being unable to find significant value in identifying high altitude features or being able to meaningfully reject B10 during training.

- *The additional of any Sentinel-2 60 m spectral band to RGB bands has minimal (and sometimes detrimental) impact on trained ML model performance.*

From Figure 37 and Figure 39 it is apparent that while additional spectral information of similar GSD improves the vessel detection performance, certain bands – particularly low spatial resolution bands – offer negligible or negative changes to the model's performance trained using YOLOv8 default training parameters. Theoretically, a perfect model should perform no worse than utilizing RGB if the additional spectral band truly added nothing. However, there are reasons why this may not occur. Tuning training parameters (e.g., learning rate, momentum, and weight decay) that help a model progress towards its trained state may improve final model performance but introduces too many model variations to evaluate and falls outside the thesis scope. Secondly, there may be an inconsistent relationship between certain spectral bands and features present in the images within the dataset. This would be a deeper problem with the dataset that is discussed further in Future Work (section 5.1)

4.1.4 Spectral Band Subtraction

Another group of models that warranted investigation was the performance of utilizing all Sentinel-2 spectral bands but with 1 spectral band excluded. The goal with these combinations were to try and discern if there were any specific spectral bands that have a positive, negligible, or negative impact on the model when absent while using the default training process. When evaluating the models for performance improvements, no change in model performance with spectral band removal was considered a slight improvement, as resultant models have faster inferencing speeds. For evaluating performance improvements, no change in model performance with spectral band removal created a slight improvement, as resultant models have faster inferencing speeds. The results from the 13 band combinations can be seen in the following figures, and were split into groups based upon their GSD of 10 m, 20 m, or 60 m.

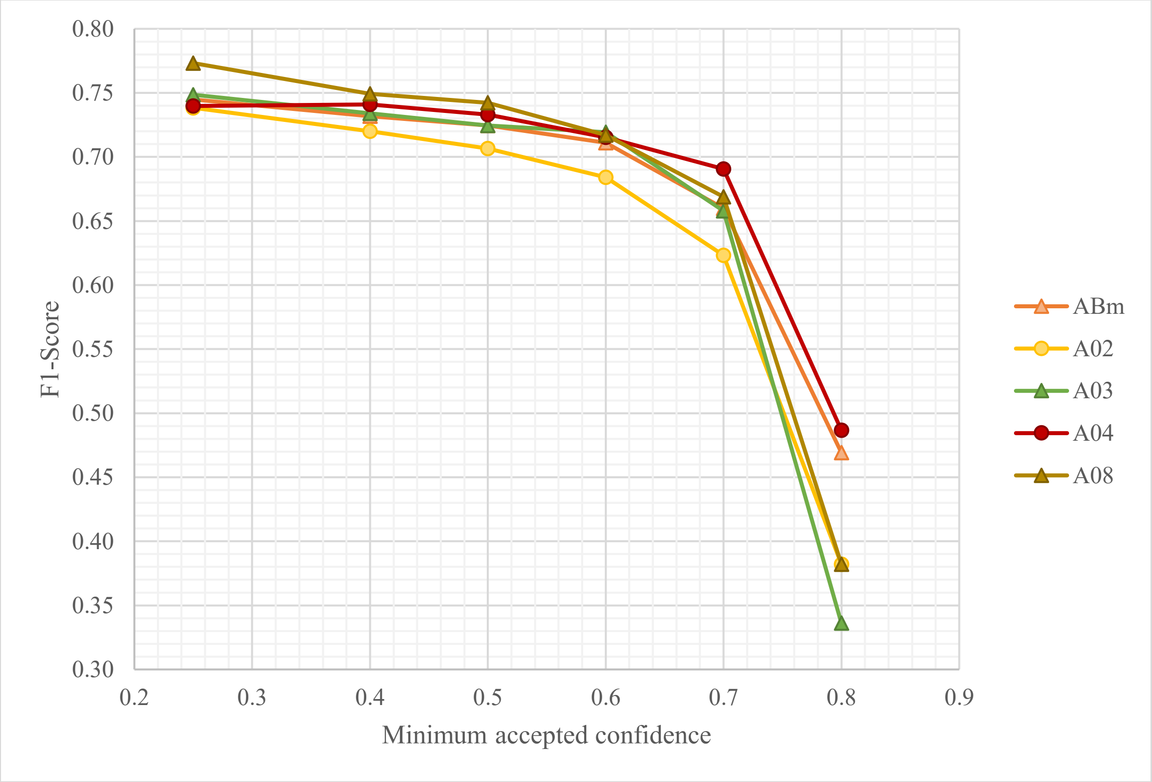


Figure 41 Comparison of F1 scores at minimum accepted confidences for models trained with combinations of 10 m GSD spectral band subtraction. Model A02 – removal of B02 – suffers distinct performance loss, while remainder of subtraction models experience slight negative or negligible performance changes.

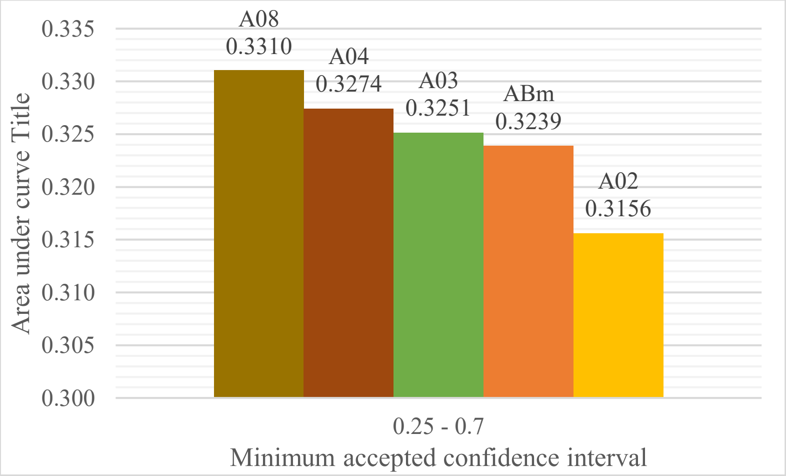


Figure 42 Comparison of AUC at MAC intervals for models trained with combinations of 10m GSD spectral band subtraction. Removing B02 sees significant

performance loss, while B08 removal is slightly beneath the threshold for performance improvement.

Figure 41 and Figure 42 shows the results of dropping different 10 m GSD spectral bands from the 13 spectral bands of Sentinel-2 prior to training. From the results, removing B02 (blue), has a distinct negative effect on performance. This is both a 10 m band, and at first thought the primary visible colour of water, so the result is not surprising. The ESA makes use of B02 in the construction of scene classification maps using algorithms included with level 2A Sentinel-2 imagery, particularly pertaining to snow, water, bare dirt, and clouds [41]. At the peak there is a 0.05 drop in F1-score compared to the ABm model, and a notable 0.0083 drop in AUC score. This is the most significant changes of any of the combinations. A08 experiences an improvement at low MACs but falls just short of the AUC threshold at 0.0071, but virtually matches the AUC of the ABl model. B08 overlaps spectral bands B07 and B8A, which may introduce over representation of the specific spectral bandwidths, in addition to B08's very large bandwidth. This may be the cause in the slight performance improvement when B8A is removed in A08A (Figure 46).

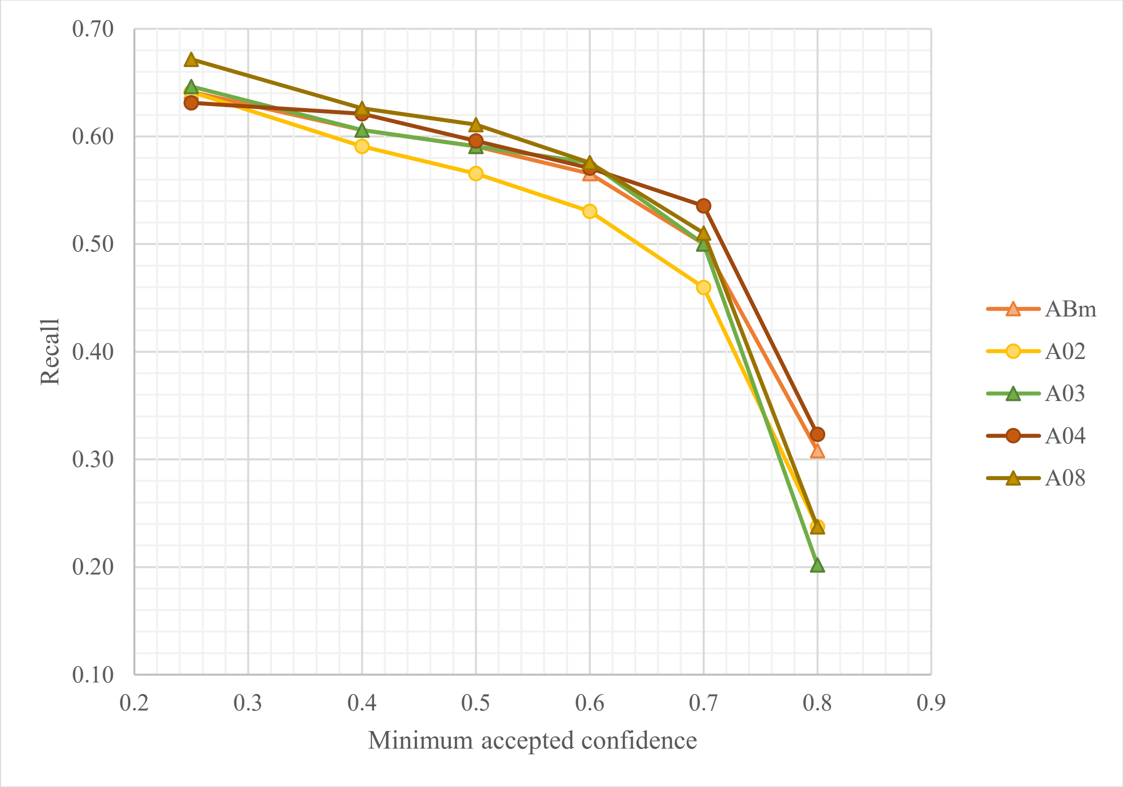


Figure 43 Comparison of recall at minimum accepted confidences, for models trained with combinations of 10 m GSD spectral band subtraction. This resultant curve and performance order closely resembles those shown in Figure 41, indicating recall is a driving factor for F1-score differences between band combinations.

Recall at different MACs (Figure 43), resembles the plot created from F1-score at different MACs (Figure 41). In fact, there are no particularly new observations that can be made. Recall performs worse with the removal of B02 (blue) and best with the removal of B08.

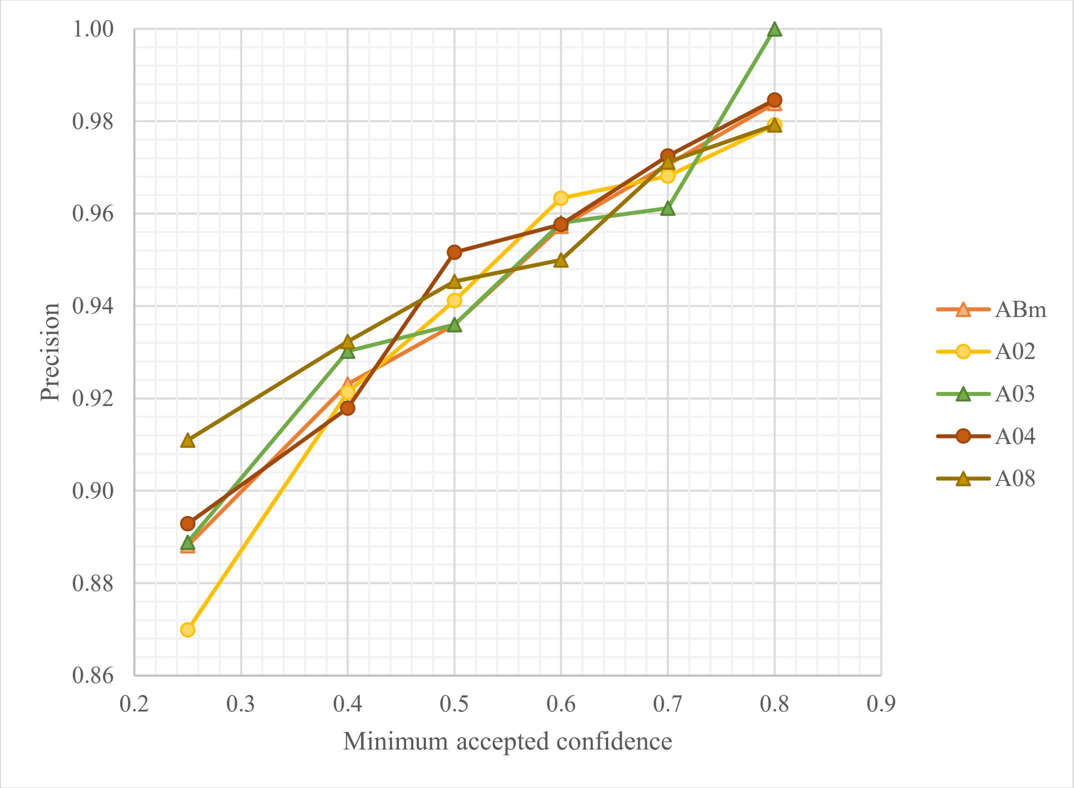


Figure 44 Comparison of precisions at minimum accepted confidences, for models trained with combinations of 10 m GSD spectral band subtraction. Precision for spectral band subtraction is not notably affected by exclusion of any single spectral band.

As shown in Figure 44, precision does not measurably favor one spectral band combination over the other. While a larger difference occurs with the subtraction of B08 and B02 at low confidence, these differences do not propagate as the MAC is increased. In fact, except for ABm which only narrowly misses out, each of the other 5 combinations shown have the best performance for at least one of the 6 MAC values.

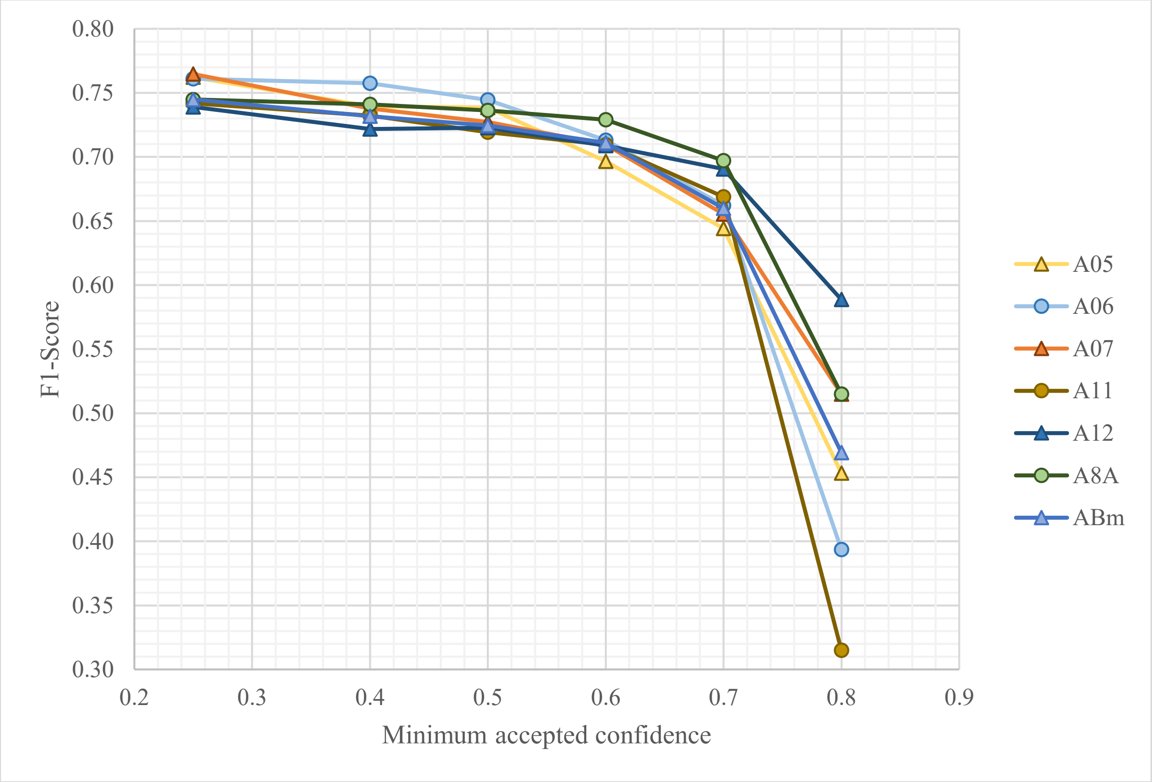


Figure 45 Comparison of F1 scores at minimum accepted confidences for models trained with combinations of 20 m GSD spectral band subtraction. The results indicate not all bands have the same or even positive value in the model performance.

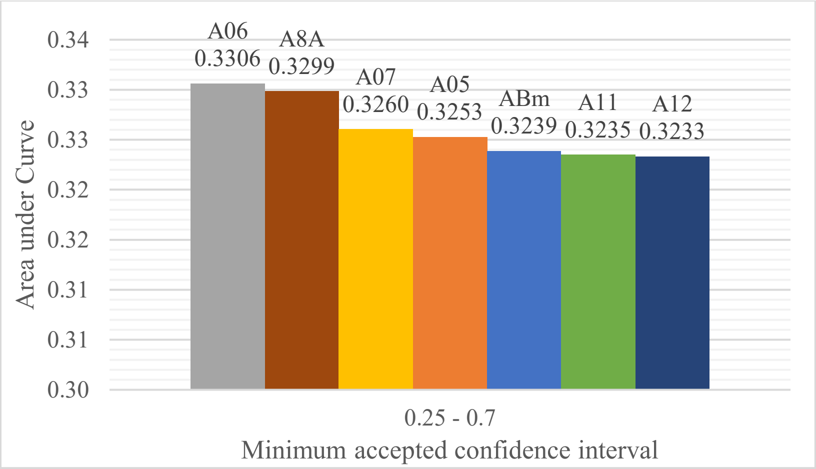


Figure 46 Comparison of AUC at MAC intervals for models trained with combinations of 20m GSD spectral band subtraction. The removal of B06 and B8A results in models with minor performance increases over ABm

The F1-score curve for removing 20 m GSD spectral bands (Figure 45) yield a mixture of positive or neutral results, with A8A and A06 being notable standouts. Both fall short of the AUC threshold for significant improvement (0.0067 and 0.006 respectively) but are still minor improvements (Figure 46). One notable result for specific MAC score is the improvement of the model's performance for 0.25 to 0.5 minimum confidence when B06 (red edge) is dropped, where the largest difference in F1- score is approximately 0.05. Another one is with the elimination of B8A, the model sees a slight performance improvement through an F1-score increase of 0.03. However, arguably the most interesting performance result in a model is the result of removing B12. At 0.8 minimum confidence, this model has over 0.55 F1-score, a feat that not even ABl or RGBl models could match, while performing similarly for the remainder of the MAC points.

The removal of B8A (model A8A) is interesting in concert with the results from 10 m GSD spectral band subtraction (Figure 41) because it supports the hypothesis that *the degree of data redundancy may be a problem* in models where B07, B08, and B8A are present. Compared to one another other the removal of B08 (model A08) offers better performance at 0.6 minimum confidence or lower, while the removal of B8A (model A8A) increases performance at or above 0.6 minimum confidence. However, both offer equal or better performance to the ABm model and are the highest AUC models for 10 and 20 m spectral band subtraction respectively.

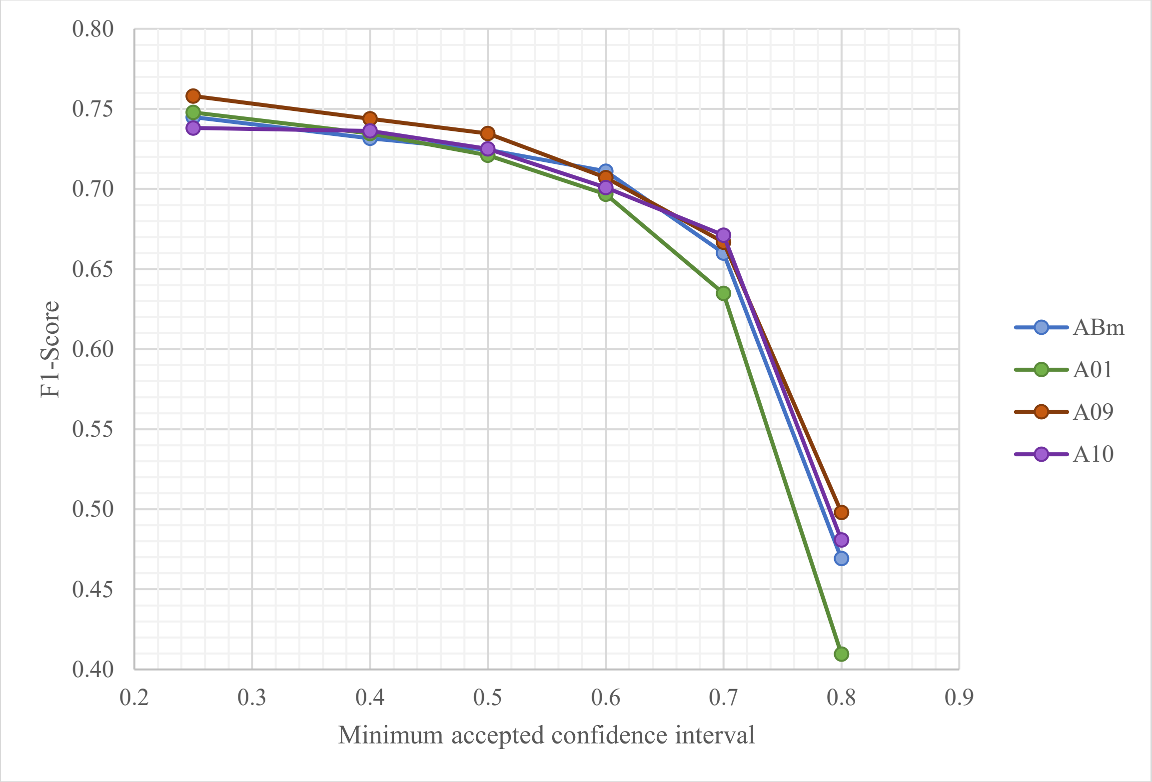


Figure 47 Comparison of F1 scores at minimum accepted confidences for models trained with combinations of 60 m GSD band subtraction. 60 m GSD spectral band removal does not show any consistent patterns of performance improvement or loss.

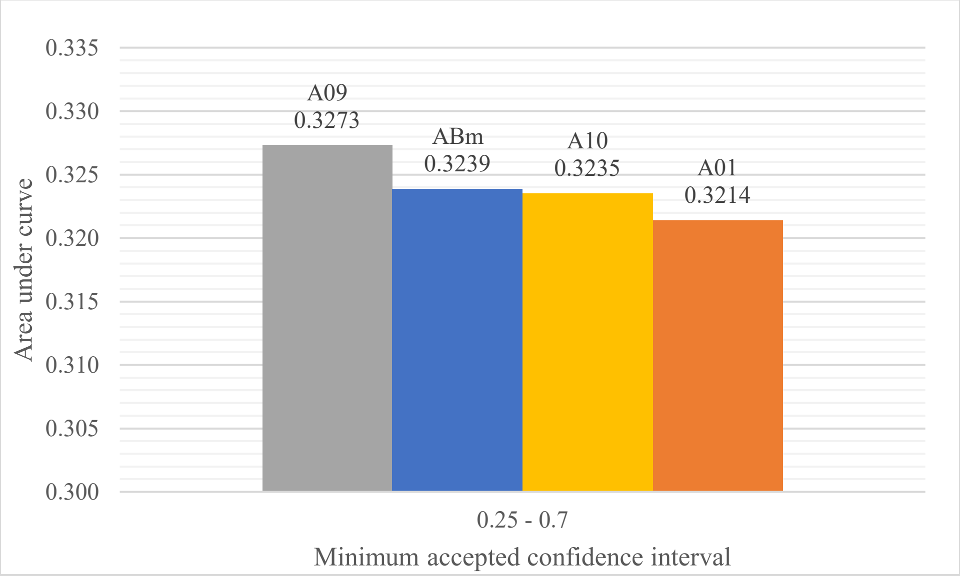


Figure 48 Comparison of AUC at MAC intervals for models trained with combinations of 60m GSD spectral band subtraction. The resultant model performance varies, but do not meet the criteria to be considered significant changes.

Removal of the 60 m GSD spectral bands (Figure 47) shows a performance loss with the elimination of B01 (aerosol), a performance improvement with the removal of B09 (SWIR) from the model at low MACs from 0.25 to 0.5. This is reflected in the AUC, though neither meet or near the 0.008 threshold to be considered significant. However, except for A01 at a MAC of 0.7 to 0.8, the changes at specific MACs are relatively minor at less than 0.02 F1-score difference. It is possible that the impact of removal of 60 m bands is relatively minor if the weights associated are minimized, resulting in a minimum performance impact whether included or removed.

- *The performance of a trained vessel model does not change uniformly with the addition or removal of specific spectral bands.*

4.1.5 Impact of Spatial Resolution

A difficulty with Sentinel-2 data is the presence of spectral bands with different GSDs, while trying to determine the value of the specific bands to the ML models. Sentinel-2 spectral bands used for this problem were provided at spatial resolutions of 10 m, 20 m, and 60 m, and resampled to 10 m if they were not already (section 3.1). This means that any conclusion is balancing two variables – GSD and spectral band interactions – instead of just spectral band interactions. It is possible to resample the 10 m and/or 20 m GSD imagery to match the spatial resolution of the 20 m and/or 60 m spectral bands, but this would call on simulating and evaluating the results on many different models (discussed further in section 5.1.1). As such, Figure 49 shows the results of utilizing different Sentinel-2 GSD spectral band groupings, and the impact on the trained model performance.

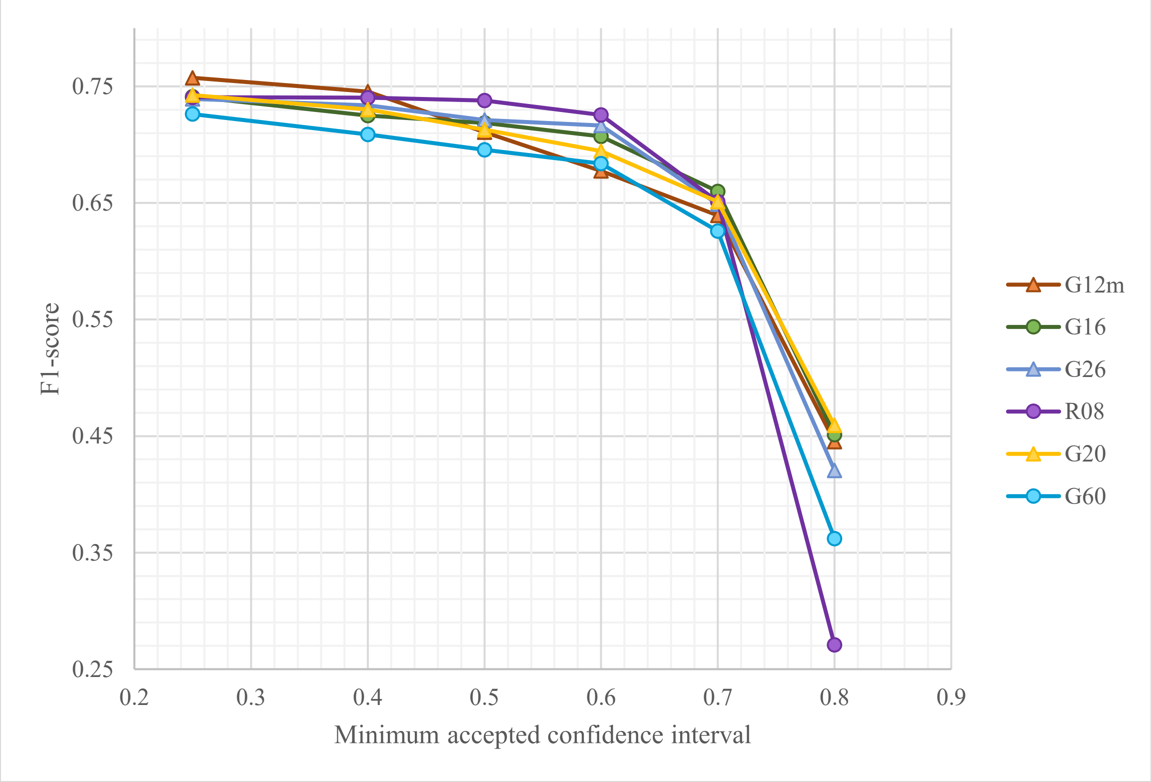


Figure 49 Comparison of F1 score at minimum accepted confidences for Sentinel-2 GSD spectral band groupings. Lower GSD and more spectral information generally improves model performance.

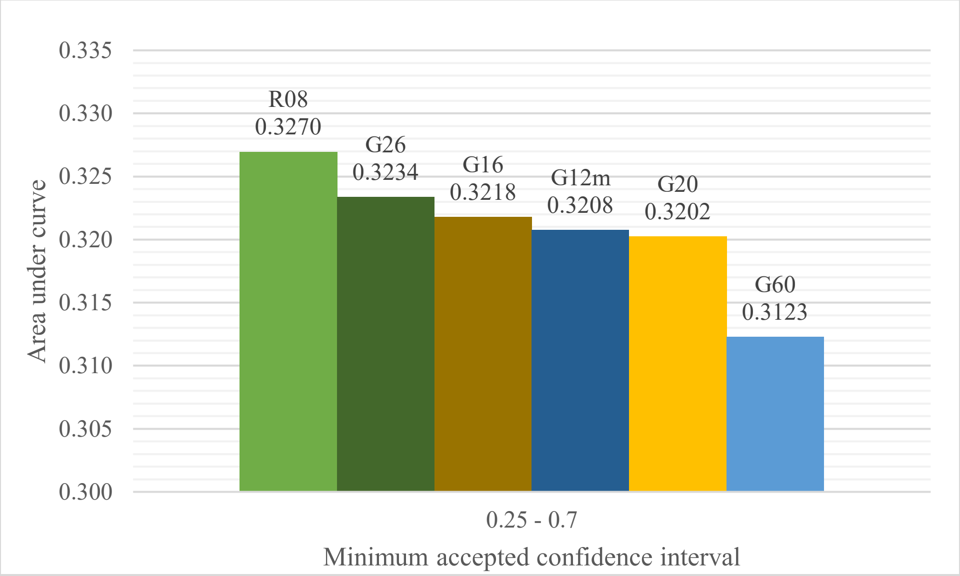


Figure 50 Comparison of AUC at MAC intervals for models of Sentinel-2 GSD spectral band groupings. While the lowest GSD bands offered the best performance (R08) and vice verse, it is difficult to draw further conclusions due to varying quantity of spectral information and value.

From Figure 49, there is a positive but rough correlation between model performance and lower GSD spectral bands along with volume of spectral information. The best model according to AUC (Figure 50) was A08 (all 10m GSD bands), while the worst by a significant margin is G60 (all 60m GSD bands). The remainder are clustered within a 0.0032 AUC range of each other. While GSD is considered, the additional variable of spectral band quantity (three 60 m, six 20 m, four 10 m GSD bands), and specific value of individual spectral bands are not accounted for. This means that any conclusion is tenuous at best, though a broad but unsurprising conclusion can be made.

- *The performance of trained vessel detection ML models improves with higher spatial resolution images.*

4.2 Model Inferencing Speed and Orbital Performance

While the performance metrics indicate the model capability, onboard computational resources for a CubeSat OBC are limited, and any solution must observe this. The inference speed for emulated hardware – whether from within a synthetic environment or a space-hardened OBC – needs to be understood. To understand the capabilities of onboard ML, the inferencing capabilities and the imaging capture rates must be considered. This section will discuss the results from inferencing images with the selected processors (4.2.1), along with contextualizing the inferencing time with respect to processing demands placed by imaging periods in orbit (4.2.2).

4.2.1 Model Inferencing Speed on Emulated OBC

Each model described in Table 7 has a slightly different process, whether in the number of parameters or the information to load into the model, consequently, *the inference time per image varies*. As discussed in Satellite Path and Hardware Processing (section 3.3) each model was loaded aboard the NVIDIA Jeston SBCs and the test portion of the dataset inferenced to obtain and measure an inferencing time. Two warm up runs were completed to place a thermal load on the model, followed by fitting the model to the test dataset and taking the median result of three additional runs for inferencing speed. Figure 51 shows F1-score as a function of inferencing time for all OBC configurations tested.

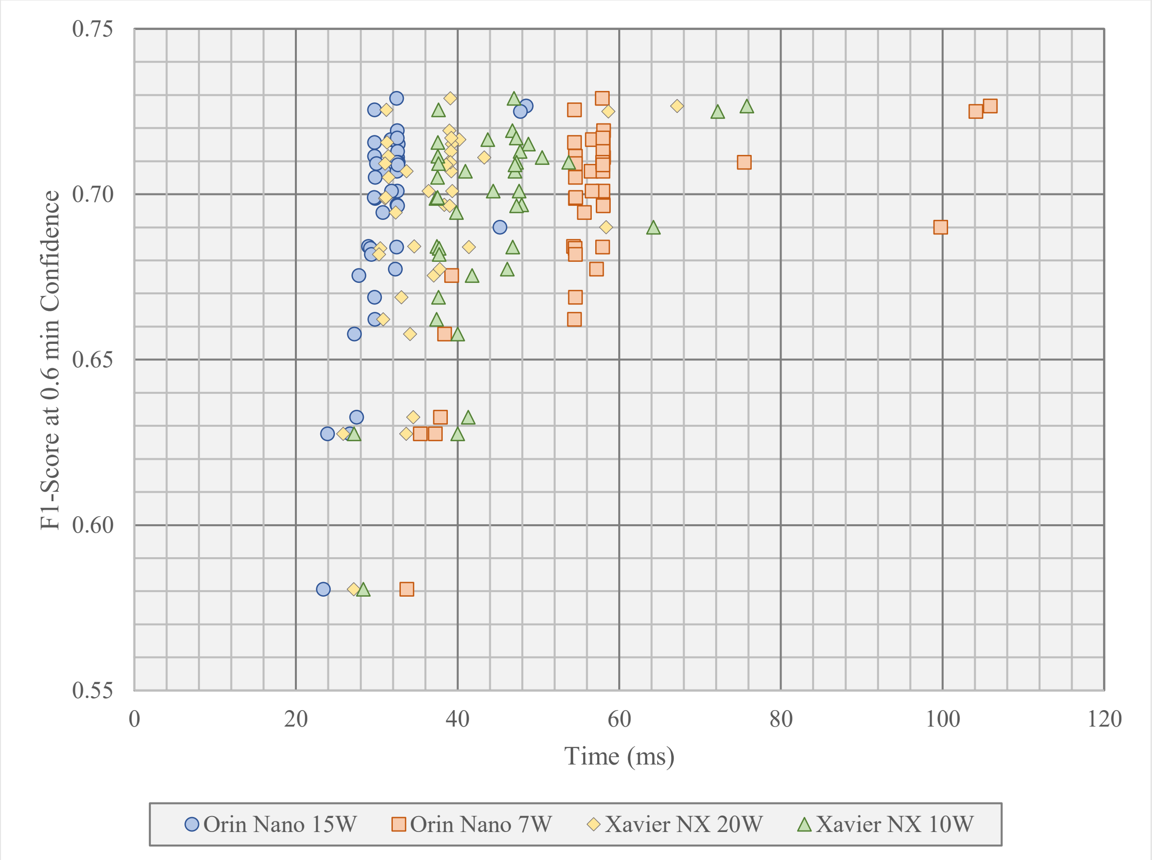


Figure 51 Inferencing speed vs F1-score at 0.6 minimum confidence for all tested hardware configurations. Inferencing occurs between 20 and 110 ms for all

models tested on all hardware configurations. This shows the relative performance of each hardware configuration.

From the inferencing speed test, the order of best OBC performance to least is the Orin Nano at 15W, Xavier at 20W, Xavier at 10W, and the Orin Nano at 7.5W. These appear as four clusters of points with minor overlap. It is important to note that the wattage represented in the name of each OBC configuration is not the actual power consumed by the SBCs, but the maximum they could consume. Characterizing the power draw of the hardware in detail through monitoring power draw from the wall or using onboard software was not completed during these steps. In the interest of legibility, Figure 51's F1-score axis is scaled so the results from R10 – the abnormal model – are not displayed. These removed points occurred at an F1-score of 0.1232 at 0.6 minimum confidence between 29.5 ms to 54.0 ms.

Figure 51, primarily shows a general trend of model performance compared to inferencing time, and relative hardware performance, but reveals nothing about model specifics. Plotting the model performance against inference time for a smaller sample of one OBC configuration will be more legible for model performance instead of a grid of over a hundred data points. Due to the number of similarly inferenced RGB plus a spectral band (R*xx* models), and all spectral bands minus a single spectral band (A*xx* models), only a portion of these will be plotted for legibility purposes. Figure 52 is the F1-score vs inferencing speed at 0.6 minimum confidence for the ML models using the Orin Nano in the 15W configuration.

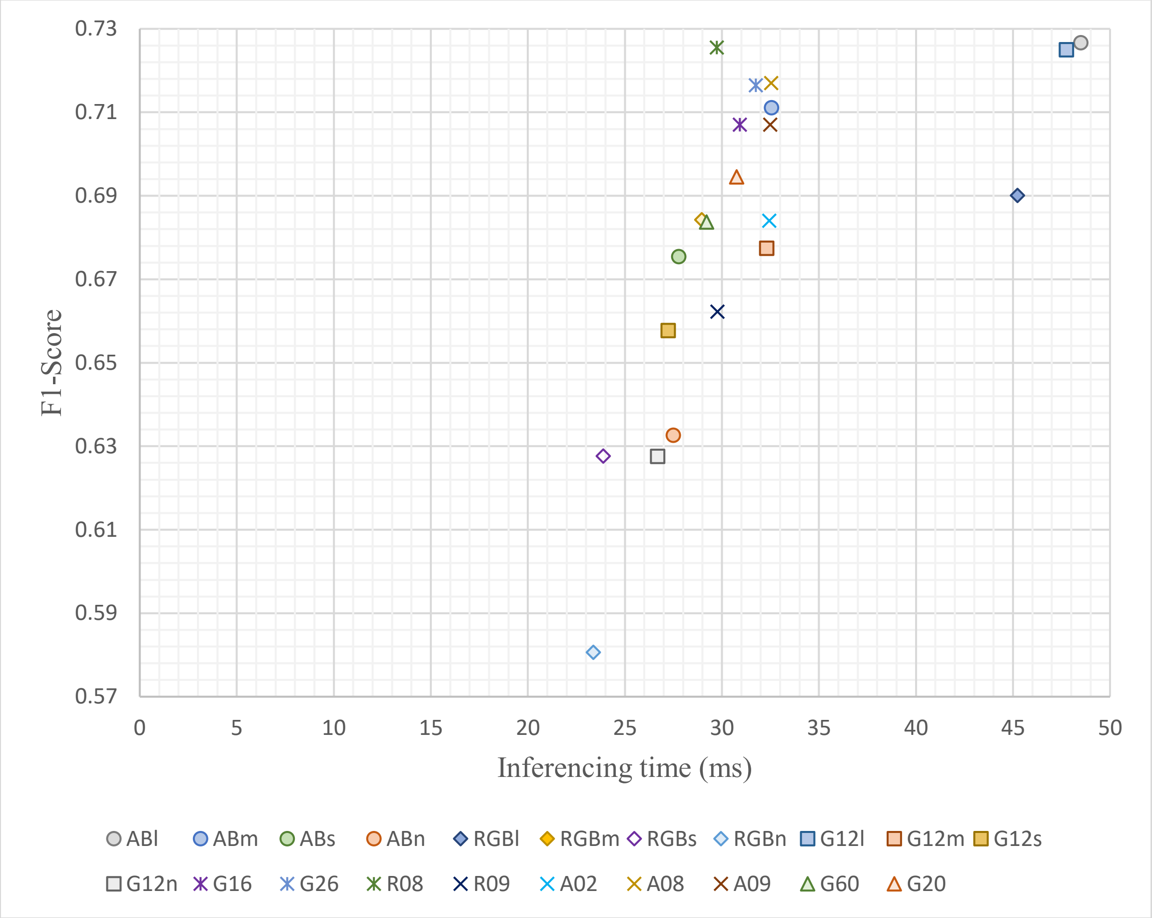


Figure 52 F1 score vs inferencing speed for ML models tested using Orin Nano in 15W configuration. Image inferencing for the ML models takes between 23 – 49 ms, equating to an inferencing rate of approximately 20 – 40 images per second. The rates suggest the models may be suitable for near real-time or real-time inferencing on the OBC.

For the Orin Nano in the 15W configuration, it was found to inference images at a rate between approximately 23 – 49 depending on the model size and spectral band combinations. Broadly, there appears to be a positive trend between inferencing time and F1-score, but two varying factors are implicitly captured in Figure 52: specific spectral band combinations and model size. To better examine the model inferencing time, the analysis is split between model size (Figure 53) and number of spectral bands (Figure 54 and Figure 55).

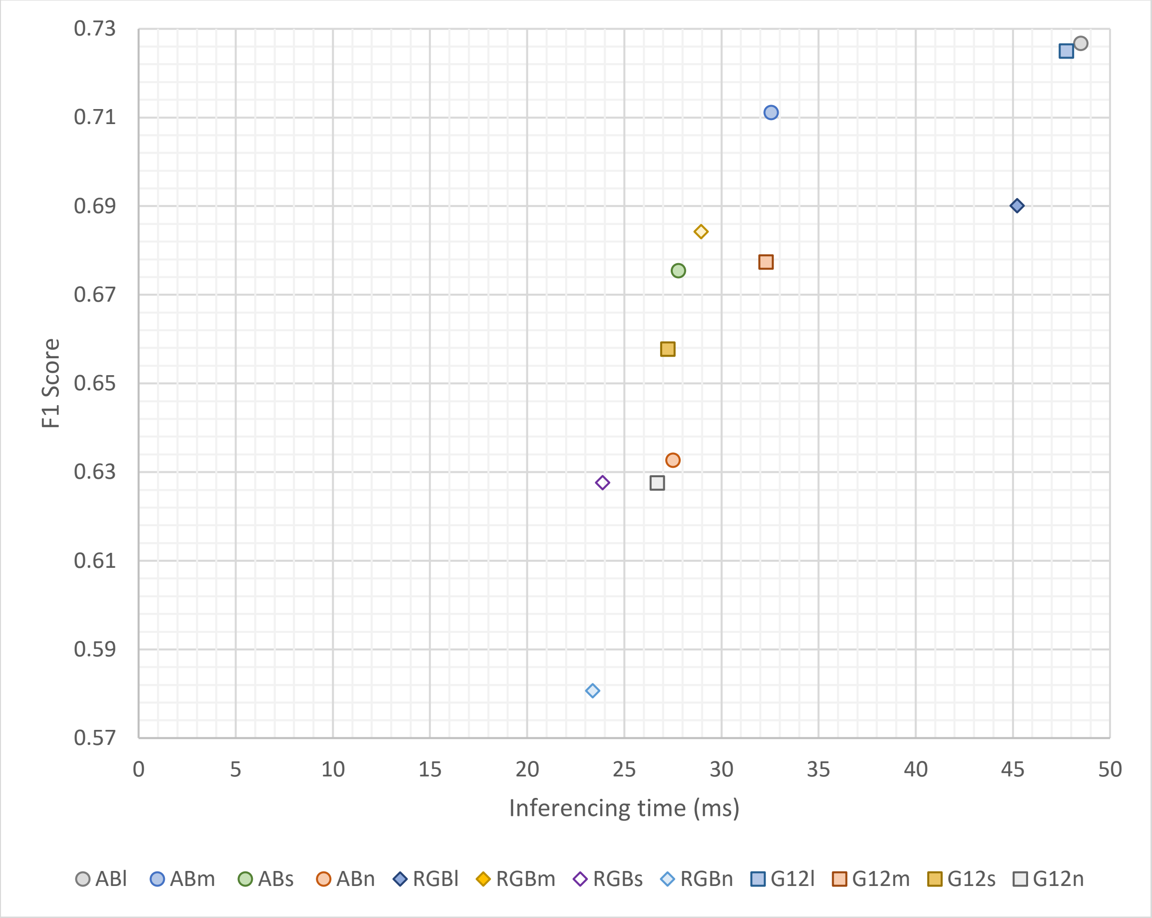


Figure 53 F1 score vs inferencing speed for different size ML models tested using Orin Nano in the 15 W configuration at 0.6 minimum accepted confidence. Inferencing time improvements decrease significantly for each model size *reduction*. This was expected behavior for the model.

As the model size shrinks, the model inferencing speed improves, but only to a point. Nano models showed very small inferencing speed improvement (<1 ms) at the cost of significant reduction in F1-Score. For the tested models, between $23 - 28$ ms is the floor for inferencing the images of any model size, with diminishing returns on performance vs inferencing time as model size increases. The performance for other OBC configurations will not be broken down like those for the Orin Nano in the 15 W configuration in Figure 52 and Figure 53, as they are similar for each OBC configuration, but with higher (worse) inferencing times.

To evaluate the effect of number of spectral bands upon inferencing speed, the models were examined using addition YOLOv8 inferencing speed results. YOLOv8 reports three measures of time during inferencing: pre-processing, processing, and post-processing, reporting the sum as the inference speed. Loading and preparing the image for insertion into the model, model processing, and handling of the results and output are pre-processing, processing, and post-processing, respectively. Up to this point, all presented inferencing times have been the sum of all three measures. The only changes to these different processing steps were the modification of the input layer of the model for the processing step and handling additional spectral channels images in the pre-processing step. The post-processing time remained constant during testing due to no changes having been made, and therefore it is not shown in this thesis. Figure 54 shows how the quantity of spectral bands influences processing time, while Figure 55 shows how the same influences pre-processing time.

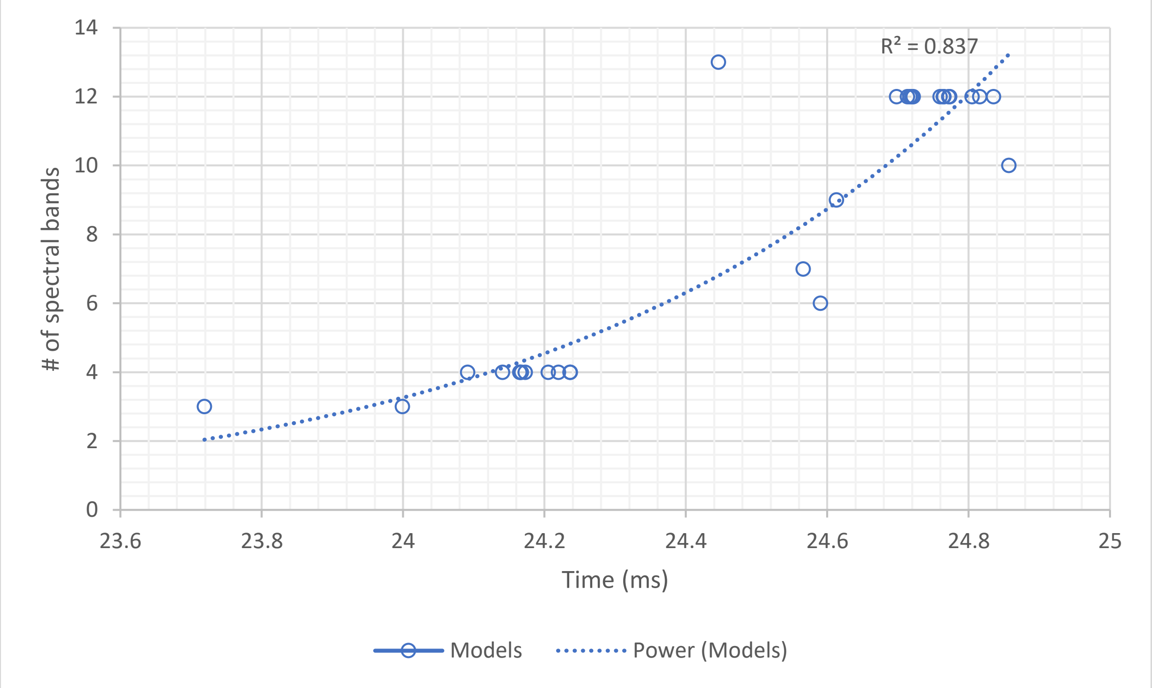


Figure 54 Number of spectral bands vs *processing* step time for medium sized YOLOv8 models tested on the Orin Nano 15W, fitted with power trend line. Additional spectral bands have minimal but non-uniform increase on inferencing speed, but do not represent a significant increase.

The effect of additional spectral bands on model processing time is minimal, with the largest between any two models to be just over 1 ms. This is not particularly significant, especially given the variance between models of the same size. The points clustered at 4 and 12 spectral bands are the Rxx models and Axx models, respectively. Additionally, the 13-band model – ABm – managed to process faster than a 6-band model. Completing more runs and averaging a wider range of results may have shown less variability between runs of the same size (instead of using the median of 3 runs for each model) but overall, the difference was unlikely to be significant.

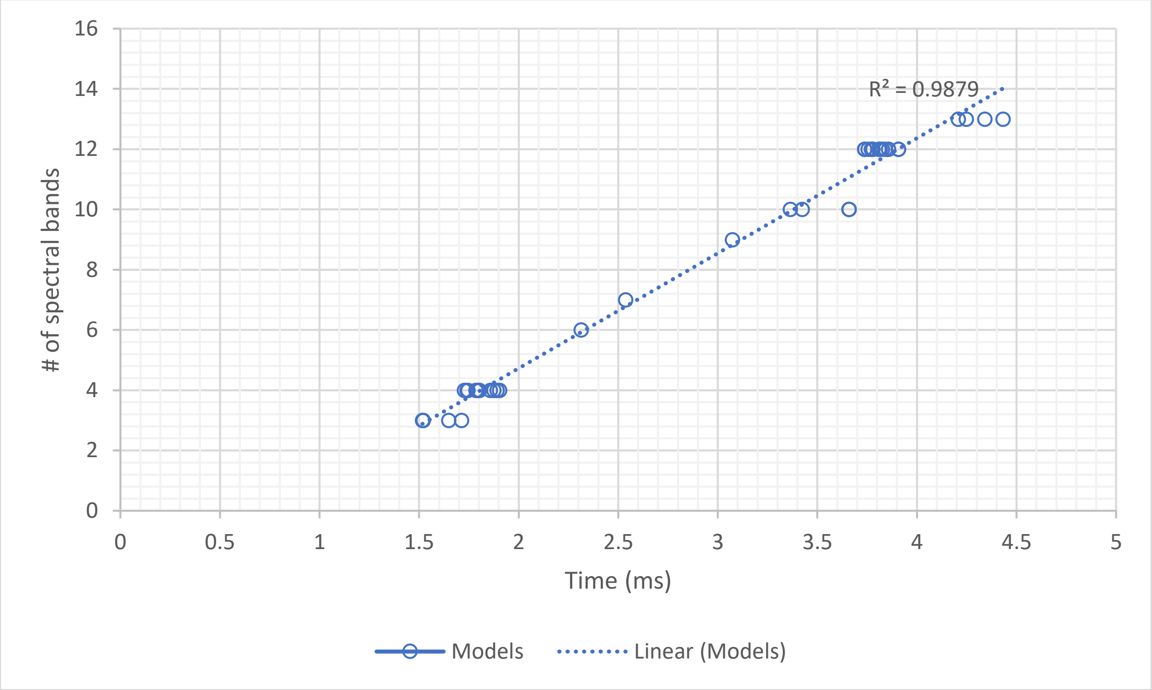


Figure 55 Number of spectral bands vs *pre-processing* step time for every YOLOv8 models tested on Orin Nano 15W with linear trend line. Additional spectral bands have a linear relation upon model inferencing time and have a larger impact than the effect of spectral bands during the processing step.

Unlike processing, the pre-processing inference time vs the number spectral bands (Figure 55) shows a more distinct relationship and experiences a larger difference of approximately 3 ms between the RGB (3 band) and 13 spectral band models. Since the model size has no impact on the pre-processing step, all trained YOLOv8 models in this thesis were included. The near-constant differences in inferencing times are not unexpected since the images loaded from file on a system running an OS and background functions. In a more accurate OBC emulation, the interface of the hardware and multispectral imaging system could be captured (section 5.1.2).

- *Additional spectral information increases the inferencing time of the model, primarily during the pre-processing step of the model.*

4.2.2 Impact of satellite orbit

Capturing images of the Earth from space with a nadir-pointing satellite can be simplified to a satellite travelling at a constant altitude over an infinite flat plan (i.e., with no curvature). Therefore, using the inferencing rate of the models aboard the tested OBCs (Figure 51) the limit of onboard processing in tandem with a simulated velocity could be determined. The MATLAB simulations (detailed in 3.3.2) calculates the velocity of each satellite in the north-east-down (NED) coordinate frame for their respective orbits.

For each of the four orbits, the largest differences between the minimum or maximum orbital velocities were less than 0.5% of the average velocity for each respective orbit. The difference of velocity over the course of an orbit was a result of the Earth rotating below the satellite, and the very small eccentricity of 0.001. As the velocities varied little over the course of the orbits, the mean velocity was taken to be constant at all points in the orbits (Table 13).

Table 13 Mean velocity of simulated orbits used to calculate inferencing rate.

orbit	**mean orbital velocity (NED coordinate frame)**
Sentinel-2	7542.6 m/s
ISS	7354.5 m/s
low SSO	7660.2 m/s
Sentinel-6	6969.0 m/2

Utilizing the inferencing rate from the tested OBC configurations, it is possible to determine the area the model can inference per second while operating with a 10 m GSD. This assumes the images were input as 128×128 multi-channel imagery from a line-scan camera feed. The results for the ABm model are presented in Table 14. Given the variations in ML models, it would be intensive to analyze results for all variations while contributing little to the end conclusion. As a result, the ABm model will be representatively used for further analysis.

Table 14 Area inferencing rate for each NVIDIA SBC evaluated with the ABm (all band, medium size) model inferencing results.

model	**Orin Nano** (15W)	**Orin Nano** (7W)	**Xavier** (20W)	**Xavier** (10W)
picture area (km^2)	1.6384	1.6384	1.6384	1.6384
inferencing time (ms)	32.5540	58.0722	43.3144	50.4489
picture rate (images/s)	30.7182	17.2200	23.0870	19.8220
inferenced area rate (km^2/s)	50.3287	28.2132	37.8258	32.4764

From this information, it is feasible to estimate what swath a corresponding OBC would be capable of inferencing. This is important as it does not matter how precise the model is if it is only able to inference a small area over the period of an orbit. Table 15 compares the inferenced area rate with respect to the imaging swath and orbit. At rates less than 100%, the OBC can pause between inferencing images. At rates of 100%, the OBC is fully saturated with respect to inferencing images, i.e., images are inferenced as fast as they are fed into the model. At rates greater than 100%, the OBC can not inference images as fast as they are captured. Note that additional OBC overhead, inefficiencies or improvements to code were not factored into the inference speed tests; it is a general emulation to coarsely determine the OBC's suitability.

In Table 15 to Table 18 the processing rate of 0-100% are marked with green, 100-200% with yellow, and higher than 200% red to improve visual clarity.

Table 15 Processor calculated saturation percent at different swaths and orbits for Orin Nano in the 15 W configuration.

camera swath (km)	**calculated processor saturation percentage (orbit)**			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
1.28	19.18%	18.70%	19.48%	17.72%
2.56	38.37%	37.41%	38.96%	35.45%
3.84	57.55%	56.11%	58.45%	53.17%
5.12	76.73%	74.82%	77.93%	70.90%

camera swath (km)	calculated processor saturation percentage (orbit)			
	Sentinel-2	ISS	low SSO	Sentinel-6
6.4	95.91%	93.52%	97.41%	88.62%
7.68	115.10%	112.23%	116.89%	106.34%
8.96	134.28%	130.93%	136.37%	124.07%
10.24	153.46%	149.64%	155.86%	141.79%
20.48	306.93%	299.27%	311.71%	283.59%
40.96	613.85%	598.55%	623.43%	567.17%
61.44	920.78%	897.82%	935.14%	850.76%
81.92	1227.71%	1197.09%	1246.85%	1134.34%
102.4	1534.64%	1496.36%	1558.56%	1417.93%

Table 16 Processor calculated saturation percent at different swaths and orbits for Orin Nano in the 7.5 W configuration

camera swath (km)	calculated processor saturation percentage (orbit)			
	Sentinel-2	ISS	low SSO	Sentinel-6
1.28	34.22%	33.37%	34.75%	31.62%
2.56	68.44%	66.73%	69.51%	63.24%
3.84	102.66%	100.10%	104.26%	94.85%
5.12	136.88%	133.47%	139.01%	126.47%
6.4	171.10%	166.83%	173.77%	158.09%
7.68	205.32%	200.20%	208.52%	189.71%
8.96	239.54%	233.57%	243.27%	221.32%
10.24	273.76%	266.93%	278.03%	252.94%
20.48	547.52%	533.86%	556.06%	505.88%
40.96	1095.04%	1067.73%	1112.11%	1011.76%
61.44	1642.56%	1601.59%	1668.17%	1517.64%
81.92	2190.08%	2135.46%	2224.22%	2023.53%
102.4	2737.60%	2669.32%	2780.28%	2529.41%

Table 17 Processor calculated saturation percent at different swaths and orbits for Xavier NX 8GB in the 20 W configuration

camera swath (km)	calculated processor saturation percentage (orbit)			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
1.28	25.52%	24.89%	25.92%	23.58%
2.56	51.05%	49.77%	51.84%	47.17%
3.84	76.57%	74.66%	77.76%	70.75%
5.12	102.09%	99.55%	103.69%	94.33%
6.4	127.62%	124.44%	129.61%	117.91%
7.68	153.14%	149.32%	155.53%	141.50%
8.96	178.67%	174.21%	181.45%	165.08%
10.24	204.19%	199.10%	207.37%	188.66%
20.48	408.38%	398.19%	414.75%	377.32%
40.96	816.76%	796.39%	829.49%	754.65%
61.44	1225.14%	1194.58%	1244.24%	1131.97%
81.92	1633.52%	1592.78%	1658.98%	1509.29%
102.4	2041.89%	1990.97%	2073.73%	1886.61%

Table 18 Processor calculated saturation percent at different swaths and orbits for Xavier NX 8GB in the 10 W configuration.

camera swath (km)	calculated processor saturation percentage (orbit)			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
1.28	29.73%	28.99%	30.19%	27.47%
2.56	59.46%	57.97%	60.38%	54.93%
3.84	89.18%	86.96%	90.57%	82.40%
5.12	118.91%	115.95%	120.77%	109.87%
6.4	148.64%	144.93%	150.96%	137.34%
7.68	178.37%	173.92%	181.15%	164.80%
8.96	208.09%	202.91%	211.34%	192.27%

camera swath (km)	calculated processor saturation percentage (orbit)			
	Sentinel-2	ISS	low SSO	Sentinel-6
10.24	237.82%	231.89%	241.53%	219.74%
20.48	475.65%	463.78%	483.06%	439.47%
40.96	951.29%	927.57%	966.12%	878.95%
61.44	1426.94%	1391.35%	1449.18%	1318.42%
81.92	1902.58%	1855.13%	1932.24%	1757.89%
102.4	2378.23%	2318.92%	2415.31%	2197.37%

Table 15 to Table 17 show OBC configurations are incapable of real-time processing beyond relatively low swath widths. For the 15W Orin Nano, for all orbits, real-time processing is only possible when the processed swath is 6.4 km or less for a 10 m GSD, and smaller for other SBC. This is a very small processing width that questions the suitability of such a system for onboard imaging process. In addition, this test ignores any additional OBC overhead from integration of a MS imaging system, and handling of data for transmission.

- *Real time MS vessel detection using developed ML models is not feasible with tested hardware.*

However, as discussed in the Orbital Simulation section (3.3.2**)** there are opportunities to reduce the load placed on the processor, and near real-time processing may be sufficient for the task of vessel detection. For instance, a vessel moving 15 knots moves less than 1 km in a minute, which would be well within the distance to horizon described in Table 12.

The orbital simulation of each orbit completed was run from 1st August 2023, 00:00 to 30th August 2023, 23:59, simulating 2,592,000 points in orbit at 1 second time steps, and correlated to GEBCO bathymetric data. Applying the various constraints (described in Table 11) to valid imaging points directly below a simulated nadir-pointed satellite, reduces the volume of imagery each satellite captures over time. The percentage of orbit that is valid for imaging under these constraints is shown in Table 19. Note that all constraints are cumulative when moving left to right.

Table 19 Percentage of 30-day simulated suitable for imaging under cumulative constraints (top to bottom)

	orbit description and suitable imaging percentage			
imaging restriction	**Sentinel-2**	**ISS**	**low SSO**	**Sentinel-6**
no restrictions	100.00%	100.00%	100.00%	100.00%
satellite not in eclipse	68.78%	61.61%	65.83%	70.99%
below sea level (0 m)	41.61%	40.68%	40.66%	46.17%
deep-sea fishing	10.66%	5.26%	9.82%	7.93%
70° N – 70° S	5.40%	5.26%	4.78%	7.93%
continental shelf	3.26%	3.05%	2.89%	4.53%
Nova Scotia region	0.13%	0.30%	0.13%	0.16%

For a satellite tasked to monitor an area such as deep-sea fishing waters between sea level (0 m) and the end of common deep sea fishing depths (-1500m), while limited to be within 70°N – 70°S, around only 5% of the orbit will be spent directly over valid imaging points. For the remainder of the time the satellite imaging would idle if no other task were assigned – this permits processing during these time periods. Returning to Table 15 to Table 17 for the OBC configurations, if the result from multiplying the processor saturation percentage with the percentage of valid imaging periods for cumulative restraints (Table 19) remains below 100% (or the corresponding eclipse valid imaging percentage if considering power management) it would be feasible to process all valid images onboard. Processing may be delayed from point-of-capture but processing the images over the course of the orbital period would be possible in near real-time processing.

A visual representation of valid positions for an orbit similar to Sentinel-2, with the restrictions of not being in eclipse and below sea level is offered in Figure 56. Another visual representation for the same orbit, but with the added constraints of between 0 m to -1500 m below sea level, and between 70°N – 70°S is presented in Figure 57.

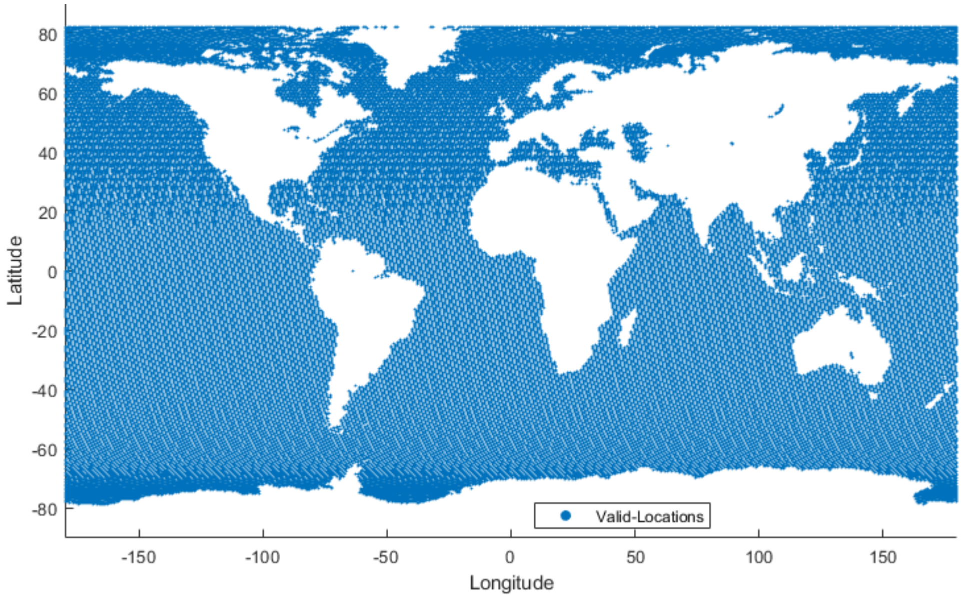


Figure 56 NADIR pointed satellite in Sentinel-2 orbit simulation for valid imaging points below sea level, and not during satellite eclipse, for a 30-day period.

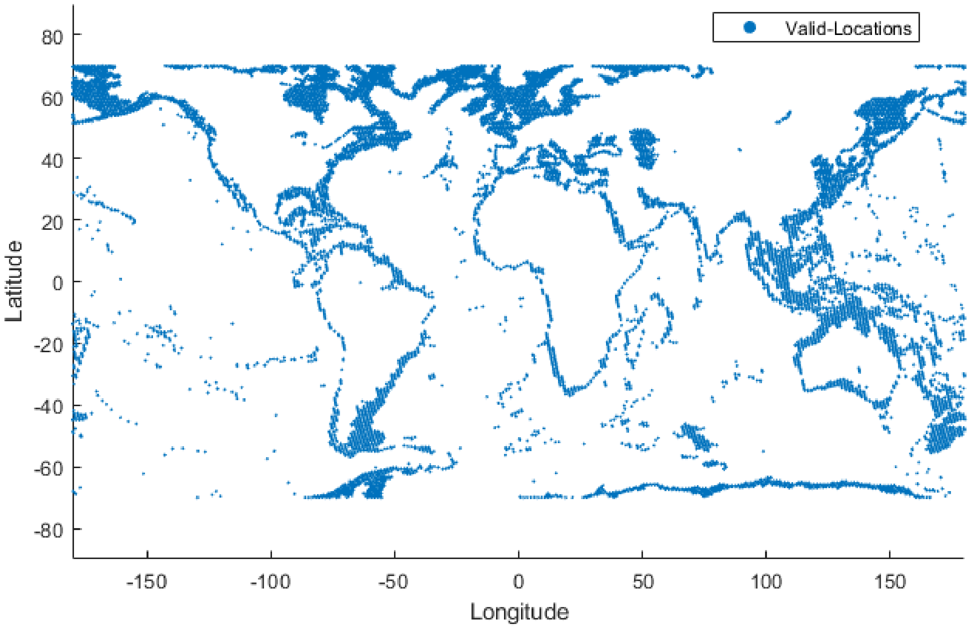


Figure 57 Nadir pointed satellite in Sentinel-2 orbit simulation for valid imaging points between 0 to 1500m below sea level, between 70°N to 70°S, and not during satellite eclipse, for a 30-day period. Significantly less area of Earth is

considered a valid imaging location, lowering captured imagery volume and what needs to be processed.

The risk with near real-time processing is the information being delivered too late to be acted upon. Just as different objectives for coastal earth monitoring for coast studies have different revisit rate requirements (Figure 7**)**, processing and timely delivery of the information is important. For dark vessels, detecting a vessel minutes after image capture is better than an hour after image capture, as the vessel may transit elsewhere during the orbital period. Another way to understand the impact of constraints upon the captured data is to consider the length of the periods where the satellite could capture valid imagery (influencing the volume of imaging captured in one capture period), and the length of the periods where it cannot (influencing the available time to process information without additional captured imagery). Statistics about the length of valid imaging periods are shown in Table 20. Statistics about the length of invalid imaging periods are shown in Table 21

Table 20 Statistics of *valid* imaging periods (s) for simulated orbits and constraints. Imaging restrictions are cumulative from top to bottom.

imaging restrictions	**orbit description**			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
eclipse				
mean period (s)	4145.9	3434.3	3766.8	4779.6
median period (s)	4137	3433	3765	4777
longest period (s)	4357	3457	3941	4909
below sea level				
mean period (s)	124.52	210.62	120.91	190.10
median period (s)	7	18	7	10
longest period (s)	2552	2522	2223	3393
between 0 - 1500m				
mean period (s)	20.52	15.20	18.51	20.16
median period (s)	5	5	5	5
longest period (s)	354	307	278	299
70°N to 70°S				
mean period (s)	16.18	15.20	14.08	20.16
median period (s)	5	5	5	5
longest period (s)	254	307	208	299
between 0-250m				
mean period (s)	11.52	10.84	10.02	12.55

imaging restrictions	orbit description			
	Sentinel-2	ISS	low SSO	Sentinel-6
median period (s)	4	4	3	4
longest period (s)	213	272	204	258
region				
mean period (s)	12.15	12.07	9.73	15.75
median period (s)	5	5	4	7
longest period (s)	78	111	53	77

Table 21 Statistics of *invalid* imaging periods for simulated orbits and constraints. Imaging restrictions are cumulative from top to bottom.

imaging restrictions	orbit description			
	Sentinel-2	ISS	low SSO	Sentinel-6
eclipse				
mean period (s)	1882.1	2139.9	1959.4	1979
median period (s)	1899	2143	1971	1965
longest period (s)	2035	2163	2075	2051
below sea level				
mean period (s)	174.72	307.10	176.44	221.60
median period (s)	7	11	7	8
longest period (s)	3134	3272	2948	3491
between 0 - 1500m				
mean period (s)	172.08	273.93	169.93	234.06
median period (s)	12	17	12	14
longest period (s)	4469	4939	2948	4474
70°N to 70°S				
mean period (s)	283.51	273.93	280.1	234.06
median period (s)	17	17	17	14
longest period (s)	6722	4939	5518	4774
between 0-250m				
mean period (s)	341.71	345.11	336.82	264.35
median period (s)	15	18	15	12
longest period (s)	6725	5163	5539	5994
region				
mean period (s)	11664	5080.3	10235	11766
median period (s)	11	9	7	13.5
longest period (s)	84433	66099	349917	87634

As the constraints for valid imaging areas increase, the mean, median and longest periods for valid imagery to be captured decrease, while invalid periods generally rise. The restriction of below sea level (0 m) still leaves upwards of 40 minutes of valid imaging time as a possibility, and the ratio of mean valid periods compared to the invalid periods is approximately 2:3. However, imposing the constraint of 0 to 1500m depth brings the longest valid period to within 6 minutes. The ratio of valid to invalid periods also sharply falls as restrictions are applied. The mean valid/invalid ratios are 1:8 to 1:15 depending on the orbit, while the median is approximately 1:3 for the constraints of between 0 to -1500m depths. This ensures there remains time for processing, power permitting.

Of course, the frequency of these periods also plays a role, which the median helps distinguish. As soon as any imaging constraint is applied beyond that of eclipse, the median valid times all plummet to less than or equal to 10 seconds, while the invalid period rises to 7 − 18 seconds. In practice, this means there are frequent bursts of imaging along with intermittent periods of invalid image processing. Therefore, any queue of images that forms is more quickly addressed and processed relative to the time the image was captured, instead of accumulating in large batches before processing.

The results of Table 20 can be used to express how long it would take an OBC to process the results valid image duration expressed in Table 20. Table 22 shows how long it would take the Orin Nano 8GB in a 15W configuration to process the imagery captured for the mean, median and longest valid imaging durations. Due to model running on the CUDA cores of the GPU, not the CPU, it is assumed that the OBC can process imagery at a reduced rate during this period. In the absence of hardware-in-the-loop testing of a camera system an arbitrary 900% penalty is added to ABm model inferencing speed for inferencing speed while imaging.

Table 22 Time required for Orin Nano in 15W configuration to process 20.48km swath imagery captured during the mean, median, and longest valid imaging durations indicated in Table 20. The results compare favorably to the invalid imaging durations that can be used for processing (Table 21) once additional imaging restrictions are applied.

imaging restrictions	**orbit description**			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
eclipse				
mean period (s)	12310.29	9934.49	11364.90	14191.92
median period (s)	12283.87	9930.73	11359.47	14184.20
longest period (s)	12937.11	10000.16	11890.48	14576.14
below sea level				
mean period (s)	369.73	609.27	364.80	564.46
median period (s)	20.78	52.07	21.12	29.69
longest period (s)	7577.58	7295.46	6707.07	10074.73
between 0 - 1500m				
mean period (s)	60.93	43.97	55.85	59.86
median period (s)	14.85	14.46	15.09	14.85
longest period (s)	1051.12	888.07	838.76	887.81
70°N to 70°S				
mean period (s)	48.04	43.97	42.48	59.86
median period (s)	14.85	14.46	15.09	14.85
longest period (s)	754.19	888.07	627.56	887.81
between 0-250m				
mean period (s)	34.21	31.36	30.23	37.26
median period (s)	11.88	11.57	9.05	11.88
longest period (s)	632.45	786.82	615.49	766.07
region				
mean period (s)	36.08	34.92	29.36	46.77
median period (s)	14.85	14.46	12.07	20.78
longest period (s)	231.60	321.09	159.91	228.63

For imaging restrictions of 0 to -1500m or further, the numbers compare favorably to those in Table 21. The times required to process mean and longest durations for valid imaging periods for these restrictions are smaller than the mean and longest periods for invalid imaging periods. Medians vary, but they are only a couple seconds difference at most. The only issue with this analysis is it doesn’t account for a valid imaging period occurring shortly after a large valid imaging period is still being processed. Returning to the simulation of the 4 orbits, and factoring when each valid image occurs, the maximum time for an image to be processed can be obtained and is shown in Table 23

Table 23 Maximum time required for an image to be processed using an 15W Orin Nano. With restrictions of only eclipse and then below sea level, the buffered images continually grow, while additional restrictions limit the maximum time to less than 36 minutes.

imaging restrictions and maximum time (s)	**orbit description**			
	Sentinel-2	**ISS**	**low SSO**	**Sentinel-6**
Eclipse	increasing	increasing	increasing	increasing
Below Sea Level	increasing	increasing	increasing	increasing
Between 0 to -1500m	2131.0	1600.5	1593.5	1700.6
70°N to 70°S	1047.2	1600.5	1042.1	1700.6
Between 0 -250m	796.4	1219.3	635.5	964.29
Region	234.2	546.3	232.0	357.8

The volume of captured but not processed (buffered) images for the restrictions of eclipse and below sea levels continues to grow during the period, indicating the OBC cannot keep up with the incoming images. By implementing imaging restrictions of from 0 to 1500 m below sea level, the OBC can keep up, with a longest (worst-case) buffered image time of 2131 seconds: under 36 minutes. In this time a vessel traveling at 15 knots (27.78 km/h) is only capable of travelling 16.44 km. Compared against Table 12, this distance is only further than distance to horizon for 5 and 10 m observation heights. For the 20 observation heights and vessels beyond the horizon for all observation heights, a vessel would still be visible. While the analysis does not address platform factors such as power draw, it indicates from an OBC perspective useful near-real time vessel detection is feasible.

- *Near real-time MS vessel detection using developed ML models is feasible for in-orbit operation with tested hardware when constraints are placed upon imaging periods.*

Chapter 5 Conclusion and Discussion

This thesis examined the suitability of utilizing multispectral imaging for vessel detection with machine learning algorithms and assessing the viability of onboard in situ satellite processing for this task. To revisit the conclusions from the results (Chapter 4):

- Multispectral imaging offers greater detection capabilities for sea level marine vessel detection beyond RGB wavelengths as implemented in an ML model.
- Additional spectral information improves the performance of ML models to detect vessels present within an image.
- The inclusion of any singular Sentinel-2 10 m or 20 m spectral band to the RGB bands has a positive impact on trained ML model performance.
- The additional of any Sentinel-2 60 m spectral band to RGB bands has minimal impact on trained ML model performance.
- The performance of a trained vessel detection ML model does not change uniformly with the addition or removal of specific spectral bands.
- Additional spectral information increases the inferencing time of the model, primarily during the pre-processing step of the model.
- Real-time MS vessel detection using developed MS ML models is not feasible for in-orbit operation with the tested hardware.
- Near real-time MS vessel detection using developed ML models is feasible for in-orbit operation with tested hardware when constraints are placed upon imaging periods.

Based on ML models trained on MS spectral band combinations, it was shown that MS imaging outperformed RGB for the designed model architecture and that there is value to multispectral imagery. Thirty-nine models trained on images created from spectral band combinations – derived from Sentinel-2 imagery – also showed that different spectral band combinations had an impact on model performance for vessel detection. The effect of MS imaging, compared to RGB, was better performance for a constant model size, or similar performance using a smaller model size which permits faster inferencing. However, the effect of adding spectral bands is not uniform as the different spectral bands of the Sentinel-2 MSI have different impact on model performance.

From the results of this research, the feasibility of performing ML onboard a satellite's OBC was determined using SBCs similar to those used by satellite companies. This portion of the work showed that – while it is not possible to perform vessel detection upon MS imaging fast enough to enable real-time processing of imagery – near-real time processing for MS ML vessel detection is feasible when imaging restrictions on image capture are in place. These restrictions were applied to four simulated orbits and included the effect of maximum ocean depth permissible for imaging, not imaging when the satellite is in eclipse, and imaging within a region, which lowered the volume of images the SBC must process and introduced frequent non-imaging periods to process the data.

If the outcome of this research could be applied to ML-capable satellite platforms coupled with integrated AIS detections, this would enable onboard autonomy for dark vessel detection from small CubeSat platforms. The data transmission required for each satellite to transmit data to a ground station for processing (if further processing is required) would be significantly reduced. Additionally, the edge compute generated data, or derived instructions, could be sent to direct other satellites without a downlink. For instance, a satellite detecting a vessel could cue satellites with a higher spatial resolution imaging system to confirm the target and classify the vessel by type. Or, if running a very lightweight cloud detection algorithm atop and before a vessel detection algorithm, a system could direct SAR satellites to cloud-covered areas, optimizing their use for when the capabilities are required. However, given the breadth of the subject, the scope of the thesis could only cover a small portion of utilizing MS imaging for vessel detection, let alone the intricacies of a multi-modal satellite detection constellation. There is room for additional research to investigate these capabilities, which is discussed in section (5.1)

5.1 Future Work

Possible future work identified during the thesis to address these areas can be divided into two categories: work that can be undertaken immediately as a continuation of the work completed in this thesis, and work that requires more extensive retooling. The first section, immediate future work (5.1.1) explores options to further improve confidence in the obtained results without extensive modification of the produced dataset or training process. Non-immediate future work (5.1.2) will discuss work that requires more fundamental changes to the process, dataset, or simulations.

5.1.1 Immediate Future Work

5.1.1.1 Vessel Size and Location on Performance

As mentioned in Chapter 3 (Methodology) all images were created using geo-referenced ships within the multispectral satellite imagery. This means that at least one vessel within each image should have an associated AIS message relating vessel size and class. The next portion of work which could be completed immediately for the trained models is associating *how* the model performs with respect to vessel size. The shortest vessel length within the dataset is 20 m, while the longest is over 350 m. Based on the results of previously trained satellite imagery vessel detection models [43, 26], there is definite correlation between the size of the vessel and the ML model's performance. With the specific training, testing and validation dataset split used, there were 198 vessels in the test portion of the dataset. This number is a reasonable amount to separate into a few discrete bins or attempt to create a more continuous analysis based around size.

Another part the previously trained satellite ship detection models measured was model performance based on vessel location, specifically, whether the vessel was docked and alongside land or in open water. While the dataset created for this thesis does not label whether a vessel is docked or in open water, it would be a straightforward label to add to the dataset for further analysis. To minimize work, only the test portion of the split dataset needs to be labelled with the docked label at minimum. If trying to avoid further labelling, future work could instead use existing "land" labels and evaluate how the performance varies based on whether water is present in the image, a mix of land and water, or if it is entirely land.

5.1.1.2 OBC Configuration Variations

In this thesis two SBCs were used to test the model inferencing time for two power configurations. However, both the NVIDIA Orin Nano 8GB and the NVIDIA Xavier NX 8GB were not the only low powered (20W or lower) SBCs available to test. Originally, the plan was to perform OBC testing on two additional SBCs: the NVIDIA Jetson TX2 8GB, and the Nvidia Jetson Nano 4GB, the latter of which was slightly accountable for the small dimensions of the image and corresponding model size. Unfortunately, the installation process for the required software and Python libraries required proved to be difficult and more convoluted on these platforms due to older versions of JetPack, corresponding software, and different compiling processes, and were subsequently dropped.

While the Orin Nano featured 2 built-in power modes, both of which were tested, the Xavier NX featured 9 built-in power modes of which only 2 were used. These modes had several distinct variations of 10W, 15W and 20W configurations that achieved the targeted power limits by using techniques such as turning off some CPU cores but running the remaining cores at higher clock speeds, to modifying the clock speed of the GPU and memory. As previously mentioned, the 15W configuration does not represent the power it consumes, but the peak power at which it can operate. To that end, running the code while monitoring power draw would allow better characterization of performance, and allow investigation of OBC power efficiency for the task.

5.1.1.3 Spectral Band Combinations

Though many band combinations were trialed during the research, they constituted only a fraction of the total number of possible band combinations. With 13 different spectral bands in Sentinel-2 imagery, there are thousands of permutations of band combinations possible. The bands combinations selected were often biased towards RGB colour bands as most common imaging systems utilize visible light in some capacity. Revisiting this problem with a wider range of spectral band combinations may yield more insight and a stronger correlation between model performance and specific bands. Additionally, model size was also only explored through three band combinations (RGB, AB, and G12), not all the variations trialed or that could be trialed. However, it is believed the models that were trailed represent a suitable sample to support the conclusions.

5.1.1.4 Alternative ML Models

In the Methodology Chapter it was mentioned that the Resnet-50-FPN failed to achieve meaningful performance and had to be abandoned for the thesis. The model did not produce results indicating where the ships accurately were, varying the bounding box in both dramatic size, shape, and general position. The training process did indicate the model moving towards a solution, but never far enough to even reasonably be considered. For instance, the worst performing model tested, R10, would have outperformed some of the results obtained from resnet-50-FPN. It would be worthwhile revisiting this model to see if the problem could be amended to evaluate performance with a different model. It would also be worthwhile to implement other models, to gauge how the performance varied with respect to inferencing time. The model was originally dropped due to split focus between fixing Resnet-50-FPN and modifying YOLOv8 affecting the quality of either produced solution, when only one working model that met near real-time processing criteria would satisfy the results and conclusion. As YOLOv8 was indicated as the faster model, Resnet-50-FPN was dropped along with any intention of developing a purpose-built lightweight CNN MS ML model to better enable real-time processing.

5.1.1.5 Training Parameters

YOLOv8 trains with many adjustable trainable parameters. These include ones previously mentioned, such as learning rate, weight decay, and momentum, in addition such as warmup bias, augmentation of images, or optimizer utilized to name a few. Tuning these values is critical for obtaining the best possible performance for a model, but this was not completed for any models evaluated. For each of the 39 models tested, there are potentially hundreds of meaningful training parameter variations. However, except for R10, all models trained and tested produced meaningful results. Testing different combinations of parameters begins to venture more into the optimization of YOLOv8 over gauging broad value (if any) of multispectral bands and may only be relevant to the specific combination of spectral bands for the specific dataset.

5.1.2 Non-Immediate Future Work

5.1.2.1 Dataset

The dataset lacks on several fronts, leaving room for refinement. An original goal was to evaluate how the presence of clouds affected vessel detection performance with different spectral band combinations. The disagreements between the vessel AIS locations and their locations within an image meant the vessels were located based on composite RGB images, reducing the number of partly obscured vessels present within the dataset. While it was possible to utilize the dataset to complete the work presented in this thesis, lack of class support for clouds within the dataset limits application of resultant models to clouded imagery. Increasing the images with clouds present would require collecting significantly more data, requiring either the purchase of additional data storage or a completely modified data processing pipeline.

The problem with the AIS may have been possible by implementing the process detailed in [26]. In this work the author identified that the time stamp from the medium resolution imager, Landsat 8, deviated from the AIS timestamp. The solution involved linear interpolation of multiple AIS reports for a vessel within an image, in concert with OpenStreetMap [80] water polygons and image pixel coordinates to interpolate a historical path and location of the vessel. However, this process still requires manual oversight for every ship position within an image. This would have been a more efficient for processing more imagery, but unlikely to affect the results of this thesis.

Next, the creation of the image dataset should have been changed and could still be improved. The dataset features a single AIS message for each image regardless of how many ships occur within the image. This was an oversight and should have been rectified at the time. Instead, when multiple ships occurred within the same image, multiple images were created and gathered leading to near duplication and human bias in their reduction from the dataset. While this method could be argued as being preferable considering the difficulties with geolocating moving vessels, it would be possible to overcome with a different data processing workflow and review, and the solution previously mentioned.

Manually cropping labelling the 128×128 MS images was performed using RGB composite imagery, instead of MS imagery. While this did not affect the cropping of stationary vessels – as AIS and MS imagery position matched – it affected cropping of moving vessels, and labelling images. It may have improved the dataset to incorporate using multiple spectral channels for these steps, but this would required reworking the cropping and per-image labelling program extensively to ensure no useful labelling information is lost or obscured during the process. For bounding box labelling, a program would need to be written or software capable of handling multispectral imaging while labelling identified.

5.1.2.2 Mixed Resolution Imagery

One of the problems with Sentinel-2 imagery for the purpose of determining which bands are suitable for vessel detection is that not all bands are equal. While all bands were resized to 10 m GSD, the original bands themselves may be 10 m, 20 m, or 60 m. It is difficult to determine if B08 – a 10m GSD NIR band – is more suitable than B01 – a 60m GSD aerosol band – for vessel detection without a more equivalent analysis. This is especially true when the features being observed may be smaller than the GSD. One uncorrected error during the process was the rescale function used to resize 20m and 60m GSD imagery was thought to have resized the imagery without interpolating values. Instead, the rescale function did interpolate the values using nearest-neighbors interpolation and it was only noticed at the very end of the thesis. This likely resulted in better performance than would be achieved if not interpolating when utilizing 20m and 60m bands. Ultimately, it is unlikely that this adversely affected the conclusions of this thesis.

5.1.2.3 Onboard Computer

Determining the inference rate of simulated hardware was not without fault. One critical problem was that inferencing rate was completed using images requiring a minimal level of processing already saved and processed onboard. The overhead of processing the raw camera data into a suitable size and format is not factored into the inferencing rate. The simulation assumes it is being “passed” data that is already correctly formatted, with nothing more than a conversion from a multi-channel TIFF image into the model as a NumPy array. Given more time and resources to integrate a physical camera system representative of a satellite camera system, the quasi hardware-in-the-loop could more closely begin to resemble actual inferencing time results.

Another factor that is as much as it is attributable to the dataset as it is the hardware is the size of the images. Because there were no guidelines existing for what sort of OBC specifications the multispectral models would demand, the dataset was designed to be fairly limited in image size. 128 by 128-pixel images were not selected because they were necessarily the desired size, but rather the resultant model would increase in size and the models produced might have been unable to run on the hardware. Originally, there were 8

OBC configurations planned, but was reduced to 4 due to difficulties setting the older boards up. The NVIDIA TX2 and more notably the Nvidia Jetson Nano with 4GB of VRAM would have been more restrictive platforms. Returning to the original dataset creation, with an idea of what computational resources were required of the simulated SBC, the dataset might have been adjusted to feature larger imagery and model sizes and process faster as a result.

5.1.2.4 Model Structure

While not a criticism of Ultralytics's work on the YOLOv8 model, it's important to recognize the model was not designed for multispectral satellite-based vessel detection. RGB object detection from terrestrial or near-terrestrial based images was the objective of YOLOv8. The adapatations to convert it to a multispectral model was the change of the input layer from 3 to a variable number. A purpose-built lightweight model for multispectral imaging might offer better performance and could feature optimizations designed to improve the speed. Similarly, balancing the model size and complexity with performance for context-sensitive optimization aboard limited satellite hardware needs to be explored further. Satellite OBC's which make use of the trialed board may utilize different frequencies and configurations for their systems, better suited for their intended role.

Bibliography

[1] Canada Centre for Remote Science - Natural Resources Canada, "Fundamentals of Remote Sensing," [Online]. Available: https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf. [Accessed 17 January 2022].

[2] K. Tewari, "A Review of Climate Change Impact Studies on Harmful Algal Blooms," *Phycology,* vol. 2(2), pp. 244-253, 2022.

[3] J. Park, J. V. Osdel, J. Turner, C. M. Farthing, N. A. Miller, H. L. Linder, G. O. Crespo, G. Carmine and D. A. Kroodsma, "Tracking elusive and shifting identities of the global fishing fleet," *ScienceAdvances,* vol. 9, no. 3, 2023.

[4] H. Welch, T. Clavelle, D. T. White, M. A. Cimino, J. V. Osdel, T. Hochberg, D. Kroodsma and E. L. Hazen, "Hot spots of unseen fishing vessels," *ScienceAdvances,* vol. 8, no. 44, 2022.

[5] U. Kanjir, H. Greidanus and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sensing of Enviroment,* vol. 2017, pp. 1-26, 2018.

[6] C. Burkhard and S. Weston, "The Evolution of CubeSat Spacecraft Platforms," NASA Ames Research Centre, Moffett Field, CA, USA, 2021.

[7] I. Taleb and A. Lassakeur, "A Survey of Compact Optical Cameras for Earth Observation CubeSat Missions," in *73rd International Astronautical Congress (IAC)*, Paris, France, 2022.

[8] G. Lentaris, K. Maragos, I. Stratakos, L. Papadopoulos, O. Papanikolaou, D. Soudris, M. Lourakis, X. Zabulis, D. Gonzalez-Arjona and G. Furano, "High-Performance Embedded Computing in Space: Evaluation of Platforms for Vision-Based Navigation," *Journal of Aerospace Information Systems,* 2018.

[9] G. Giuffrida, L. Fanucci, G. Meoni, M. Batic, L. Buckley, A. Dunne, C. V. Dijk, M. Esposito, J. Hefele, N. Vercruyssen, G. Furano, M. Pastena and J. Aschbacher, "The Φ-Sat-1 Mission: The First On-Board Deep Neural Network Demonstrator for Satellite Earth Observation," *IEE Transactions on Geoscience and Remote Sensing,* vol. 60, 2022.

[10] J. Park, J. Lee, K. Seto, T. Hochberg, B. A. Wong, N. A. Miller, K. Takasaki, H. Kubota, Y. Oozeki, S. Doshi, M. Midzik, Q. Hanich, B. Sullivan, P. Woods and D. A. Kroodsma, "Illuminating dark fishing fleets in North Korea," *SCIENCE ADVANCES,* vol. 6, no. 30, 2020.

[11] C. Chiego, "Maritime Cybersecurity: AIS Manipulation Motivations in the Maritime Domain," 2022.

[12] NATO Shipping Centre , "AIS (Automatic Identification System) overview," 2021. [Online]. Available: https://shipping.nato.int/nsc/operations/news/2021/ais-automatic-identification-system-overview#:~:text=The%20Automatic%20Identification%20System%20(AIS,both%20vessels%20and%20shore%20stations.. [Accessed 18 12 2023].

[13] International Maritime Organization, *Resolution A.917(22) - GUIDELINES FOR THE ONBOARD OPERATIONAL USE OF SHIPBORNE AUTOMATIC IDENTIFICATION SYSTEMS (AIS),* 2002.

[14] Bureau of Ocean Energy Management (BOEM) and National Oceanic and Atmospheric Administration (NOAA)., "Vessel Traffic Data," MarineCadastre.gov, [Online]. Available: https://marinecadastre.gov/ais/. [Accessed 28 September 2022].

[15] C. Triebert, B. Migliozzi, A. Cardia, M. Xiao and a. D. Botti, "Fake Signals and American Insurance: How a Dark Fleet Moves Russian Oil," The New York Times, 30 May 2023. [Online]. Available: https://www.nytimes.com/interactive/2023/05/30/world/asia/russia-oil-ships-sanctions.html. [Accessed 23 November 2023].

[16] United Nations Conference on Trade and Development, "Review of Maritime Transport 2018," United Nations Conference on Trade and Development, Geneva, Switzerland , 2018.

[17] FAO, "The State of World Fisheries and Aquaculture," Food and Agriculture Organization of the United Nations, Rome, 2022.

[18] D. A. Kroodsma, T. Hochberg, P. B. Davis, F. S. Paolo, R. Joo and B. A. Wong, "Revealing the global longline fleet with satellite radar," *Scientific Reports,* vol. 12, 2022.

[19] S. Widjaja, T. Long, H. Wirajuda and e. al, "Illegal, Unreported and Unregulated Fishing and Associated Drivers," World Resources Institute, Washington, 2020.

[20] N. A. Miller, A. Roan, T. Hochberg, J. Amos and D. A. Kroodsma, "Identifying Global Patterns of Transshipment Behavior," *Frontiers in Marine Science,* vol. 5, p. 240, 2018.

[21] T. Clavelle, "Hotspots of Unseen Fishing Vessels Illuminate Areas of Concern for Illegal, Unreported and Unregulated Fishing," Global Fishing Watch, 2 November 2022. [Online]. Available: https://globalfishingwatch.org/research/hotspots-of-unseen-fishing-vessels-qa/. [Accessed 23 November 2023].

[22] UN. Security Council President; UN. Panel of Experts pursuant to Security Council Resolution 1874 (2009), "Note [transmitting final report of the Panel of Experts Established pursuant to Security Council Resolution 1874 (2009) concerning the Democratic People's Republic of Korea]," United Nations, New York, 2021.

[23] J. C. Iacarella, G. Clyde and A. Dunham, "Vessel Tracking Datasets For Monitoring Canada’s Conservation Efforts," Fisheries and Oceans Canada, Sidney, BC, 2020.

[24] MIS Marine, "The Use of AIS Data to Identify Dark Activity for Marine Auditing Purposes," MIS Marine, Birmingham, United Kingdom, 2020.

[25] H. Welch, T. Clavelle, T. D. White, M. A. Cimino, J. V. Osdel, T. Hochberg, D. Kroodsma and E. L. Hazen, "Hot spots of unseen fishing vessels," *ScienceAdvances,* vol. 8, no. 44, 2022.

[26] S. Voinov, "Deep Learning-based Vessel Detection from Very High and Medium Resolution Optical Satellite Images as Component of Maritime Surveillance Systems," Universität Rostock, Rostock, 2020.

[27] W. Wang, J. Ren and M. Huang, "Ship Detection in Multispectral Remote Sensing Images via Saliency Analysis," *Applied Ocean Research,* vol. 106, 2021.

[28] M. F. Fingas and C. E. Brown, "Review of Ship Detection from Airborne Platforms," *Canadian Journal of Remote Sensing,* vol. 27, no. 4, pp. 379-385, 2001.

[29] G. Rowlands, J. Brown, B. Soule, P. T. Boluda and A. Rogers, "Satellite Surveillance of Fishing Vessel Activity in the Ascension Island Exclusive Economic Zone and Marine Protected Area," *Marine Policy,* vol. 101, pp. 39-50, 2019.

[30] European Space Agency, "Sentinel-1 SAR User Guide," European Space Agency, [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar. [Accessed 3 January 2024].

[31] L. Yiding, S. Zhang and W.-Q. Wang, "A lightweight faster R-CNN ship detection," *IEE Geoscience and remote sensing letters (2022),* vol. 19, pp. 1-5, 2022.

[32] M. I. Pettersson, V. T. Vu and T. K. Sjögren, "New Forms of Likelihood Ratio Test for SAR," *IEEE Access,* vol. 9, pp. 127906-127916, 2021.

[33] M. Ma, J. Chen, W. Liu and W. Yang, "Ship Classification and Detection Based on CNN Using GF-3 SAR Images," *Remote Sensing,* vol. 10, no. 12, 2018.

[34] T. Zhang and X. Zhang, "High-Speed Ship Detection in SAR Images Based on a Grid Convolutional Neural Network," *Remote Sensing,* vol. 11, no. 10, 2019.

[35] ICEYE, "Completeting the Picture, SAR Product Guide," 1 May 2021. [Online]. Available: https://earth.esa.int/eogateway/documents/20142/37627/ICEYE-SAR-Product-Guide-V4.pdf. [Accessed 18th December 2023].

[36] K. Hirako, S. Shirasaka, T. Obata, S. Nakasuka, H. Saito, S. Nakamura and T. Tohara, "Development of small satellite for X-Band compact," in *Journal of Physics: Conference Series*, Jakarta, Indonesia, 2018.

[37] K.-A. Park, J.-J. Park, J.-C. Jang, J.-H. Lee, S. Oh and M. Lee, "Multi-Spectral Ship Detection Using Optical, Hyperspectral, and Microwave SAR Remote Sensing Data in Coastal Regions," *Sustainability,* vol. 10, no. 11, p. 4064, 2018.

[38] M. Zhao, Y. Zhou, X. Li, W. Cao, C. He, B. Yu, X. Li, C. Elvidge, W. Cheng and C. Zhou, "Applications of Satellite Remote Sensing of Nighttime Light Observations," *Remote Sensing,* vol. 11, no. 17, 2019.

[39] R. Bhattarai, P. Rahimzadeh-Bajgiran, A. Weiskittel and D. A. MacLean, "Sentinel-2 based prediction of spruce budworm defoliation using red-edge spectral vegetation indices," *Remote Sensing Letters,* vol. 11, no. 8, pp. 777-786, 2020.

[40] J. Bramich, B. J. S. Christopher and A. Fischer, "Improved red-edge chlorophyll-a detection for Sentinel 2," *Ecological Indicators,* vol. 120, January 2021.

[41] European Space Agency, "Sentinel-2 MSI User Guide," [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi. [Accessed 04 12 2023].

[42] Y. Zhong, X. Wang, S. Wang and L. Zhang, "Advances in spaceborne hyperspectral remote sensing in China," *Geo-Spatial Information Science,* vol. 24, no. 1, pp. 95-120, 2021.

[43] A. Ciocarlan and A. Stoian, "Ship Detection in Sentinel 2 Multi-Spectral Images with Self-Supervised Learning," *Remote Sens.,* vol. 13, no. 21: 4255, 2021.

[44] H. J. Kramer and A. P. Cracknell, "An overview of small satellites in remote sensing," *International Journal of Remote Sensing,* vol. 29, no. 15, 2008.

[45] R. Botelho and A. Xavier, "A Unified Satellite Taxonomy Proposal Based on Mass and Size," *Advances in Aerospace Science and Technology,* vol. 4, no. 4, 2019.

[46] Small Spacecraft Systems Virtual Institute, "State-of-the-Art Small Spacecraft Technology," NASA, Moffett Field, California, 2020.

[47] J. McDowell, "General Catalog of Artificial Space Objects," 2023. [Online]. Available: https://planet4589.org/space/gcat/. [Accessed 14 Jan 2024].

[48] Bryce Tech, "Smallsats by the Numbers," Bryce Tech, Alexandria, Virginia, 2023.

[49] D. Querejazu and L. Randazzese, "Small satellites - Big Missions," CSIS, Washington, D.C., 2017.

[50] O. Dubovik, G. L. Schuster, F. Xu, Y. Hu, H. Bosch and Z. Li, "Grand Challenges in Satellite Remote Sensing," *Frontiers in Remote Sensing,* vol. 2, January 2021.

[51] R.Sandau and K. Brieß, "Potential for advancements in remote sensing using small satellites," in *ISPRS Congress*, Bejing, China, 2008.

[52] X. Luo, M. Wang, G. Dai and X. Chen, "A Novel Technique to Compute the Revisit Time of Satellites and Its Application in Remote Sensing Satellite Optimization Design," *International Journal of Aerospace,* vol. 2017, p. 9, 2017.

[53] A. P. Trishchenko, L. Garand and a. L. D. Trichtchenko, "Observing Polar Regions from Space: Comparison between Highly Elliptical Orbit and Medium Earth Orbit Constellations," *Journal of Atmospheric and Oceanic Technology,* pp. 1605-1621, 2019.

[54] H. W. Jones, "The Recent Large Reduction in Space Launch Cost," in *International Conference on Environmental Systems*, Albuquerque, New Mexico, 2018.

[55] N. Crisp, K. Smith and P. Hollingsworth, "Small Satellite Launch to LEO: A Review of Current and Future Launch Systems," *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan,* vol. 12, no. 29, pp. 39-47, 2014.

[56] SatCatalog, "CubeSat Launch Costs," SatCatalog, 2024. [Online]. Available: https://satcatalog.com/insights/cubesat-launch-costs/. [Accessed 30 April 2024].

[57] EnduroSat, "EnduroSat 3U Platform," EnduroSat, [Online]. Available: https://endurosat.com/products/3u-platform/. [Accessed 30 April 2024].

[58] J. W. Blackwell and S. Braun, "Overview of the NASA TROPICS CubeSat Constellation," 2 Nov 2023. [Online]. Available: https://ntrs.nasa.gov/api/citations/20190001205/downloads/20190001205.pdf.

[59] S. Yavari, "RADARSAT Constellation Mission: From Design to Launch to Operation," GoGeomatics, 17 February 2022. [Online]. Available: https://gogeomatics.ca/radarsat-constellation-mission-from-design-to-launch-to-operation/. [Accessed 30 April 2024].

[60] P. Zong and S. Kohani, "Optimal Satellite LEO Constellation Design Based on Global Coverage in One Revisit Time," *International Journal of Aerospace Engineering,* p. 12, 2019.

[61] J. A. Nichols, H. W. H. Chan and M. A. B. Baker, "Machine learning: applications of artificial intelligence to imaging and diagnosis," *Biophys Reviews,* vol. 11, no. 1, pp. 111-118, 2019.

[62] Intel, "Intel Powers First Satellite with AI on Board," 20 October 2020. [Online]. Available: https://www.intel.com/content/www/us/en/newsroom/news/first-satellite-ai.html#gs.1rm5m9. [Accessed 18 12 2023].

[63] SmartSat CRC, "Machine Learning Onboard Satellites," SmartSat, Adelaide, Australia, 2021.

[64] ACC Clyde Space, "PULSAR HSTX," ACC Clyde Space, [Online]. Available: https://www.aac-clyde.space/what-we-do/space-products-components/communications/pulsar-hstx. [Accessed 30 April 2024].

[65] Paradigma Technologies, "K Band Transmitter," Paradigma Technologies, [Online]. Available: https://paradigma-tech.com/k-band-transmitter/. [Accessed 30 April 2024].

[66] J. J. Senecal, J. W. Sheppard and J. A. Shaw, "Efficient Convolutional Neural Networks for Multi-Spectral Image Classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019.

[67] S. Ren, G. Chen, T. Li, Q. Chen and S. Li, "A Deep Learning-Based Computational Algorithm for Identifying Damage Load Condition: An Artificial Intelligence Inverse Problem Solution for Failure Analysis," *Computer Modeling in Engineering and Sciences,* vol. 117, no. 3, pp. 287-307, 2018.

[68] D. Helder, B. Markham, R. Morfitt, J. Storey, J. Barsi, F. Gascon, S. Clerc, B. LaFrance, J. Masek, D. Roy, A. Lewis and N. Pahlevan, "Observations and Recommendations for the Calibration of Landsat 8 OLI and Sentinel 2 MSI for Improved Data Interoperability," *Remote Sensing,* vol. 10, no. 9, 2018.

[69] R. Ruiloba, F. D. Vieilleville, A. Lagrange and B. Saux, "Sentinel-2 dataset for ship detection," 30 Jun. 2020.

[70] S. Gilles and Others, *Rasterio: geospatial raster I/O for {Python} programmers,* 2013--.

[71] Danish Maritime Authority, "AIS Data," Danish Maritime Authority, [Online]. Available: https://dma.dk/safety-at-sea/navigational-information/ais-data#:~:text=3.-,Historical%20AIS%20data,into%20a%20user%20friendly%20presentation.. [Accessed 27 September 2022].

[72] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, 2019.

[73] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 6, pp. 1137 - 1149, 2016.

[74] G. Jocher, A. Chaurasia and J. Qiu, *Ultralytics YOLO,* 8.0.0 ed., 2023.

[75] A. Murray, "python-pillow/Pillow: 10.2.0," *Zenodo,* 02 01 2024.

[76] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools,* 2000.

[77] R. King, "Brief summary of YOLOv8 model structure - Issue 189," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics/issues/189.

[78] NVIDIA, "Nvidia Jetson Developer Guide," NVIDIA, 14 August 2023. [Online]. Available: https://docs.nvidia.com/jetson/archives/r35.4.1/DeveloperGuide/index.html.

[79] GEBCO, "Gridded Bathymetry Data," GEBCO, 2023. [Online]. Available: https://www.gebco.net/data_and_products/gridded_bathymetry_data/.

[80] OpenStreetMap, "OpenStreetMap," 2023. [Online]. Available: https://www.openstreetmap.org/.

[81] A. Pellegrino, M. G. Pancalli, A. Gianfermo, P. Marzioli, F. Curiano, F. Angeletti, F. Piergentili and F. Santoni, "HORUS: Multispectral and Multiangle CubeSat Mission Targeting Sub-Kilometer Remote Sensing Applications," *Remote Sensing,* vol. 13, 2021.

[82] J. Transon, R. d'Andrimont and P. Defourny, "Survey of Hyperspectral Earth Observation Applications from Space in the Sentienel-2 Context," *Remote Sensing,* vol. 10, no. 2, 2018.

[83] J. Transon, R. D'Andrimont, A. Maugnard and P. Defourny, "Survey of Hpyerspectral Earth Observation Applications from Space in the Sentinel-2 Context," *Remote Sensing,* vol. 10, no. 2, p. 157, 2018.

[84] H. Dierssen, S. Ackleson, K. Joyce, E. Hestir, A. Castagna, S. Lavender and M. McManus, "Living up to the Hype of HypeSpectral Aquatic Remote Sensing: Science, Resources and Outlook," *Frontiers in Enviromental Science,* vol. 9, 2021.

[85] B. Lal, E. d. l. R. Blanco, J. R. Behrens, B. A. Corbin, E. K. Green, A. J. Picard and A. Balakrishnan, "Global Trends in Small Satellites," Science and Technology Policy Institute., Washington DC, 2017.

[86] K. O'Donnell and G. Richardson, "Small Satellite Trending & Reliability 2009-2018," in *Small Satellite Confrence*, Logan Utah, 2020.

[87] S. C. Burleigh, T. D. Cola, S. Morosi, S. Jayousi, E. Cianca and C. Fuchs, "From connectivity to advanced internet services A Comprehensive Review of Small Satellites Communications and Networks," *Wireless Communications and Mobile Computing,* vol. 2019, p. 17, 2019.

[88] S. Nag, J. LeMoigne and O. d. Weck, "Cost and Risk Analysis of Small Satellite Constellations for Earth Observation," in *2014 IEEE Aerospace Conference*, Big Sky, Montana, 2014.

[89] ICEYE, "Detecting Dark Vessels," Iceye, Espoo, Finland, 2019.

[90] K. Hirako, S. Shirasaka, T. Obata, S. Nakasuka, H. Saito, S. Nakamura and T. Tohara, "Development of small satellite for X-Band compact synthetic aperture radar," *Journal of Physics: Conference Series,* p. NEEDS TO BE CHECKED, 2018.

[91] M. Song, H. Qu, G. Zhang and G. Jin, "Detection of small ship targets from an optical remote sensing image," *Frontiers of Optoelectronics,* vol. 11, no. 3, pp. 275-284, 2018.

[92] H. J. Kramer and A. P. Cracknell, "An overview of small satellites in remote sensing," *International Journal of Remote Sensing,* vol. 29, no. 15, pp. 4285-4337, 2008.

Appendix A Model Performance Results

			Vessel Confusion Matrix 198 Vessels								
		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
AllBand	M	0	127	16	71	0	198	0.8881	0.6414	0.5935	0.7449
	M	0.4	120	10	78	0	198	0.9231	0.6061	0.5769	0.7317
	M	0.5	117	8	81	0	198	0.9360	0.5909	0.5680	0.7245
	M	0.6	112	5	86	0	198	0.9573	0.5657	0.5517	0.7111
	M	0.7	99	3	99	0	198	0.9706	0.5000	0.4925	0.6600
	M	0.8	61	1	137	0	198	0.9839	0.3081	0.3065	0.4692
AllBand	N	0	122	15	76	0	198	0.8905	0.6162	0.5728	0.7284
	N	0.4	109	7	89	0	198	0.9397	0.5505	0.5317	0.6943
	N	0.5	101	3	97	0	198	0.9712	0.5101	0.5025	0.6689
	N	0.6	93	3	105	0	198	0.9688	0.4697	0.4627	0.6327
	N	0.7	71	2	127	0	198	0.9726	0.3586	0.3550	0.5240
	N	0.8	28	0	170	0	198	1.0000	0.1414	0.1414	0.2478
AllBand	S	0	128	14	70	0	198	0.9014	0.6465	0.6038	0.7529
	S	0.4	118	7	80	0	198	0.9440	0.5960	0.5756	0.7307
	S	0.5	113	5	85	0	198	0.9576	0.5707	0.5567	0.7152
	S	0.6	103	4	95	0	198	0.9626	0.5202	0.5099	0.6754
	S	0.7	85	2	113	0	198	0.9770	0.4293	0.4250	0.5965
	S	0.8	38	0	160	0	198	1.0000	0.1919	0.1919	0.3220
AllBand	L	0	133	19	65	0	198	0.8750	0.6717	0.6129	0.7600
	L	0.4	125	12	73	0	198	0.9124	0.6313	0.5952	0.7463
	L	0.5	120	8	78	0	198	0.9375	0.6061	0.5825	0.7362
	L	0.6	117	7	81	0	198	0.9435	0.5909	0.5707	0.7267
	L	0.7	107	5	91	0	198	0.9554	0.5404	0.5271	0.6903
	L	0.8	73	1	125	0	198	0.9865	0.3687	0.3668	0.5368
RGB	M	0	124	15	74	0	198	0.8921	0.6263	0.5822	0.7359
	M	0.4	116	7	82	0	198	0.9431	0.5859	0.5659	0.7227
	M	0.5	113	7	85	0	198	0.9417	0.5707	0.5512	0.7107
	M	0.6	104	2	94	0	198	0.9811	0.5253	0.5200	0.6842
	M	0.7	83	2	115	0	198	0.9765	0.4192	0.4150	0.5866
	M	0.8	51	0	147	0	198	1.0000	0.2576	0.2576	0.4096
RGB	N	0	114	10	84	0	198	0.9194	0.5758	0.5481	0.7081
	N	0.4	100	1	98	0	198	0.9901	0.5051	0.5025	0.6689
	N	0.5	91	1	107	0	198	0.9891	0.4596	0.4573	0.6276
	N	0.6	81	0	117	0	198	1.0000	0.4091	0.4091	0.5806
	N	0.7	63	0	135	0	198	1.0000	0.3182	0.3182	0.4828
	N	0.8	27	0	171	0	198	1.0000	0.1364	0.1364	0.2400
RGB	S	0	111	8	87	0	198	0.9328	0.5606	0.5388	0.7003

Vessel Confusion Matrix
198 Vessels

		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
	S	0.4	102	5	96	0	198	0.9533	0.5152	0.5025	0.6689
	S	0.5	95	1	103	0	198	0.9896	0.4798	0.4774	0.6463
	S	0.6	91	1	107	0	198	0.9891	0.4596	0.4573	0.6276
	S	0.7	81	1	117	0	198	0.9878	0.4091	0.4070	0.5786
	S	0.8	53	1	145	0	198	0.9815	0.2677	0.2663	0.4206
RGB	L	0	126	12	72	0	198	0.9130	0.6364	0.6000	0.7500
	L	0.4	118	10	80	0	198	0.9219	0.5960	0.5673	0.7239
	L	0.5	113	8	85	0	198	0.9339	0.5707	0.5485	0.7085
	L	0.6	108	7	90	0	198	0.9391	0.5455	0.5268	0.6901
	L	0.7	99	4	99	0	198	0.9612	0.5000	0.4901	0.6578
	L	0.8	76	1	122	0	198	0.9870	0.3838	0.3819	0.5527
10+20m	M	0	128	12	70	0	198	0.9143	0.6465	0.6095	0.7574
	M	0.4	123	9	75	0	198	0.9318	0.6212	0.5942	0.7455
	M	0.5	113	7	85	0	198	0.9417	0.5707	0.5512	0.7107
	M	0.6	105	7	93	0	198	0.9375	0.5303	0.5122	0.6774
	M	0.7	94	2	104	0	198	0.9792	0.4747	0.4700	0.6395
	M	0.8	57	1	141	0	198	0.9828	0.2879	0.2864	0.4453
10+20m	N	0	111	6	87	0	198	0.9487	0.5606	0.5441	0.7048
	N	0.4	105	3	93	0	198	0.9722	0.5303	0.5224	0.6863
	N	0.5	99	2	99	0	198	0.9802	0.5000	0.4950	0.6622
	N	0.6	91	1	107	0	198	0.9891	0.4596	0.4573	0.6276
	N	0.7	78	0	120	0	198	1.0000	0.3939	0.3939	0.5652
	N	0.8	29	0	169	0	198	1.0000	0.1465	0.1465	0.2555
10+20m	S	0	121	7	77	0	198	0.9453	0.6111	0.5902	0.7423
	S	0.4	111	5	87	0	198	0.9569	0.5606	0.5468	0.7070
	S	0.5	106	4	92	0	198	0.9636	0.5354	0.5248	0.6883
	S	0.6	98	2	100	0	198	0.9800	0.4949	0.4900	0.6577
	S	0.7	76	2	122	0	198	0.9744	0.3838	0.3800	0.5507
	S	0.8	24	0	174	0	198	1.0000	0.1212	0.1212	0.2162
10+20m	L	0	135	22	63	0	198	0.8599	0.6818	0.6136	0.7606
	L	0.4	130	12	68	0	198	0.9155	0.6566	0.6190	0.7647
	L	0.5	123	9	75	0	198	0.9318	0.6212	0.5942	0.7455
	L	0.6	116	6	82	0	198	0.9508	0.5859	0.5686	0.7250
	L	0.7	103	5	95	0	198	0.9537	0.5202	0.5074	0.6732
	L	0.8	74	1	124	0	198	0.9867	0.3737	0.3719	0.5421
10+60m_Old	M	0	126	16	72	0	198	0.8873	0.6364	0.5888	0.7412
	M	0.4	120	13	78	0	198	0.9023	0.6061	0.5687	0.7251
	M	0.5	115	7	83	0	198	0.9426	0.5808	0.5610	0.7188
	M	0.6	111	5	87	0	198	0.9569	0.5606	0.5468	0.7070

Vessel Confusion Matrix
198 Vessels

		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
	M	0.7	99	3	99	0	198	0.9706	0.5000	0.4925	0.6600
	M	0.8	58	1	140	0	198	0.9831	0.2929	0.2915	0.4514
20+60m	M	0	129	22	69	0	198	0.8543	0.6515	0.5864	0.7393
	M	0.4	124	16	74	0	198	0.8857	0.6263	0.5794	0.7337
	M	0.5	119	13	79	0	198	0.9015	0.6010	0.5640	0.7212
	M	0.6	115	8	83	0	198	0.9350	0.5808	0.5583	0.7165
	M	0.7	97	4	101	0	198	0.9604	0.4899	0.4802	0.6488
	M	0.8	53	1	145	0	198	0.9815	0.2677	0.2663	0.4206
RGB+B01	M	0	118	10	80	0	198	0.9219	0.5960	0.5673	0.7239
	M	0.4	113	8	85	0	198	0.9339	0.5707	0.5485	0.7085
	M	0.5	104	6	94	0	198	0.9455	0.5253	0.5098	0.6753
	M	0.6	100	3	96	0	196	0.9709	0.5102	0.5025	0.6689
	M	0.7	79	2	119	0	198	0.9753	0.3990	0.3950	0.5663
	M	0.8	36	0	162	0	198	1.0000	0.1818	0.1818	0.3077
RGB+B05	M	0	128	18	70	0	198	0.8767	0.6465	0.5926	0.7442
	M	0.4	120	11	78	0	198	0.9160	0.6061	0.5742	0.7295
	M	0.5	115	7	83	0	198	0.9426	0.5808	0.5610	0.7188
	M	0.6	110	4	88	0	198	0.9649	0.5556	0.5446	0.7051
	M	0.7	87	3	111	0	198	0.9667	0.4394	0.4328	0.6042
	M	0.8	35	0	163	0	198	1.0000	0.1768	0.1768	0.3004
RGB+B06	M	0	129	20	69	0	198	0.8658	0.6515	0.5917	0.7435
	M	0.4	119	13	79	0	198	0.9015	0.6010	0.5640	0.7212
	M	0.5	114	6	84	0	198	0.9500	0.5758	0.5588	0.7170
	M	0.6	111	3	87	0	198	0.9737	0.5606	0.5522	0.7115
	M	0.7	94	2	104	0	198	0.9792	0.4747	0.4700	0.6395
	M	0.8	49	0	149	0	198	1.0000	0.2475	0.2475	0.3968
RGB+B07	M	0	130	19	68	0	198	0.8725	0.6566	0.5991	0.7493
	M	0.4	121	9	77	0	198	0.9308	0.6111	0.5845	0.7378
	M	0.5	116	7	82	0	198	0.9431	0.5859	0.5659	0.7227
	M	0.6	112	3	86	0	198	0.9739	0.5657	0.5572	0.7157
	M	0.7	96	3	102	0	198	0.9697	0.4848	0.4776	0.6465
	M	0.8	68	0	130	0	198	1.0000	0.3434	0.3434	0.5113
RGB+B08	M	0	130	23	68	0	198	0.8497	0.6566	0.5882	0.7407
	M	0.4	124	13	74	0	198	0.9051	0.6263	0.5877	0.7403
	M	0.5	121	9	77	0	198	0.9308	0.6111	0.5845	0.7378
	M	0.6	115	4	83	0	198	0.9664	0.5808	0.5693	0.7256
	M	0.7	97	3	101	0	198	0.9700	0.4899	0.4826	0.6510
	M	0.8	31	0	167	0	198	1.0000	0.1566	0.1566	0.2707
RGB+B09	M	0	125	13	73	0	198	0.9058	0.6313	0.5924	0.7440

Vessel Confusion Matrix
198 Vessels

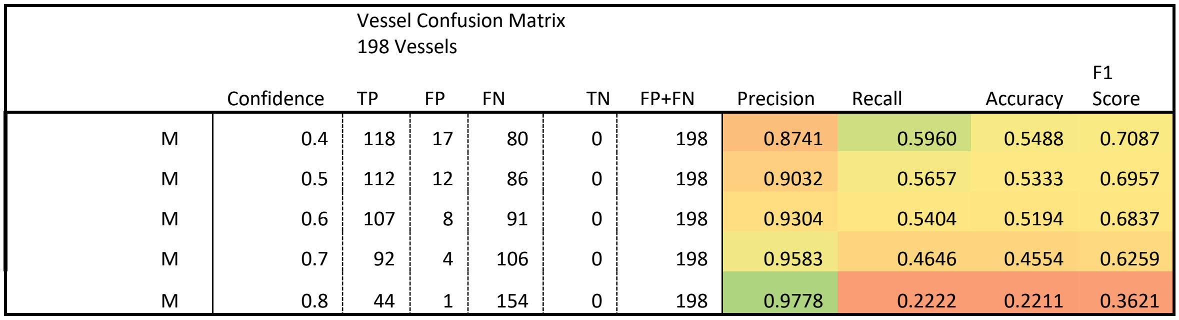
		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
	M	0.4	117	8	81	0	198	0.9360	0.5909	0.5680	0.7245
	M	0.5	109	6	89	0	198	0.9478	0.5505	0.5343	0.6965
	M	0.6	100	4	98	0	198	0.9615	0.5051	0.4950	0.6623
	M	0.7	83	2	115	0	198	0.9765	0.4192	0.4150	0.5866
	M	0.8	39	0	159	0	198	1.0000	0.1970	0.1970	0.3291
RGB+B10	M	0	58	23	140	0	198	0.7160	0.2929	0.2624	0.4158
	M	0.4	46	6	152	0	198	0.8846	0.2323	0.2255	0.3680
	M	0.5	30	0	168	0	198	1.0000	0.1515	0.1515	0.2632
	M	0.6	13	0	185	0	198	1.0000	0.0657	0.0657	0.1232
	M	0.7	0	0	198	0	198	#DIV/0!	0.0000	0.0000	#DIV/0!
	M	0.8	0	0	198	0	198	#DIV/0!	0.0000	0.0000	#DIV/0!
RGB+B11	M	0	127	15	71	0	198	0.8944	0.6414	0.5962	0.7471
	M	0.4	121	11	77	0	198	0.9167	0.6111	0.5789	0.7333
	M	0.5	117	9	81	0	198	0.9286	0.5909	0.5652	0.7222
	M	0.6	109	5	89	0	198	0.9561	0.5505	0.5369	0.6987
	M	0.7	95	2	103	0	198	0.9794	0.4798	0.4750	0.6441
	M	0.8	45	1	153	0	198	0.9783	0.2273	0.2261	0.3689
RGB+B12	M	0	128	18	70	0	198	0.8767	0.6465	0.5926	0.7442
	M	0.4	122	10	76	0	198	0.9242	0.6162	0.5865	0.7394
	M	0.5	115	7	83	0	198	0.9426	0.5808	0.5610	0.7188
	M	0.6	108	3	90	0	198	0.9730	0.5455	0.5373	0.6990
	M	0.7	99	2	99	0	198	0.9802	0.5000	0.4950	0.6622
	M	0.8	64	1	134	0	198	0.9846	0.3232	0.3216	0.4867
RGB+08A	M	0	131	18	67	0	198	0.8792	0.6616	0.6065	0.7550
	M	0.4	121	9	77	0	198	0.9308	0.6111	0.5845	0.7378
	M	0.5	118	7	80	0	198	0.9440	0.5960	0.5756	0.7307
	M	0.6	111	4	87	0	198	0.9652	0.5606	0.5495	0.7093
	M	0.7	97	2	101	0	198	0.9798	0.4899	0.4850	0.6532
	M	0.8	45	1	153	0	198	0.9783	0.2273	0.2261	0.3689
All Band - B01	M	0	126	13	72	0	198	0.9065	0.6364	0.5972	0.7478
	M	0.4	119	7	79	0	198	0.9444	0.6010	0.5805	0.7346
	M	0.5	115	6	83	0	198	0.9504	0.5808	0.5637	0.7210
	M	0.6	108	4	90	0	198	0.9643	0.5455	0.5347	0.6968
	M	0.7	93	2	105	0	198	0.9789	0.4697	0.4650	0.6348
	M	0.8	51	0	147	0	198	1.0000	0.2576	0.2576	0.4096
All Band - B02	M	0	127	19	71	0	198	0.8699	0.6414	0.5853	0.7384
	M	0.4	117	10	81	0	198	0.9213	0.5909	0.5625	0.7200
	M	0.5	112	7	86	0	198	0.9412	0.5657	0.5463	0.7066
	M	0.6	105	4	93	0	198	0.9633	0.5303	0.5198	0.6840

Vessel Confusion Matrix
198 Vessels

		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
	M	0.7	91	3	107	0	198	0.9681	0.4596	0.4527	0.6233
	M	0.8	47	1	151	0	198	0.9792	0.2374	0.2362	0.3821
All Band - B03	M	0	128	16	70	0	198	0.8889	0.6465	0.5981	0.7485
	M	0.4	120	9	78	0	198	0.9302	0.6061	0.5797	0.7339
	M	0.5	117	8	81	0	198	0.9360	0.5909	0.5680	0.7245
	M	0.6	114	5	84	0	198	0.9580	0.5758	0.5616	0.7192
	M	0.7	99	4	99	0	198	0.9612	0.5000	0.4901	0.6578
	M	0.8	40	0	158	0	198	1.0000	0.2020	0.2020	0.3361
All Band - B04	M	0	125	15	73	0	198	0.8929	0.6313	0.5869	0.7396
	M	0.4	123	11	75	0	198	0.9179	0.6212	0.5885	0.7410
	M	0.5	118	6	80	0	198	0.9516	0.5960	0.5784	0.7329
	M	0.6	113	5	85	0	198	0.9576	0.5707	0.5567	0.7152
	M	0.7	106	3	92	0	198	0.9725	0.5354	0.5274	0.6906
	M	0.8	64	1	134	0	198	0.9846	0.3232	0.3216	0.4867
All Band - B05	M	0	133	18	65	0	198	0.8808	0.6717	0.6157	0.7622
	M	0.4	122	10	76	0	198	0.9242	0.6162	0.5865	0.7394
	M	0.5	120	7	78	0	198	0.9449	0.6061	0.5854	0.7385
	M	0.6	109	6	89	0	198	0.9478	0.5505	0.5343	0.6965
	M	0.7	95	2	103	0	198	0.9794	0.4798	0.4750	0.6441
	M	0.8	58	0	140	0	198	1.0000	0.2929	0.2929	0.4531
All Band - B06	M	0	129	12	69	0	198	0.9149	0.6515	0.6143	0.7611
	M	0.4	125	7	73	0	198	0.9470	0.6313	0.6098	0.7576
	M	0.5	121	6	77	0	198	0.9528	0.6111	0.5931	0.7446
	M	0.6	113	6	85	0	198	0.9496	0.5707	0.5539	0.7129
	M	0.7	100	4	98	0	198	0.9615	0.5051	0.4950	0.6623
	M	0.8	49	2	149	0	198	0.9608	0.2475	0.2450	0.3936
All Band - B07	M	0	130	12	68	0	198	0.9155	0.6566	0.6190	0.7647
	M	0.4	121	9	77	0	198	0.9308	0.6111	0.5845	0.7378
	M	0.5	116	5	82	0	198	0.9587	0.5859	0.5714	0.7273
	M	0.6	110	2	88	0	198	0.9821	0.5556	0.5500	0.7097
	M	0.7	97	1	101	0	198	0.9898	0.4899	0.4874	0.6554
	M	0.8	69	1	129	0	198	0.9857	0.3485	0.3467	0.5149
All Band - B08	M	0	133	13	65	0	198	0.9110	0.6717	0.6303	0.7733
	M	0.4	124	9	74	0	198	0.9323	0.6263	0.5990	0.7492
	M	0.5	121	7	77	0	198	0.9453	0.6111	0.5902	0.7423
	M	0.6	114	6	84	0	198	0.9500	0.5758	0.5588	0.7170
	M	0.7	101	3	97	0	198	0.9712	0.5101	0.5025	0.6689

Vessel Confusion Matrix
198 Vessels

		Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
	M	0.8	47	1	151	0	198	0.9792	0.2374	0.2362	0.3821
All Band - B09	M	0	130	15	68	0	198	0.8966	0.6566	0.6103	0.7580
	M	0.4	122	8	76	0	198	0.9385	0.6162	0.5922	0.7439
	M	0.5	119	7	79	0	198	0.9444	0.6010	0.5805	0.7346
	M	0.6	111	5	87	0	198	0.9569	0.5606	0.5468	0.7070
	M	0.7	100	2	98	0	198	0.9804	0.5051	0.5000	0.6667
	M	0.8	66	1	132	0	198	0.9851	0.3333	0.3317	0.4981
All Band - B10	M	0	124	14	74	0	198	0.8986	0.6263	0.5849	0.7381
	M	0.4	120	8	78	0	198	0.9375	0.6061	0.5825	0.7362
	M	0.5	116	6	82	0	198	0.9508	0.5859	0.5686	0.7250
	M	0.6	109	4	89	0	198	0.9646	0.5505	0.5396	0.7010
	M	0.7	101	2	97	0	198	0.9806	0.5101	0.5050	0.6711
	M	0.8	63	1	135	0	198	0.9844	0.3182	0.3166	0.4809
All Band - B11	M	0	125	14	73	0	198	0.8993	0.6313	0.5896	0.7418
	M	0.4	119	8	79	0	198	0.9370	0.6010	0.5777	0.7323
	M	0.5	114	5	84	0	198	0.9580	0.5758	0.5616	0.7192
	M	0.6	110	2	88	0	198	0.9821	0.5556	0.5500	0.7097
	M	0.7	100	1	98	0	198	0.9901	0.5051	0.5025	0.6689
	M	0.8	37	0	161	0	198	1.0000	0.1869	0.1869	0.3149
All Band - B12	M	0	126	17	72	0	198	0.8811	0.6364	0.5860	0.7390
	M	0.4	118	11	80	0	198	0.9147	0.5960	0.5646	0.7217
	M	0.5	116	7	82	0	198	0.9431	0.5859	0.5659	0.7227
	M	0.6	112	6	86	0	198	0.9492	0.5657	0.5490	0.7089
	M	0.7	106	3	92	0	198	0.9725	0.5354	0.5274	0.6906
	M	0.8	83	1	115	0	198	0.9881	0.4192	0.4171	0.5887
All Band - B8A	M	0	127	16	71	0	198	0.8881	0.6414	0.5935	0.7449
	M	0.4	123	11	75	0	198	0.9179	0.6212	0.5885	0.7410
	M	0.5	120	8	78	0	198	0.9375	0.6061	0.5825	0.7362
	M	0.6	117	6	81	0	198	0.9512	0.5909	0.5735	0.7290
	M	0.7	107	2	91	0	198	0.9817	0.5404	0.5350	0.6971
	M	0.8	69	1	129	0	198	0.9857	0.3485	0.3467	0.5149
20m	M	0	124	12	74	0	198	0.9118	0.6263	0.5905	0.7425
	M	0.4	119	9	79	0	198	0.9297	0.6010	0.5749	0.7301
	M	0.5	113	6	85	0	198	0.9496	0.5707	0.5539	0.7129
	M	0.6	108	5	90	0	198	0.9558	0.5455	0.5320	0.6945
	M	0.7	97	3	101	0	198	0.9700	0.4899	0.4826	0.6510
	M	0.8	59	0	139	0	198	1.0000	0.2980	0.2980	0.4591
60m	M	0	126	23	72	0	198	0.8456	0.6364	0.5701	0.7262



Vessel Confusion Matrix
198 Vessels

	Confidence	TP	FP	FN	TN	FP+FN	Precision	Recall	Accuracy	F1 Score
M	0.4	118	17	80	0	198	0.8741	0.5960	0.5488	0.7087
M	0.5	112	12	86	0	198	0.9032	0.5657	0.5333	0.6957
M	0.6	107	8	91	0	198	0.9304	0.5404	0.5194	0.6837
M	0.7	92	4	106	0	198	0.9583	0.4646	0.4554	0.6259
M	0.8	44	1	154	0	198	0.9778	0.2222	0.2211	0.3621