UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Optical Remote Sensing of Oil Spills by using Machine Learning Methods in the Persian Gulf: A Multi-Class Approach

Martin H. Evenseth

EOM-3901 Master's thesis in energy, climate and environment 30 SP
June 2023

Abstract

Marine oil spills are harmful for the environment and costly for society. Coastal areas are particularly vulnerable since they provide habitats for organisms, animals and marine ecosystems. This thesis studied machine learning methods to classify thick oil in a multi-class case, using remotely sensed multi-spectral data in the Persian Gulf. The study area covers a large area between United Arab Emirates (UAE) and Iran. The dataset is extracted from 10 Sentinel-2 tiles on six spectral bands between 492 nm to 2202 nm. These images were annotated for four classes, namely thick oil, thin oil, ocean water and turbid water by using the Bonn Agreement to analyse true color composite images. A variety of machine learning methods were trained and evaluated using this dataset. Then a robustness evaluation was done by using selected machine learning methods on an independent dataset. Initially multiple machine learning methods were included; three decision trees, six K-Nearest Neighbor (KNN) models, two Artificial Neural Network (ANN) models, two Naive bayes models, and two discriminant models. Two KNN models and two ANN models were then picked for further evaluation. The results show that the fine KNN approach with two nearest neighbors had the best performance based on the computed statistical measures. However, the robustness evaluation showed that the tri-layered NN performed better. This thesis has shown that supervised machine learning with a multi-class approach can be used for oil spill monitoring using multi-spectral remote sensing data in the Persian Gulf.

Acknowledgements

Without the help and support from fantastic people I don't know what would have become of this thesis, thank you!

First, I would like to express my greatest gratitude to my supervisor at UiT, Katalin Blix. Thank you for your extensive knowledge about the subject, for your guidance with encouraging feedback, for your enthusiasm trying to recruit me to different seminars, for your continuous good mood, for proofreading my thesis and for your availability, always having time for questions and meetings.

I would like to thank Martine Espeseth and Hugo Isaksen at Kongsberg Satellite Services for their expertise and help composing ideas and putting together a framework for the thesis, as well as helping to identify oil slicks using the Bonn Agreement Oil Appearance Code. Thanks to Silje Birgitte Segrem Grue for help and for proofreading.

Thanks to my colleagues at TEOS who the last two months willingly have traded shifts with me, which made it easier to juggle between work and studies.

Finally I would like to thank my girlfriend, Julie Høie Nygård, for her emotional support and for taking extra care of me and our home in the final stages of the writing process, and our cat Arja for always reminding me to take breaks.

Martin H. Evenseth
Tromsø, June 2023.

Contents

List of Figures

List of Tables

Abbreviations

AC	Atmospheric Correction
ANN	Artificial Neural Network
AOI	Area Of Interest
BAOAC	Bonn Agreement Oil Appearance Code
BRANN	Bayesian Regularized Neural Network
CDOM	Colored Dissolved Organic Matter
DWH	DeepWater Horizon
EM	ElectroMagnetic
EVOS	Exxon Valdez Oil Spill
FBNN	Feed Backward Neural Network
FFNN	Feed Forward Neural Network
FNR	False Negative Rate
HSOS	Hebei Spirit Oil Spill
KNN	K-Nearest Neighbor
ML	Machine Learning
MLP	Multi Layer Perceptron
NDWI	Normalized Difference Water Index
NIR	Near InfraRed
NN	Neural Network
OA	Overall Accuracy
RBFN	Radial Basis Function Network
ReLU	Rectified linear unit
RF	Random Forest
RRC	Reyleigh Roughness Criterion
RMS	Root Mean Square
SAR	Synthetic Aperture Radar
SOM	kohonen's SelfOrganizing Map
SVM	Support Vector Machine
SWIR	Short-Wave InfraRed
TOA	Top Of Atmosphere
TPR	True Positive Rate
USGS	United States Geological Survey
VIS	VISible

1

Introduction

Marine anthropogenic oil spills have a negative impact on life and the surrounding environment [1, 2, 3]. This can affect ocean bottom habitats, sea corals and a wide range of organism including fish, birds and plankton [3]. If the oil reaches coastal areas the negative effects increases impacting habitats for organisms, animals and marine ecosystems [3]. Larger oil spills comes from catastrophic or accident events, such as the Exxon Valdez oil spill (EVOS) in 1989, the Hebei Spirit oil spill (HSOS) in 2007, and the Deepwater Horizon oil spill (DWH) in 2010 [3]. More often marine anthropogenic oil spills appears in smaller batches released from oil platforms, tankers or other vessels [4].

The effects are not limited to an environmental impact, it also has an economical impact on society. In case of an oil spill there is potential economic cost associated with assessing, monitoring and cleaning up the oil spill [4, 5].

In case of a marine anthropogentic oil spill, response teams needs information about the spill in order to efficiently react, limiting the spread and the environmental impact [4]. This can be done by visual inspection of the affected area onboard vessels and airplanes. Location and the size of an oil spill is important considerations for a response team. Potentially the oil spill can be located in the middle of a large ocean, and spread out over a large area. Vessels and airplanes may use a long time reaching and assessing the oil spill, sometimes also inaccurately [4], before response teams can act. A quick and correct assessment of the oil spill is necessary to identify the thickest parts of the oil spill. This is usually the most critical part, which a response team wants to control first

due to the risk of spread and drift. Remote sensing can be a more efficient way to gather information about oil spills compared to airborne sensors or in-situ observations. Remote sensing can cover large maritime areas with good spatial resolution, providing accurate information about the size and thickness of oil spills [4, 6]. Remote sensing can be used to regularly monitor large marine area for oil spill, but also as a response service for oil spill emergencies. However, large scale monitoring makes it difficult to get in-situ data for all detection, especially smaller oil spills. Assessment is then dependent on the expertise and experience of the oil analysis expert. In optical imagery, the use of the Bonn Agreement Oil Appearance Code (BAOAC) [7] may be used as a decent replacement for in-situ data, looking for oil in true color composite images [4].

Oil spills are affected by several factors at the ocean surface, including wind and sea currents [4]. Together these factors spread the oil spill on the ocean surface by time. Due to this, time used analysing a remotely sensed image for oil spills must be minimized. The analysis is mainly done manually by trained observers, visually inspecting the remotely sensed images [4]. ML classifiers can aid the observers to quickly identify potential oil spills, and aid in setting a confidence for detected features believed to be oil spills [8, 9]. However, due to the limited time, a trained observer can't wait too long for a ML classifier, so the ML classifier needs to be simple regarding computational time, yet effective enough to be of aid. For instance, Sentinel-2 images are often stitched together creating a mosaic of four or more tiles to cover a larger area compared to only using one tile, yet keeping the fine 10 m resolution. This creates a large demand of computational power for complex and time consuming ML classifiers.

There are several types of sensors used for oil spill remote sensing. Synthetic Aperture Radar (SAR) is one of the most popular sensors due to its ability to operate under a large variety of weather conditions [6, 10, 11]. It is operational in nighttime as well as daytime, and can penetrate clouds. The main disadvantages are the limited spectral information collected due to that SAR sensors rely on radar backscatter, and that SAR sensors generally are costly due to the advanced radar technology and the complex data processing involved [6]. Another type of sensors are optical sensors. They are multi-spectral or hyper-spectral scanners, meaning that they capture data across a range of spectral bands, with the potential to detect and classify oil spills based on their spectral signature. Optical sensors can be cheaper than SAR sensors [6, 12], and more free access data is available such as data from the Sentinel satellites and the Landsat satellites. The main disadvantages of optical sensors are the dependency of daylight, and the possible interference of sunglint when sun angle is similar to the view angle [13]. Other sensors, like passive microwave sensors, where not considered for this work. The detection of oil in remotely sensed imagery is often done manually by trained operators, both for SAR and

optical sensors [4]. This is often time consuming and resource-intensive. To save time and resources, a goal would then be to automate this service, or if this is not possible, create a tool that can aid the operators in their analysis. This would require the use of ML methods that could identify oil with high accuracy. Especially thick oil is of interest since this is the most critical part of an oil spill and have the best potential to be cleaned up [4].

Many ML methods have been used for oil spill classification. This include Neural networks (NN) [9], using a binary approach with two classes, K-nearest Neighbor (KNN) [9, 14] , Support Vector Machine (SVM) [14], Random Forest (RF) [14] and Decision Trees [14] using a multi-class approach with five classes. These methods are based on various fundamental principles and have different advantageous properties. Despite that both SAR and optical sensor are popular sensors used for oil spill detection[10], oil spill classification using ML methods are mostly applied on SAR data [11]. A thorough comparison study on various methods for multi-class classification, including thick oil might reveal the accuracy and robustness of the machine learning models on multi-spectral data. Therefore, the goal of this work is to compare and evaluate several machine learning approaches to accurately classify and separate thick oil from thin oil, oceanic and turbid waters in multi-spectral imagery. An additional class for vessels was also included due to the large number of ships present in the area of interest (AOI). The studied methods include NNs, KNNs, Decition trees, Naive bayes classifiers and discriminant classifiers. The Persian Gulf was chosen for the AOI due to frequent occurrence of oil spills. A secondary goal of this work is to assess if the evaluated ML models could be of aid in an oil spill monitoring services, such as KSAT's oil spill detection service. To achieve the main objective of this work we first created the dataset, second we trained a variety of different ML models, third we tuned the parameters of the ML models showing best result. Fourth we trained and applied four ML models which were found to have the highest accuracy in classifying thick oil. Finally we evaluated the results and applied the ML models to labeled and unlabeled data.

1.1 Outline of the thesis

This thesis is structured in 8 chapters.

Chapter 2 outlines the fundamentals of optical remote sensing applied for oil spill detection, presenting a theoretical background needed to understand the main objective of the thesis.

Chapter 3 presents the experimental setup of this thesis, outlining the steps.

Chapter 4 describes the data included in this thesis, how it was selected, how it was acquired, and the specifications.

Chapter 5 presents the methodology, describing how the dataset was created, how the training, testing and classification was performed and the machine learning setup for this thesis.

Chapter 6 presents the results of comparison and evaluation of the ML and their application to unlabeled data.

Chapter 7 discusses the main results.

Chapter 8 concludes the thesis and suggest further work.

2 Theory

In this project optical remote sensing is used for oil spill detection in ocean waters. Therefore, this chapter gives an overview of the relevant theoretical background on electromagnetic radiation and it's behaviour interacting with the atmosphere and ocean, on passive remote sensing of oceans, on passive remote sensing of marine oil slicks, and on pattern recognition and the machine learning (ML) classifiers used in this work.

2.1 Background physics

Optical remote sensing of oil spills is based on the physical properties of electromagnetic radiation. Therefor, this section provides an overview on electromagnetic radiation and its interactions with atmosphere and ocean.

2.1.1 Electromagnetic radiation

Electromagnetic (EM) radiation are waves of electromagnetic energy. The sun emits a complete spectrum of electromagnetic radiation. Microwaves, visible light, X-rays and gamma rays are all examples of different types of electromagnetic radiation. The electromagnetic spectrum, shown in Fig.2.1 describes the whole range of electromagnetic radiation, where the electromagnetic radiation is divided into different classes based on their wavelength and frequency

[6]. Optical remote sensing is passive remote sensing using the wavelengths from the ultraviolet to the infrared spectrum, which corresponds to the range between 10 μm - 1 mm [15].

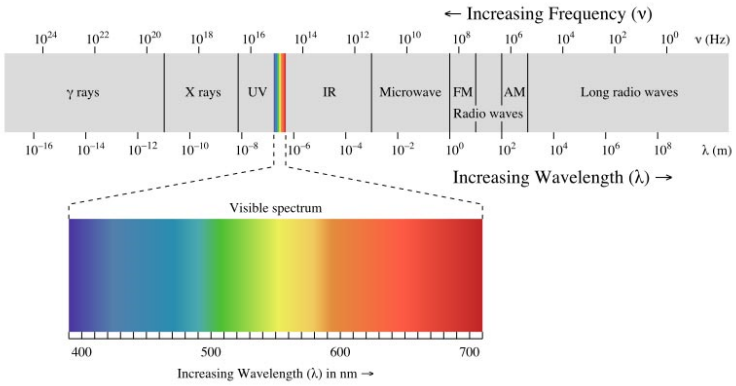


Figure 2.1: The electromagnetic spectrum, highlighting the visible spectrum [16].

Electromagnetic energy propagates in the form of a sine wave which is a geometric form that oscillates around an axis. A sine wave can be described by its amplitude, wavelength and frequency. *Wavelength* is the distance between to identical points in the wave, for instance the distance between crests. *Amplitude* represents the magnitude of the wave. Often the amplitude measures the energy level known as spectral irradiance. *Frequency* is a measure for counting number of wavelengths passing a fixed point in unit time. Hertz is often used as the unit of measurement, which counts passing's per second. For instance, 1 Hz means that one wavelength passes a fixed point in 1 second. Wavelength and frequency are inversely affecting each other. When the wavelength decreases, the frequency increases [6]. The parameters of the sine wave can be measured by a satellites sensor. An optical sensor only receives EM waves from within the optical range, ignoring EM waves from outside the optical range. The parameters are then used to estimate the surface reflectance values for optical imagery.

EM radiation consists of two sine waves. First is an electric sine wave which makes an electric field as it propagates. Second is a magnetic sine wave, which makes a magnetic field. Electromagnetic radiation is made up of an electric field and a magnetic field which both vary in magnitude and are oriented at fixed angles. The magnitudes vary along an axis perpendicular to the axis of propagation [6]. Figure 2.2 illustrates electromagnetic radiation propagating in x-direction. The electric field is denoted by E and is measured in volts per meter (V/m), which is equivalent to newtons per coulomb (N/C). The magnetic field is denoted by the vector fields B and H. In vacuum, their relationship is connected by the vacuum permeability, showed in Eq.2.1,

$$B = H * \mu_0, \tag{2.1}$$

where B is the magnetic flux density, and is measured in Newton per meter per ampere (Tesla units). H is the magnetic field strength, which is measured in ampere per meter, and μ_0 is the vacuum permeability constant. Further, we will describe how the electric and magnetic fields can be expressed by Maxwell's equations.

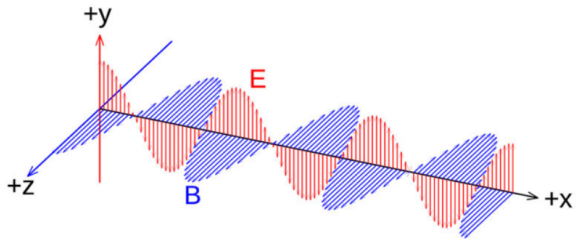


Figure 2.2: Electromagnetic wave propagating in a 3D space in x direction, showing the Electric field E with red colors, and the magnetic field B with blue colors [17].

Maxwell's equations describes how electromagnetic waves behaves in vacuum. Assuming no charges (p = 0) or currents (J = 0), the equations can be written by

$$\nabla \cdot E = 0 \tag{2.2}$$

$$\nabla \cdot B = 0 \tag{2.3}$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \tag{2.4}$$

$$\nabla \times B = \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \tag{2.5}$$

where E is the electric vector, B is the induction vector, μ_0 is the permeability constant in vacuum, μ_r is the relative permeability, ε_0 is the permittivity constant in vacuum and ε_r is the relative permittivity [18, 19]. Given an isotropic, homogeneous and nonmagnetic media, the wave equation can be derived combining Maxwell's equations. This can be expressed by

$$\nabla^2 E - \mu_0 \varepsilon_0 \mu_r \varepsilon_r \frac{\partial^2 E}{\partial t^2} = 0. \tag{2.6}$$

Solving the differential equation in Eq. (2.6) results in the expression

$$E = Ae^{i(kr-\omega t+\phi)}, \tag{2.7}$$

where A is the wave amplitude, k is the wave vector in the medium, ω is the angular frequency and ϕ is the phase [18].

Further on, we will describe the term reflectance, which is a measure of reflected energy from the earth's surface. This can be used to identify and separate what we observe at the ocean surface, such as oil and algae. To be able to explain reflectance, first we will describe relevant quantities measuring EM radiation.

Radiometry

Here we define quantities of EM radiation that will be used in this work. Radiant energy is the measure of the energy in electromagnetic radiation, with symbol Q_e and unit Joule (J). Radiant flux is radiant energy per second, with symbol ϕ_e and unit Watt, which is joule per second (J/s). Radiance is the radiant flux per unit solid angle per unit area, with symbol L_e and unit watt per steridian per square meter ($Wsr^{-1}m^{-2}$). Irradiance is the radiant flux received per unit area, with symbol E_e and unit watt per square meter W/m^2. Emittance is the radiant flux emitted per unit area, with symbol J_e and unit watt per square meter W/m^2.

Reflectance

Remote sensed reflectance (R_{rs}) is a measure on how much of the radiance that hits a surface are reflected to a sensor. This can be written by

$$R_{rs} = L_e^r/L_e^i, \tag{2.8}$$

where R_{rs} is the remote sensing reflectance, L_e^r is the reflected radiance at the surface, and L_e^i is the incident radiance to the surface. Due to that reflectance is a measure of reflected radiance divided by the incident radiance, the reflectance has no units and can be expressed in percentage or decimals. Using a sensor with one band results in a single reflectance value for each data point. Using a sensor with multiple bands gives us knowledge about the spectral behaviour, how much reflected radiation there is in different parts of the electromagnetic spectrum. When showing the reflectance as values dependent on wavelength, we get a reflectance spectrum. Reflectance spectrums are unique, and can

therefor act as a signature for the material[20]. This signature can be effected by several factors, for instance weather conditions and sun angle. For the case of oil slicks, thickness and type can have an impact on the spectral response[4, 21]. Fig.2.3 shows an example of a reflectance spectrum and illustrates the case where thickness of the oil slick has an impact on the reflectance. It can be seen that the reflectance decreases for oil slicks and increases for very thin oil in comparison to seawater in the range between 400-900 nm. Figure 2.4 shows another example of a reflectance spectrum [20]. This figure shows a stable reflectance for black oil beneath the more varying reflactance for seawater. Here we see that the reflectance for oil increases in the range between 400-600 nm, and descreases in the range between 600-1000 nm in comparison to seawater.

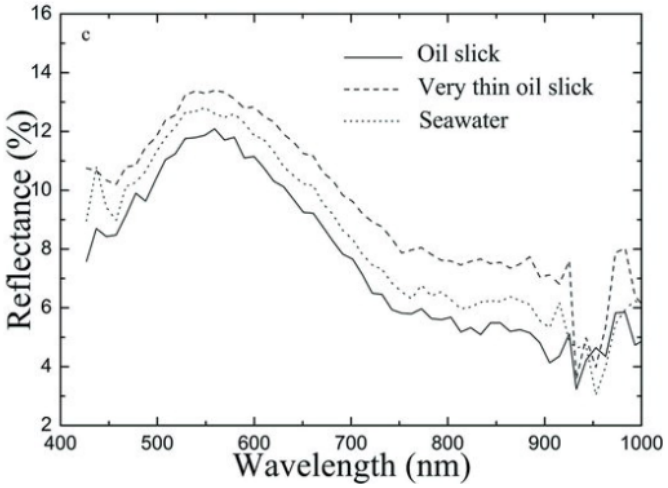


Figure 2.3: Reflectance spectrum of seawater, oil slick and very thin oil slick [20].

2.1.2 Atmospheric interactions

The atmosphere is a medium consisting of gases and aerosols which affects the electromagnetic radiation that propagates through it. These affects can be described by four interaction mechanisms; scattering, absorption, emission and refraction [18]. When scanning the earth's surface, we are only interested in the reflectance values from the surface, but these mechanisms causes additional radiation to the sensor. To be able to only observe the desired reflected radiation from the ocean surface, we have to do a correction to filter out all other radiation received. This is called the atmospheric correction (AC). The AC algorithm calculates the contributions from the unwanted radiation.

Radiative transfer theory can be used to model the contributions from the atmo-

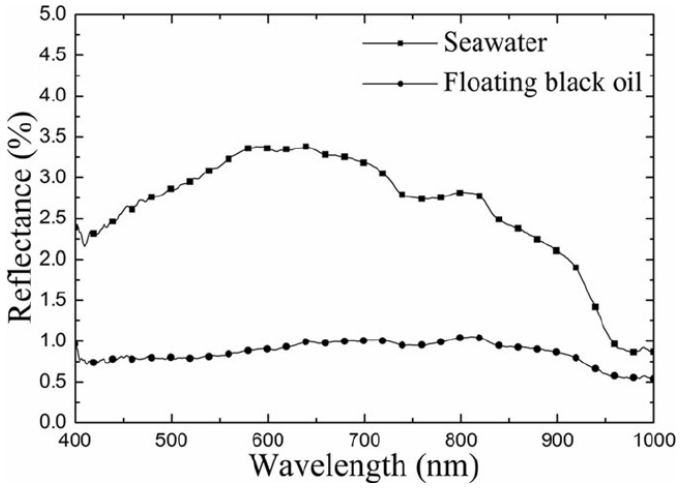


Figure 2.4: Reflactance spectrum for seawater (upper curve) and floating black oil [20].

sphere and to establish atmospheric correction algorithms [18]. The Radiative Transfer Equation (RTE), Eq. (2.9) is radiative transfer theory that describes how EM-radiation propagates in an absorbing and scattering medium and can be used to calculate gain and loss for radiation propagating through the atmosphere [18]. The equation can be written by

$$\frac{dI}{dz} = -\alpha_a I - \alpha_s I + \alpha_a B + \alpha_s J, \tag{2.9}$$

where the intensity loss in the wave due to absorption is given by

$$\frac{dI}{dz} = -\alpha_a I \tag{2.10}$$

and the intensity loss due to scattering is given by

$$\frac{dI}{dz} = -\alpha_s I. \tag{2.11}$$

The intensity gained from thermal emission in the medium is given by

$$\frac{dI}{dz} = \alpha_a B, \tag{2.12}$$

and intensity gained from scattering of waves incident from other directions is given by

$$\frac{dI}{dz} = \alpha_s J. \tag{2.13}$$

The term I is the original intensity before entering the medium. dI is the change in intensity at a specified point in the medium and dz is the distance the wave travels. The term α_a is the absorption coefficient summarized for all gas and particles in the medium. The term α_s is the sum of the scattering coefficient of all gases and particles in the medium. B is the the thermal energy from the medium, and J is the intensity from waves incident from other directions [18].

2.1.3 Ocean interactions

This section will describe the physics concerning reflection and scattering of electromagnetic waves on the ocean surface. The reflected radiation carries the information from the surface that is desired to retrieve by the spaceborne sensor.

Surface scattering

Radiation hitting the ocean surface can be reflected, scattered, absorbed or transmitted to different degrees and are depending on the surface properties such as roughness and dielectric constant. On a flat surface, the reflections are specular, similar to a reflection on a mirror. This is shown in Fig. 2.5 (a). The incident angle and the reflection angle are then identical. This is also called coherent scattering, meaning that the reflected radiation has a specific phase relation with the incident beam [22]. A non-flat surface is more complicated and is refereed to as a rough surface. A method to describe rough surfaces are the Reyleigh Roughness Criterion (RRC), which can be expressed by

$$(\sigma_\eta \, cos\, \theta)/\lambda < 1/8, \tag{2.14}$$

and it determines if the water surface is rough or smooth. Here σ_η is the Root Mean Square (RMS) surface height, θ is the angle and λ is the wavelength. The reflection is specular if Eq.2.14 is satisfied, and rough otherwise [23].

If the RRC cannot be satisfied at any angle, the reflection is Lambertian. This

is shown in Fig.2.5 (d) and is the case of an ideal rough surface with equal distribution in all directions. Fig. 2.5(b) shows reflection happening partly by specular scattering and partly by diffuse scattering. Diffuse scattering is also called incoherent scattering, meaning it has a random phase relation with the incident beam. As the ocean surface gets more rough, the specular scatter decreases and the diffuse scatter increases. Fig. 2.5(c) shows the last case not mentioned, where most of the scatter is diffuse in all directions, but still some specular scatter exists [23].

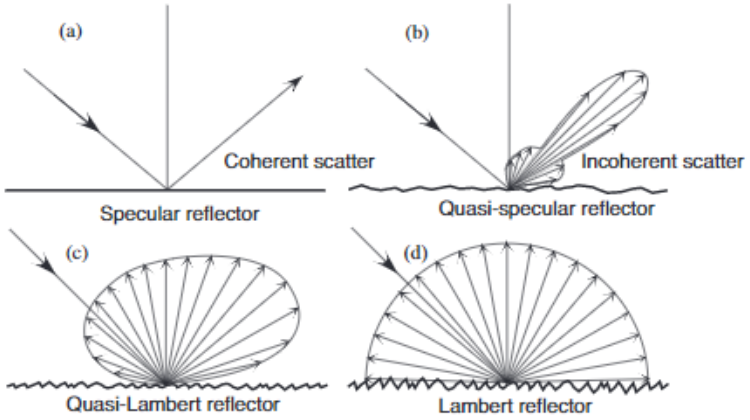


Figure 2.5: Four different cases of scattering: a) Specular reflection, b) Quasi-specular reflection, c) Quasi-Lambert reflection and d) Lambert reflection [23].

Transmission and interior scattering

Parts of the solar radiation that reaches the ocean surface is transmitted through the water body to different depths. The radiation is then either absorbed by the water body or scattered. In the specular case, two different equations describes the transmission and reflection of radiation. The first is Snell's law which can be written by

$$n_w/n_a \equiv n = sin\ \theta_i/sin\ \theta_t. \tag{2.15}$$

It describes which angles the radiance is reflected and refracted. The second is Fresnel equations, were the unpolarized case can be expressed by

$$r(\lambda, \theta_r) = L_r(\lambda, \theta_r)/L_i(\lambda, \theta_i). \tag{2.16}$$

Fresnel equations concerns the magnitude for the reflected and refracted radi-

ances depending on the incidence angle [23].

In Eq.2.15, n_w and n_a are the real part of the index of refraction for water and air and θ_i and θ_t are the angles of incident and transmitted radiation. Fig.2.6 illustrates the concept of reflection and transmission at a planar surface with relevant angles. In Eq.2.16 $r(\lambda, \theta_r)$ is the reflectance. The reflectance equals the ratio of the reflected radiance $L_r(\lambda, \theta_r)$ and the incident radiance $L_i(\lambda, \theta_i)$. λ is the wavelength, θ_i is the incident angle and θ_r is the reflected angle.

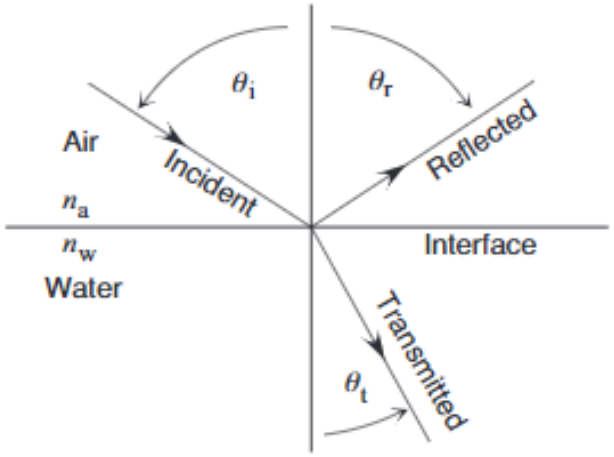


Figure 2.6: Concept of incident, reflected and transmitted radiation with angles on a planar surface [23].

A method to describe scattering in the ocean interior is the plane irradiance based reflectance $R(\lambda, z)$. This reflectance is used to relate the incident solar irradiance to the water-leaving emittance, the absorption properties and the seawater scattering. $R(\lambda, z)$ is defines as the the ratio of the upcoming plane irradiance (E_u) to the downgoing irradiance (E_d), and it can be written by

$$R(\lambda, z) = E_u(\lambda, z)/E_d(\lambda, z). \tag{2.17}$$

Here λ represents wavelength and z represents depth into the water. Fig.2.7 illustrates the concept of plane irradiance based reflectance [23].

This section gave an overview of background physics related to this work. This is important due to that remote sensing is based on the principles mentioned above.

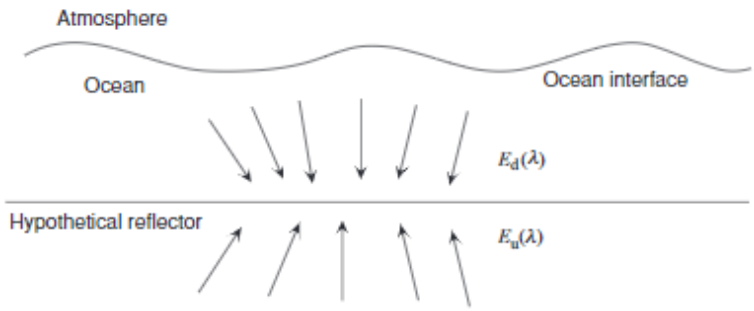


Figure 2.7: Concept of plane irradiance based reflectance [23].

2.2 Passive optical remote sensing of oceans

2.2.1 Radiation - From the sun to data products

Remote sensors receives and processes radiation. For optical remote sensing the source of illumination is commonly reflected radiation from the sun [19]. Figure 2.8 illustrates the journey of the radiation from the sun to the sensor.

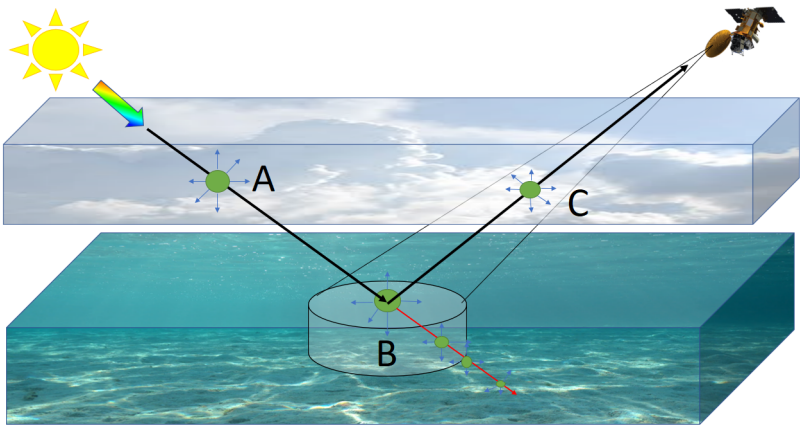


Figure 2.8: From sun to sensor. The black arrow shows the surface reflecting radiation. The red arrow shows the radiation transmitting into the water body, while the small blue arrows shows scattering in the atmosphere and in the water interior.

Solar radiation travels undisturbed through space before entering the Earth's atmosphere. Parts of the radiation is then scattered or absorbed by atmospheric gases and aerosols, as seen in Fig2.8, A. The remaining radiation reaches the earths ocean surface. Here parts of the radiation is scattered or absorbed, while other parts are transmitted through the water body and then scattered or absorbed at different depths. The rest are reflected at the surface. The interaction at the surface and ocean interior is shown in Fig2.8, B. The reflected radiation

travels back towards the sensor. Again parts of the radiation is scattered and absorbed in the atmosphere while the rest reaches the sensor, as shown in Fig2.8, C. The radiance are recorded and digitized by the sensor before being pre-processed and prepared to be sent back to earth for processing valuable data products.

In addition to the desired surface reflectance, a sensor also receives radiation from other source. Figure 2.9 illustrates all contributions to the sensor.

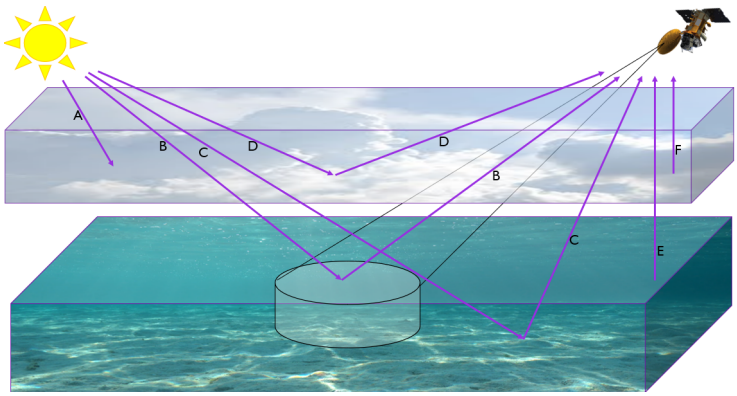


Figure 2.9: Radiation to satellite. A) Absorption in atmosphere. B) Ocean surface reflectance C) Scattering in the ocean interior. D) Scattering in atmosphere. E) Emittance from Earth. F) Emittance from the atmosphere.

The additional radiation that reaches the sensor is; scattered and emitted radiation from the atmosphere (Fig.2.9, D and F), scattered radiation from the ocean interior (Fig.2.9, C) and emitted radiation from the Earth's surface (Fig.2.9, E).

2.2.2 Passive optical sensing

A passive sensor senses radiation, mainly reflected radiation from the sun. This thesis works with passive optical remote sensors. Passive optical sensors are multispectral or hyperspectral scanners, meaning they obtain data from several bands of the electromagnetic spectrum. Combining different bands they are able to extract a large variety of spectral content from the data [6].

An advantage of multispectral data, is the ability to create colorful imagery, using bands in different combinations, such as a true color composites, different false color composites, and Normalized Difference Water Index (NDWI). This way the spectral content can be fully exploited, enhancing features in the image such as water, vegetation and urban areas for visual interpretation. The

spectral information combined with high spatial resolution makes it possible to interpret optical imagery for the human eye, possible to identify what we observe and to separate objects from each other. Other common uses of passive sensors can be to measure surface temperatures using infrared bands, measuring the concentration of chlorophyll, measuring colored dissolved organic matter (CDOM), or mineral particles close to or on the surface [19].

The alternative to a passive sensor is an active sensor. Common active remote sensors are Synthetic aperture radars (SAR). Detection and classification of oil is more often done in SAR imagery in practice [24]. An active sensor transmits it's own signal and senses the reflection of the same signal. For instance, an active microwave sensor sends out microwaves which is reflected at the surface before getting back to the sensor. Active sensors have fewer constraints compared to a passive sensor, and can be used under a wide range of operational conditions due to that they do not depend on radiation from the sun and is operational at nighttime as well as daytime. Also, active microwave sensors have the ability to penetrate clouds. This is what often makes SAR imagery a more attractive choice for companies such as KSAT, in maritime surveillance for services such as oil detection and vessel detection [24]. The downside of active sensors is that they come at a higher price and more investment is needed [6].

2.2.3 Multispectral imaging

Multispectral satellite imaging is a technology that uses sensors onboard satellites, airplanes, drones, e.g, to capture images of the Earth's surface in multiple wavelengths of light. The sensor can detect radiation in various spectral bands, including the visible, near-infrared, and thermal infrared regions of the electromagnetic spectrum. Each spectral band of the electromagnetic spectrum provides different information about the Earth's surface. For instance, the visible bands can capture information about the reflectance of different land cover types like vegetation, water bodies, and urban areas. The near-infrared bands can capture information about vegetation health and biomass. The thermal infrared band can capture information about the temperature of the Earth's surface [18, 23].

Multispectral satellite imaging can be used for a range of applications, including agriculture, forestry, environmental monitoring, and natural resource management. It can provide valuable information about crop health, vegetation cover, water quality, and land changes. One of the benefits of multispectral satellite imaging is its ability to capture images over large areas combined with a high spatial resolution, making it useful for monitoring changes over time and identifying trends and patterns. The technology has become more advanced over the years, with new sensors and algorithms enabling more accurate and detailed

imagery [25, 6].

2.2.4 Affects on detection capabilities for oil in optical imagery

There are several factors that can affect the detection capabilities of oil slicks in the optical imagery. In this section we describe these factors.

Sunglint

Sunglint is a phenomenon observed in optical satellite imagery where the angle of the sun equals the viewing angle of the satellite. It is a reflection causing the affected ocean area to appear silver-like. Often seen as a thick line through parts of the image, with the unaffected areas in contrast. Sunglint can cause a reduced detection capability when looking for surface constituents on the ocean surface in an optical image [13].

Clouds and haze

Oil spill detection by optical sensors can be affected by clouds and haze. This is due to the fact that clouds or haze can cover oil slicks, either partially or completely. The sensor is then unable to see the reflectance from the surface underneath [26].

Aquatic vegetation

The ocean is full of aquatic vegetation, often refereed to as the blue forest [27]. Algae, kelp, and seaweeds are examples of marine vegetation. When marine vegetation in various forms lays on the ocean surface, like for instance Sargassum or algae, the reflection can be visible from optical remote sensors. The chlorophyll in plant materials reflects light in the green part of the VIS spectrum, and absorbs in the red and blue part of the VIS spectrum, which make it appear green colored in true color composites. Other plant materials may have different reflections. This reflection can sometime be mistaken for oil in optical imagery. Examples of this is shown in Fig. 2.15, described in Sec.2.3.3 [4].

Other disturbances in remote sensed optical imagery

Other factors that can affect the ocean surface appearance in an optical remotely sensed image are wind conditions, ice, internal waves, cloud shadows, and oceanic fronts[4]. Altogether there are many factors that contribute to the possible complexity for trained observers and machine learning algorithms to detect oil spills at the ocean surface using passive optical remote sensing data.

2.2.5 Optical properties

The optical properties of oil and water depends on a variety of factors. For oil, thickness, type and oil:water emulsion ratio are important. For ocean water, temperature and salinity are key factors. The optical properties affects the remote sensing reflectance ending up at the sensor. This section presents some optical properties of oil and water.

Properties of oil

Oil is a wide term including many different unctuous liquids. In this work our goal is to separate petroleum oil from the ocean surface. The oil composition will vary depending on several properties, bot physical and chemical. Physical properties include density, sulphur content, viscosity, and water content. Chemical properties are the composition of different elements, such as hydrocarbons, nitrogen, oxygen, and nickel [28]. Examples of oil types are crude oil, lubricant oil and fuel oil. Figure 2.10 shows absorption at different wavelength in the visible spectrum for the three oil types. It can be seen that crude oil has the highest absorption coefficient, while fuel oil has the lowest.

Properties of ocean water

The optical properties of clear water varies with temperature and salinity. Fig. 2.11a shows an example of the absorption coefficient at different wavelength and Fig.2.11b shows an example of the scattering coefficient at different wavelengths.

Figure 2.12 show an example of the absorption and scattering of oil-in-water emulsion and sea water.

The oil:water emulsion have a higher scattering coefficient than sea water in the VIS spectrum. When viewing the absorption, we see that one sort of

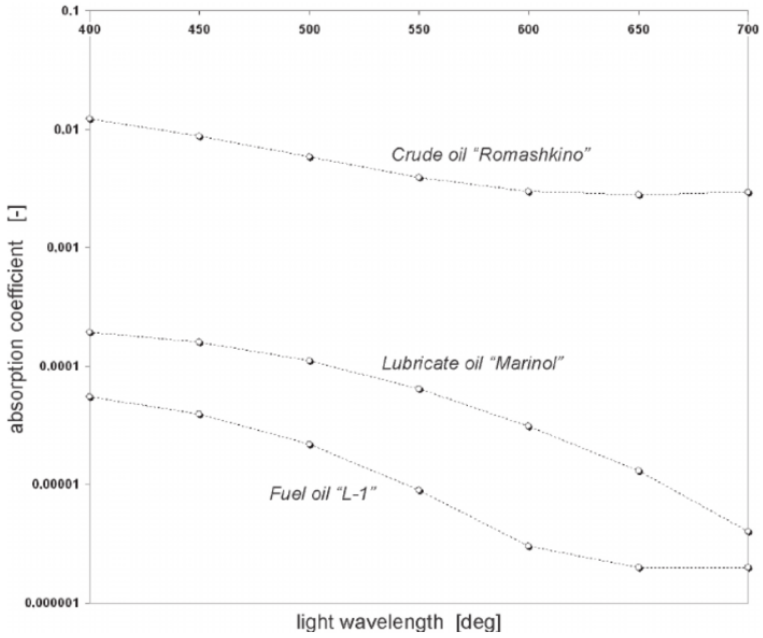


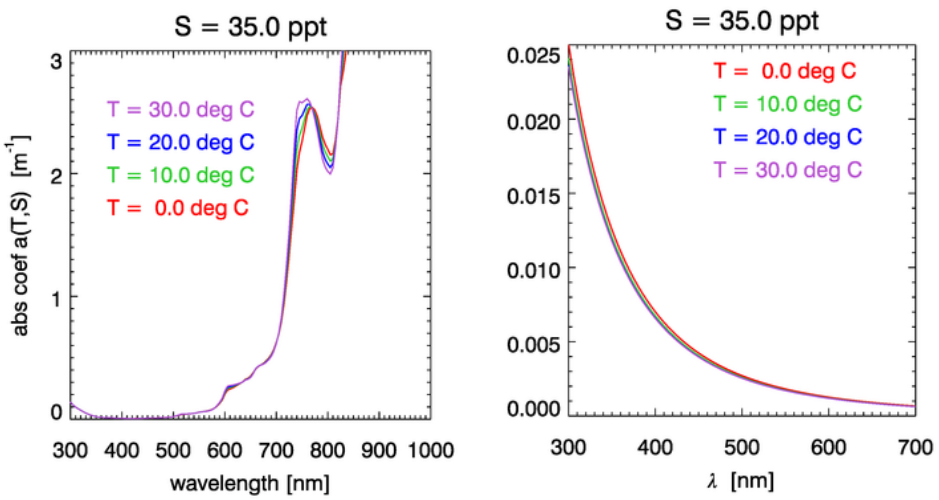
Figure 2.10: Absorption of three different oil types [21].

oil:water emulsion, Romashkino has higher absorption than sea water, while the other oil:water emulsion, Petrobaltic has a lower absorption coeficcient than sea water.

This section showed some examples for the optical properties of oil and water. It can be seen in Fig. 2.12 that the scattering of oil differs from water in the VIS. The same is valid for the absorption coefficient, but the absorption coefficient seem to be greatly impacted by oil:water emulsion type.

2.2.6 Sensors

This section describes some of the commonly used free sensors and satellites including Sentinel-2, which was used in this project. Table 2.1 shows the satellites along with their sensors, swath width, two different types of resolution, bands, spectrum interval and current lifespan. Swath width can be describes as the largest width the sensor possible can obtain when sensing an image. Spatial resolution describes the size of each pixel in the image. With higher spatial resolution more details and smaller objects can be sensed on the surface. Temporal resolution describes the revisiting time of a particular location on the Earth's surface. A revisiting time of 10 days means that it take 10 days for the satellite to come back to the same location moving in the same direction.



(a) Example of water absorption coefficient [19].

(b) Example of water scattering coefficient [19].

Figure 2.11: Examples of water absorption in the VIS and NIR spectrum and scattering in the VIS spectrum [19].

Another resolution worth mentioning is the spectral resolution. Spectral resolution describes the smallest spectral interval that can be properly sensed by the instrument [6]. Due to that each band has different spectral resolution, it is not included in the table. Ideally we want highest possible spatial and spectral resolution to obtain as much information from the surface as possible, combined with a high temporal resolution and a large swath width for more available images covering larger areas. Unfortunately this is difficult to obtain. For instance, a larger swath leads to a lower spatial resolution and vise versa. We have to decide which parameters are more important for the given task.

*MultiSpectral Instrument
**Operational Land Imager
***Moderate Resolution Imaging Spectroradiometer
****Visible Infrared Imaging Radiometer Suit
*****PRecursore IperSpettrale della Missione Applicativa

2.2.7 Sentinel-2

The Sentinel-2 constellation consist of Sentinel-2A and Sentinel-2B. They are a pair of European Space Agency (ESA) Earth observation satellites launched in 2015 and 2017, respectively. They are part of the Copernicus Programme,

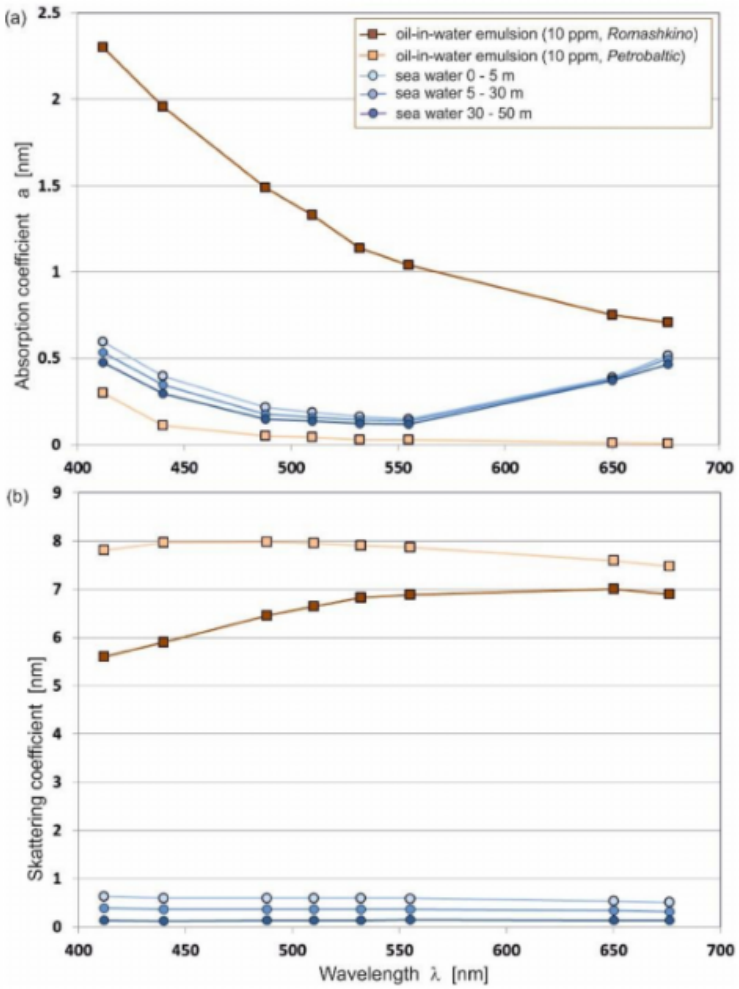


Figure 2.12: Example of absorption and scattering for two oil-in-water emulsion and sea water [29].

which is a joint initiative between the European Union and the European Space Agency to monitor the Earth's environment and security using space-based technologies. They are equipped with a multispectral imaging instrument called the MultiSpectral Instrument (MSI), which captures data in 13 spectral bands including wavelengths from the VIS (442 nm) to the SWIR spectrum (2202 nm). The MSI has a ground resolution of 10, 20, or 60 meters depending on the spectral band, which makes it ideal for a wide range of applications, including land cover mapping, agriculture, forestry, and coastal monitoring [30].

The Sentinel-2 satellites orbit the Earth at an altitude of 786 km and have a revisit time of 5 days at the equator, which means they can image the entire Earth's land surface every 5 days. This high revisit time, combined with

Table 2.1: Overview of sensors available for free optical remote sensing data in the Persian Gulf [30, 31, 32, 33, 34, 35, 36].

Satellite and sensor	Swath width	Spatial resolution	Bands and spectrum	Temporal resolution	Lifespan
Sentinel-2A/2B: MSI*	290 km	10 m, 20 m 60 m	13 bands 442-2202 nm	5 days	23.06.15/ 07.03.17-
Landsat-8: OLI**	185 km	15 m, 30 m	9 bands 430-2290 nm	16 days	11.02.13-
Landsat-9: OLI-2	185 m	15 m, 30 m	9 bands 430-2290 nm	16 days	27.09.21-
Aqua/Terra: MODIS***	2330 km	250 m, 500 m 1000 m	36 bands 405 nm - 14.385 μm	2 days	18.12.99(Terra)- 04.05.02(Aqua)-
NOAA-20 VIIRS****	3060 km	375 m, 750 m	22 bands 400 nm - 12.5 μm	10 days	18.11.17-
PRISMA***** PRISMA	30 km	5 m, 30 m	238 bands 400 nm - 2505 nm	29 days	22.03.19-

the high spatial resolution of the MSI, allows for monitoring of changes to the Earth's surface in near-real-time. The Sentinel-2 satellites are also equipped with a unique feature called the "revisit and tasking" capability, which allows users to request specific areas to be imaged with a high priority. This is particularly useful for emergency response and disaster management applications [30].

2.2.8 Processing

To extract useful information from multispectral data, the radiance measured by the sensor must be processed. The processing is done in several steps which results in different levels of processed data. The levels can be described as follows [30]:

- Level-0 (L0): Unprocessed instrument data.
- Level-1A (L1A): Unprocessed instrument data, but appended information such as geometric and radiometric calibration coefficients and georeferencing parameters.
- Level-1B (L1B): The beginning of science data. Applied the calibration coefficients from L1A and processed to sensor units.
- Level-1C (L1C): L1B data that include new variables to describe the spectra. For instance Top of Atmosphere (TOA) reflectance.

- Level-2A (L2A): Geophysical sensor data and surface reflectance (e.g. surface ground elevation or chlorophyll concentration).
- Level-2B (L2B): L2A data that have been processed to sensor units.
- Level-3 (L3): Variables mapped onto uniform space-time grids (e.g. create large scale maps by combining images from multiple orbits, like mapping the entire earth.)

In this work we used surface reflectance level-2A data. Using L2A data means that any uncertainties in the pre-processing chain might propagate to the product.

2.3 Passive remote sensing of marine oil slicks

2.3.1 Oil slick science

Oil slicks come from natural sources (seeps) as well as man made sources, such as tanker vessel spills, pipeline leaks and platform spills [4]. The type of oil and composition affect how the oil changes over time, as well as marine factors such as weather, current, waves and water [37]. Figure 2.13 shows the different effects oil can be affected by in the ocean. These effects are; "*wind and wave advection, compression from waves and currents (into wind rows and narrow slicks), spreading and surface diffusion, sedimentation and dissolution into the water column, emulsification, evaporation and photo-chemical and biological degradation*" [4].

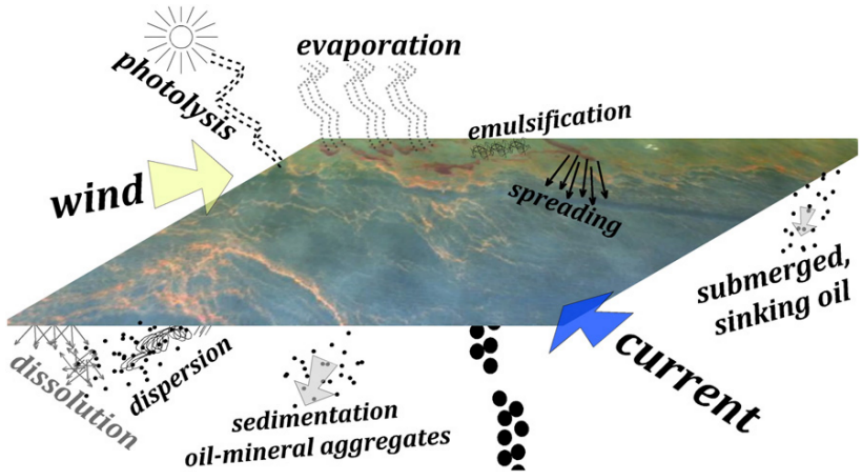


Figure 2.13: Marine effects on oil [4].

2.3.2 Environmental monitoring - machine learning

Passive optical remote sensing is used to routinely monitor ocean areas for oil spills. Also, passive optical remote sensing can be an important tool for a quick response in an emergency oil spill event [4]. Trained observers detect marine oil slicks by observation. Machine learning can be a great asset to help detecting oil spills, and also to prevent the detection of false positives. False positive detection can be costly for the environment and society in a time critical situation, allowing the real oil spill to spread more. Also, the resources set into action using oil spill response teams are economically costly for society [4].

In an event of an oil spill emergency we want to identify the thickest part of the spill so that the ground response teams can minimize the spread by engaging the thickest areas first. The thickest part is the most actionable part. [4]. Machine learning can be an asset classifying the oil. Using a trained machine learning model to do a preliminary or additional screening of the image, looking for oil slick, can possibly save time for trained observers and help classify detections correctly. There are many different machine learning (ML) methods available. Commonly used ML methods for optical remote sensing are [38]:

- Artificial Neural Network (ANN)
- Support Vector Machine (SVM)
- Fuzzy logic
- K-nearest neighbor
- Ensemble learning

The ML models used in this works are described in Sec. 2.4.2

2.3.3 Detecting oil slicks

Visual appearance

Identifying oil slicks in the visible part of the EM spectrum are complex due to the number of look-alikes and the number of factors affecting the visual appearance. Figure 2.14 shows aerial photos from the DWH spill in 2010, with examples of thin sheen slicks and thicker oil slicks. To help standardizing reporting of oil slicks, the slick thickness can be defined based on the visual

appearance on the ocean surface through an oil appearance code system; the Bonn Agreement [4]. See "Bonn Agreement oil appearance code under Sec. 2.3.3.

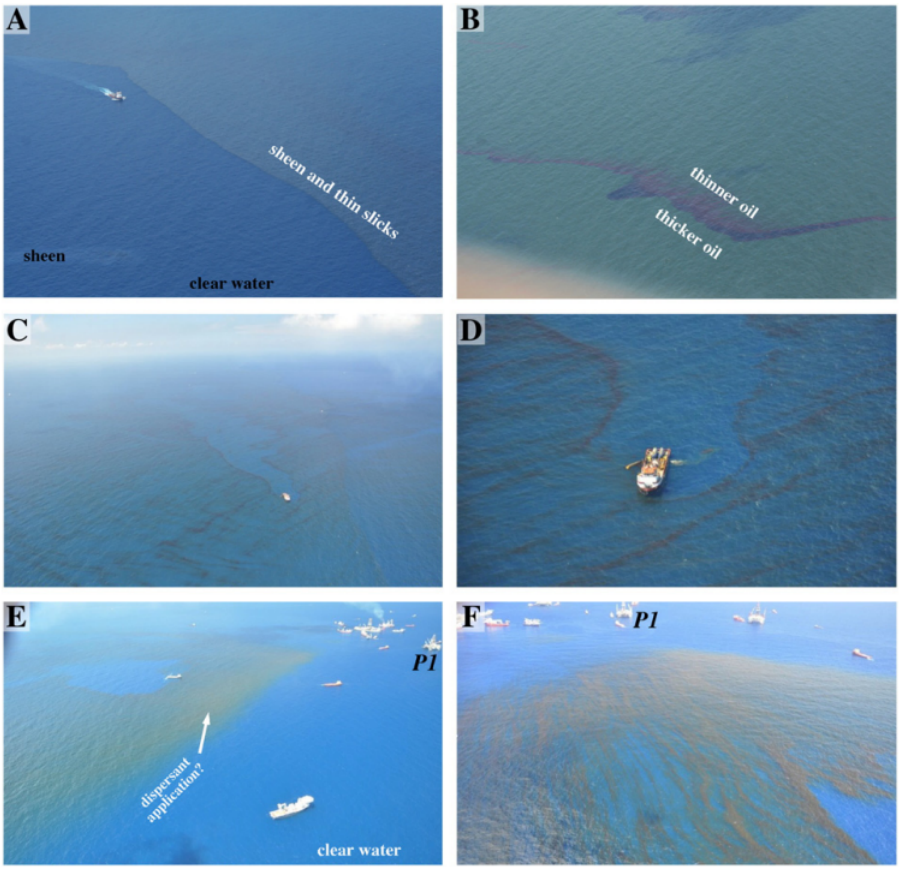


Figure 2.14: Examples of the visual appearance of thick and thin oil [4].

Identifying oil spills can sometimes be challenging due to look-alikes, especially identifying thin "sheen" oil slicks. Marine phenomena such as ice, kelp beds, plankton blooms, cloud shadows, weed beds, wind shadows, weather fronts and algae, appear similar to thin sheens in true image composites [4]. Studies [39] have shown that the sea surface's upper 1 mm microlayer is a gelatinous biofilm. Figure 2.15 shows examples of marine phenomena similar to oil spills. Other non-marine factors affecting the visual appearance are described in Sec.2.2.4.

Fig.2.15 (B) shows red oil that is similar to the algae and wake shown in Fig.2.15 (A). The orange oil in Fig.2.15 (D) is similar to the Sargassum shown in Fig.2.15 (C). Last the brown oil in Fig.2.15 (F) is similar to the brown algae in Fig.2.15 (E).

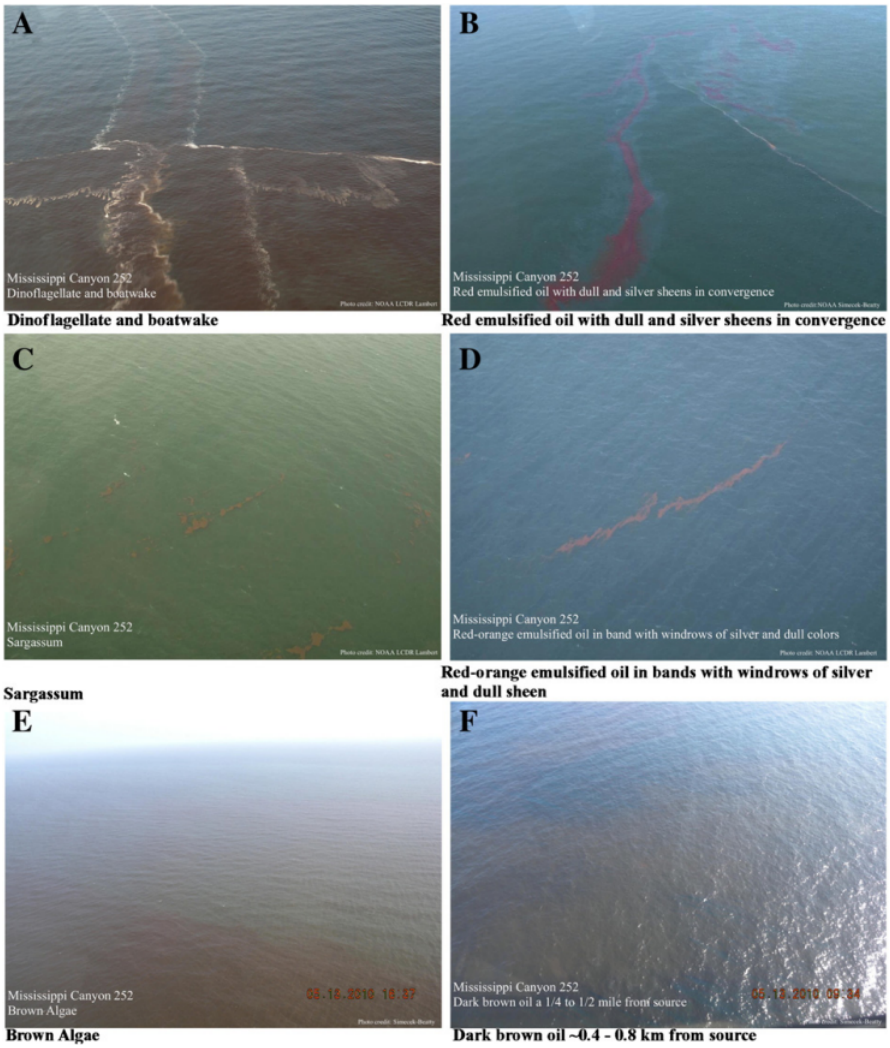


Figure 2.15: Examples of marine phenomena similar to oil spills next to oil spills. [4].

Bonn Agreement Oil Appearance Code

The Bonn Agreement Oil Appearance Code (BAOAC) is a method for detection and estimation of oil volumes at sea by visual interpretation of oil spills, developed by the Bonn agreement committee. The Bonn agreement is an agreement between ten governments and the European Union cooperate dealing with pollution in the north sea. The first agreement was signed in year 1969, and the first agreement code was created in 1993 [7]. Fig.2.16 is a caption of the most relevant part of the current oil appearance code, including a table showing the different thickness categories. The appearance for codes 1 to 3 are based on optical effects, while codes 4 and 5 are defined by the true color of the oil slick. Code 1, sheen, is for the thinnest slicks, up to 0.3 μm. They have

a silvery appearance due to a higher reflectance relative to seawater. Code 2 slicks, appear with rainbow colors, and are defined with thickness up to 5 μm. From 5 to 50 μm (code 3) we have oil color, but the reflection from the sky is dominant. This is described as metallic. Codes 4 and 5, with oil slicks thicker than 50 μm, appears with the oil's true color. Absorption is dominant in these slicks, so sky reflection and optical effects are small [7, 4]. Figure 2.17 shows an example of the sheen, rainbow and metallic oil appearance, and Fig. 2.18 shows an example of true colored oil appearance. In these examples, it can be seen that the reflectance decreases with increasing oil thickness.

11.2 The Bonn Agreement Oil Appearance Code

4. Since the colour of the oil itself as well as the optic effects are influenced by meteorological conditions, altitude, angle of observation and colour of the sea water, an appearance cannot be characterised purely in terms of apparent colour and therefore an 'appearance' code, using terms independent of specific colour names, has been developed.
5. The Bonn Agreement Oil Appearance Code has been developed as follows:
   - In accordance with scientific literature and previously published scientific papers,
   - Its theoretical basis is supported by small scale laboratory experiments,
   - It is supported by mesoscale outdoor experiments,
   - It is supported by controlled sea trials.
6. Due to slow changes in the continuum of light, overlaps in the different categories were found. However, for operational reasons, the code has been designed without these overlaps.
7. Using thickness intervals provides a biased estimation of oil volumes that can be used both for legal procedures and for response.
8. Again for operational reasons grey and silver have been combined into the generic term 'sheen'.
9. Five levels of oil appearances are distinguished in the code detailed in the following table:

Code	Description - Appearance	Layer Thickness Interval (µm)	Litres per km²
1	Sheen (silvery/grey)	0.04 to 0.30	40 – 300
2	Rainbow	0.30 to 5.0	300 – 5000
3	Metallic	5.0 to 50	5000 – 50,000
4	Discontinuous True Oil Colour	50 to 200	50,000 – 200,000
5	Continuous True Oil Colour	More than 200	More than 200,000

Figure 2.16: Bonn Agreement oil appearance code [7].

Visible, near infrared and short wave infrared spectrum

Studies [4] have shown that oil can have a variying spectral response in the NIR and SWIR spectrum, based on the emulsion thickness and the oil:water ratio for thick emulsions. This is shown in Fig.2.19. Observing Fig. 2.19A, showing the reflectance for different oil:water emulsions, we see three distinct apexes. The first with highest reflectance, is in the NIR range, about 0.8-1.3 μm. Here there is

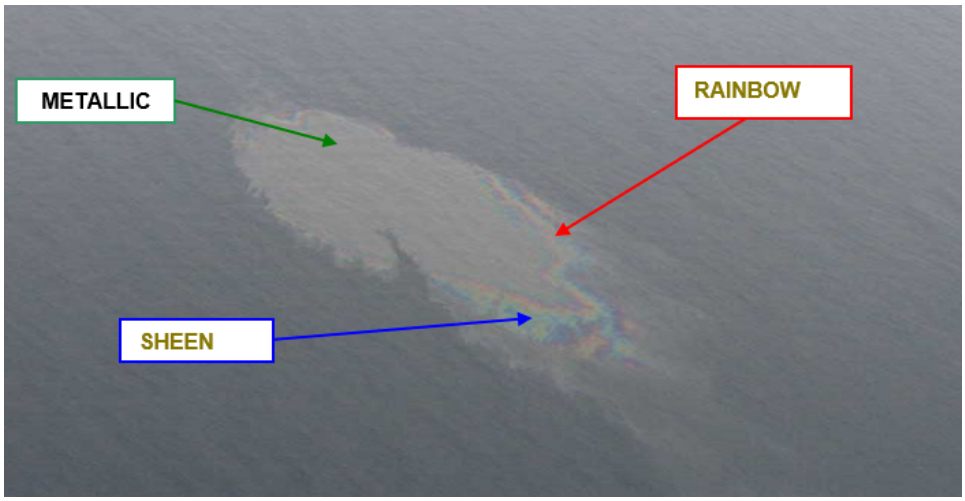


Figure 2.17: Example showing code 1 (Sheen), 2 (Rainbow) and 3 (Metallic) [7].

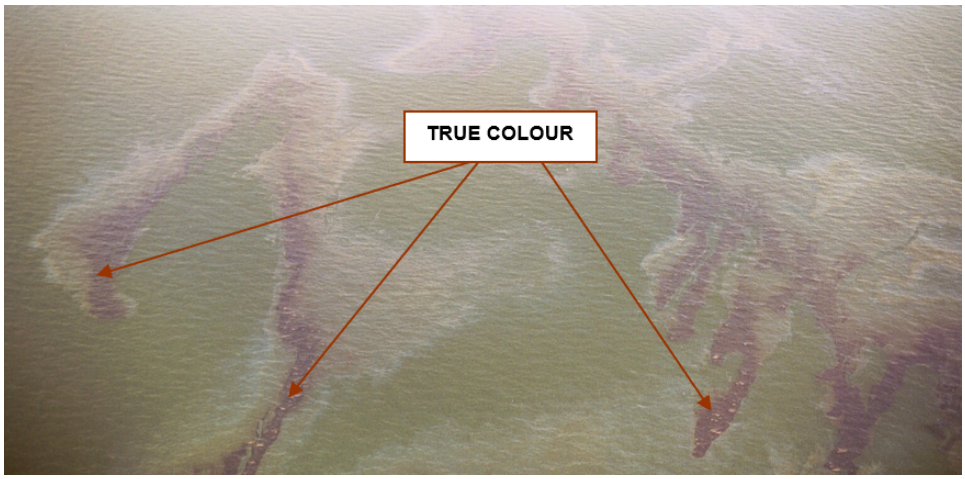


Figure 2.18: Example showing code 5, true colored oil [7].

a distinct difference in the reflectance for the different oil:water emulsions. The 40:60 oil:water ratio, has the highest reflectance, while the 92:8 oil:water ratio, has the lowest reflectance. In the second apex, in the SWIR range between 1.6-1.85 μm, the difference in reflectance between the 40:60 and 60:40 oil:water ratio is small. There is still a gap down to the 75:25 and 92:8 oil:water emulsions. In the last apex with the lowest reflection of the apexes, in SWIR range betweem 2.1-2.3 μm, the differences between the emulsions are similar to that of the second apex.

Observing Fig. 2.19 B, showing reflectance for 60:40 oil:water emulsion with different thicknesses, we see a similar curve of that in Fig. 2.19 A. Again, we

see three distinct apexes. One between 0.7-1.4 μm, a second between 1.55-1.85 μm, and a third between 2.1-2.3 μm. It appears to be little or no difference in reflection values for the 8 mm, 4 mm and 1.85 mm thickness in all wavelengths. In the range between 1.4-2.5 μm, the difference in reflection values is small for thicknesses between 0.5-8 mm. In the range between 1.9-2.5 μm the difference in reflection values is small for thicknesses between 0.1-8 mm.

From Fig. 2.19A and Fig. 2.19B we clearly see the impact the oil:water emulsion ratio and the oil:water emulsion thickness has on the reflectance from oil in optical remote sensing.

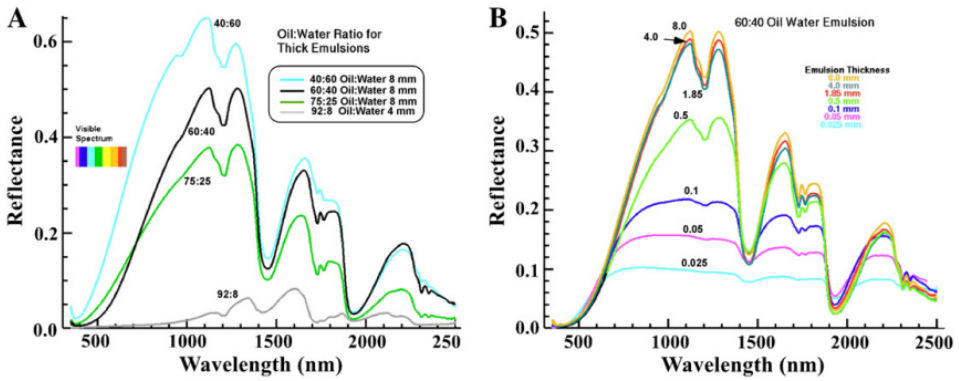


Figure 2.19: Figure A shows the reflectance for different oil:water emulsions. Figure B shows the reflectance for 60:40 oil:water emulsion with different thicknesses. Both showing reflectance values within the visible, NIR and SWIR range [4].

2.4 Pattern recognition

Machine learning (ML) is the scientific field of automated learning. ML methods use training data consisting of an input and output dataset. This training data is used to establish the ML algorithm for estimating the desired output. There are many ML algorithms in pattern recognition. Various methods are used for different applications. For instance it can be as simple as a spam filter searching for keywords in emails, or more advanced a facial recogniser looking for wanted humans in surveillance videos[40]. In this work, classification algorithms were used for detecting oil spills. This section will describe the methods used in this thesis and the statistical measures used to measure their performance.

2.4.1 Pattern recognition

Pattern recognition is a branch of machine learning, where the goal is to classify data into classes. In this field, algorithms are often referred to as classifiers. We separate between supervised and unsupervised classification. Supervised classification uses training data with known classes to train the classifier before it can be used. Unsupervised classification does not use any training data and classifies based on similarities in the data [41].

For supervised classification, the data is usually divided into training, validation and test data. The split percentage is user-specific, and the size of the training data often vary from 50 % to 90 % of the dataset. There is no optimal split percentage that covers all cases. The split percentage must be adjusted to meet the ML models needs to get the best performance possible [42]. The classification algorithm is then trained, tested and evaluated on this data. One can also choose to exclude the validation part, such that the data are divided into training and test data. For instance for a ML model with set parameters and there is no need to tune the model using validation data. The performance of the classifier algorithm can be evaluated by using confusion matrices and accuracy measures, such as the overall accuracy [41, 40]. It is also possible to apply a model on unlabeled data to visually inspect the model. Next, confusion matrices, the overall accuracy and prediction are described.

A confusion matrix is a tool for measuring the performance of a classifier. After a classifier has been trained, the performance is evaluated, either for validation or testing, predicting the correct class for the data. The classified data is then compared to the true classes in a specific table layout, the confusion matrix, which shows the percentage of the each class being classified correctly and the percentage of each class being classified falsely into other classes. The confusion matrix can also be presented using the number of observations instead of percentage. Figure 2.20 shows an example of a confusion matrix, presented in percentage. The diagonal going from top left to bottom right in Fig. 2.20 is where the predicted class have been classified correctly as the true class. This is the true positive rate (TPR) for each class, also referred to as sensitivity. All other information in the table shows misclassified observations. When the misclassified observations in a row are added together, it is called the false negative rate (FNR) for that class. Confusion matrices are used to measure how accurately a ML model classifies each class and to locate possible errors [41, 40].

The overall accuracy (OA) is another statistic for measuring the performance of a classifier. OA is defined by summing up all the TPR values for each class, and divide the sum on the number of classes. This gives an OA, a tool to measure the overall performance. It is a great tool to compare the model against other

True class					
	Class 1	80 %	1 %	2 %	17 %
	Class 2	1 %	95 %	1 %	3 %
	Class 3	5 %	4 %	76 %	15 %
	Class 4	2 %	2 %	2 %	93 %
		Class 1	Class 2	Class 3	Class 4
		Predicted class			

Figure 2.20: Example of a confusion matrix.

models, but it tells us little about the individual performance of each class. In cases where certain classes are more important than others, or when we have a large difference in observations for each class, TPR can be of greater value than the OA. For instance in this thesis, the number of observation for ocean water (class 3) and turbid water (class 4) is much greater than the number of observations for thick and thin oil, see Sec. 4.3.3. If the OA is high, it might be due to the fact that these classes have a high TPR. In this work, we are interested in classifying a specific class, namely thick oil with high accuracy. Therefore, both the confusion matrix and the OA are computed to evaluate the performance of the method.

Once the classifier is established, we want to apply the model on a dataset without labels. This is called prediction and are used in order to assess the performance of the model visually. In our case, the models are applied to Sentinel-2 data, using the same bands as the model was trained with. The model evaluates the data, and classifies it into classes based on the training, resulting in a classified image with one label color representing each class. Since we now are working with data without annotation, we have no statistical measures to assess if the classification is true or false. We can only compare visually the classified image against the equivalent true color composite. Masks may also be used to filter out areas that we want excluded from the classification. For instance, in the case of oil classification a landmask will be useful if the ML model is only trained over marine areas.

2.4.2 Algorithms

In this work, several algorithms were evaluated for oil slick classification. These methods include Decision trees, K-Nearest Neighbors, Neural Networks, Naive Bayes classifier and discriminant classifiers, a simple efficient linear SVM and a simple efficient logistic regression classifier. Based on the computed statistical measures, it was found that the best performers are KNN and ANN. Therefore, these methods are briefly described below. More details on the methodology can be found in Sec. 5.7.

K-nearest neighbour

The K-nearest neighbour (KNN) is a simple yet effective machine learning classifier. Before using a KNN classifier, training data should be normalized to avoid large differences between different features in a dataset. The KNN classifier uses classified neighborhood values to determine the class of new datapoints. The term K specifies how many neighbourhood data observations the classifier uses in the estimation. By measuring distance between the new observation and neighborhood observations, it finds the K-closest classified data observation, and classifies the new observation to the class that is dominant in the defined neighbourhood [43, 44].

Two factors affect the creation of a KNN classifier; the value of K and the measure of distance. The classification performance of KNN is greatly influenced by the determination of the K-value. If K is small, "*the local estimate tends to be very poor owing to the data sparseness and the noisy, ambiguous or mislabeled points*" [45]. A large K number on the other hand can include outliers from other classes. The distance measure can also affect the performance of the classifier, where different distance measures can include different neighborhood datapoints. There are many type of distance measures, for instance Euclidean, Euclidean squared, City-block and Chebychev [45]. The choice of the distance measure and K depend on the type of the data.

Artificial neural networks

Artificial neural networks (ANN) are classifiers inspired by the brain. The brain consist of 10^{11} neurons working in parallel as processor units. In the brain they are connected with synapses, each having connections to around 10000 other neurons. In artificial neural networks the "neuron" is called a perceptron or node, and is the basic processing element. It receives an input and produces an output. The perceptron takes input from the environment or from other perceptrons, where each input has an associated weight. Together they can

form a network. In the simplest case the output is a weighted sum of the inputs. Fig.2.21 shows an example of a simple perceptron [43] including input nodes (x), an output node (y) and weigths (w).

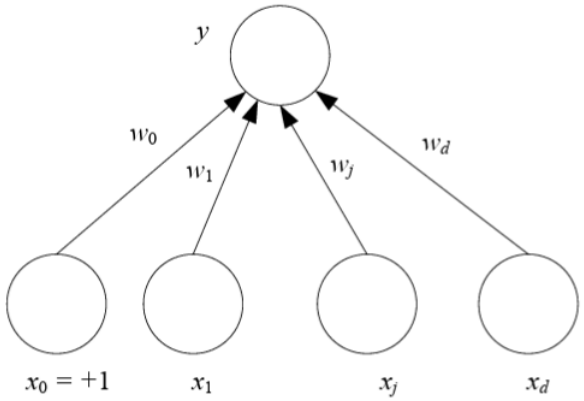


Figure 2.21: Simple Perceptron with input layer x, weights w and output layer y [43].

A node in an ANN has an activation function. In simple terms this function decides if a node should be activated or not. The activation function takes the summed weighted inputs inserted to the node, and creates an output value depending on the function, to insert into the next layer or output. Figure 2.22 shows this case. Here x are the inputs, w are the weights, z is the sum of the input and weights, f is the activation function, and a is the output. Examples of activation function are Sigmoid, Rectified linear unit (ReLU) and tangent hyperbolic (tanh) [46, 40].

There are many different type of ANN models, including Multi Layer Perceptron (MLP), Radial basis function network (RBFN), Bayesian regularized neural network (BRANN) and Kohonen's selforganizing map (SOM) [48]. They can be classified into Feed forward neural networks (FFNN) or Feed backward neural networks (FBNN), also called backpropagation. In FFNN, the information is transmitted in only one direction, from input node, via potential hidden nodes, to an output node. Examples of FFNN are single layer or multilayer perceptrons, and RBFN. In FBNN, the information can also be transmitted backwards to adjust the network's weights and biases. Here a cost function are used to measure the performance of the network, comparing the output with the training sample. The cost function is used to calculate the error of the output layer, which is used to train the network [41]. Examples of FBNN are BRANN and SOM.

When training an ANN classifier we need to build the network. This involves deciding how many nodes we want and how many layers of nodes we want. Also we need to decide which activation function to use. An ANN can greatly

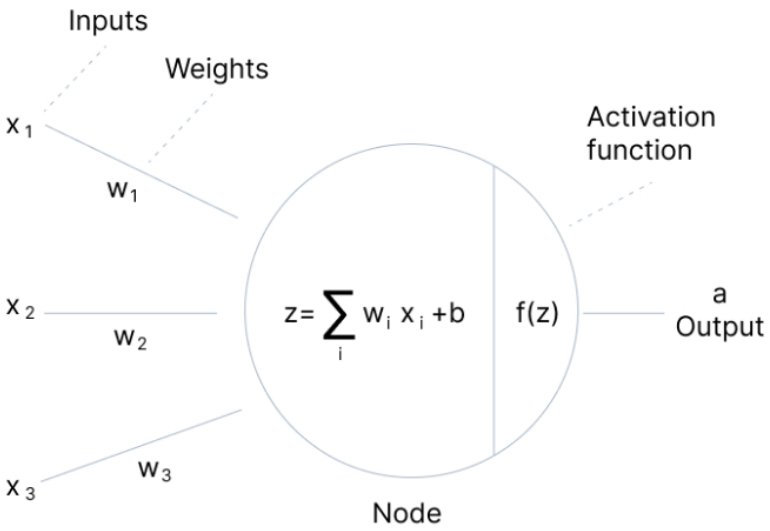


Figure 2.22: Node, the artificial neuron with input x, weights w, z which is sum of the inputs, weights and bias, the activation function f(z) and output a [47].

vary in complexity depending on the model choice and the configurations. More layers and nodes leads to more complexity and greater need of computational power.

3

Experimental setup

Chapter 3 gives an overview of the different steps involved working with this thesis; the experimental setup.

This work experimented with multiple supervised ML methods applied to multispectral satellite data with the main goal of detecting thick oil. Figure 3.1 shows a flowchart explaining the experimental setup.

As shown in Fig. 3.1, the whole process can be divided into two parts, data and methodology, and are described by; image selection, data collection, dataset creation (annotation, band selection, up-sampling), training, testing, evaluating, and prediction. Image selection describes how we chose the remote sensing data included in this works and data collection describes how we collected the data and what type of data we collected. Annotation describes how the labeled data was created. Band selection describes how we chose the bands included in this works, and up-sampling describes how we performed the up-sampling on two of the bands. Then the dataset section explains how the extracted remote sensing data and the labels were merged together, forming the dataset used for training the classifiers in this works. Training, testing and evaluation describes how the training and testing was performed and evaluated, and the different stages involved. Confusion matrix and the overall accuracy where computed. Last, in the prediction step, the ML models were applied to unlabeled data for visual interpretation. The programming implementations in this works were performed in Matlab software by Mathworks.

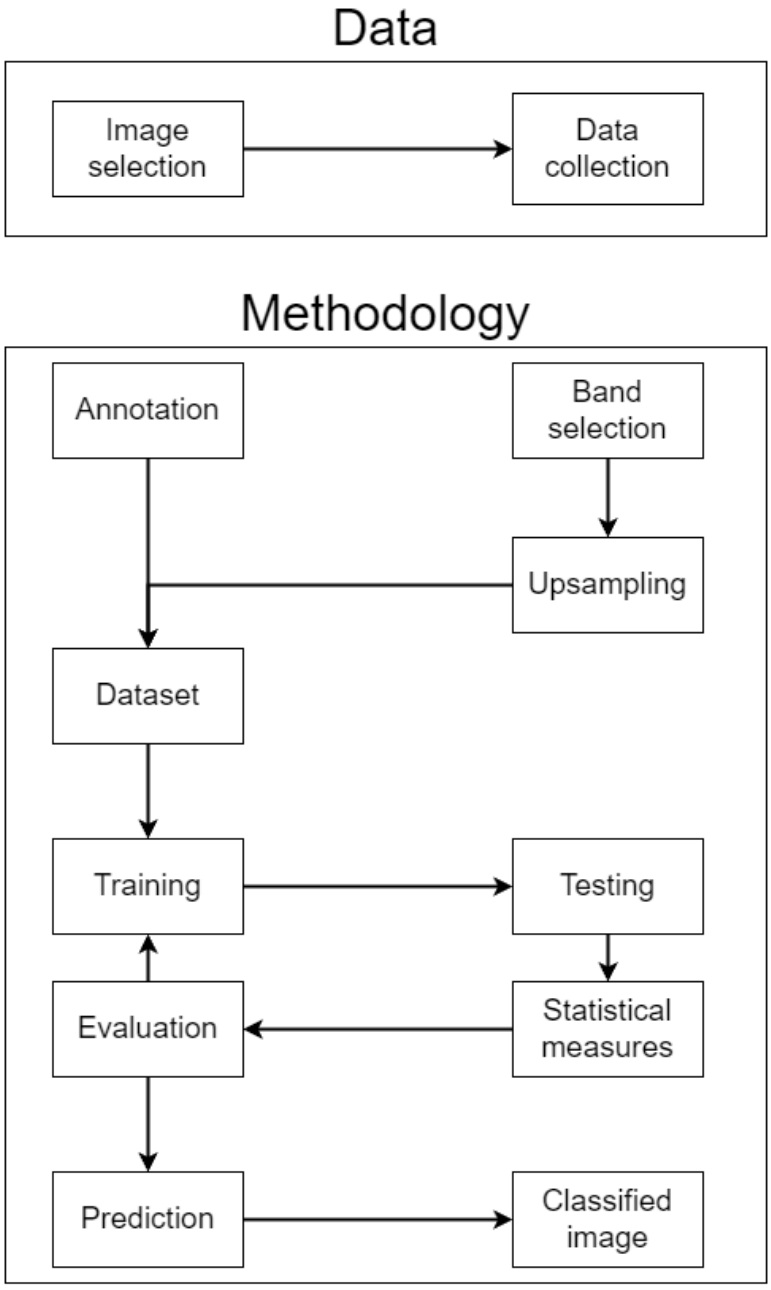


Figure 3.1: Flowchart divided into data and methodology showing the steps involved working with this thesis.

4

Data

The ideas and subject for this thesis were developed in collaboration with the company Kongsberg Satellite Services AS (KSAT). The initial thoughts were to use pattern recognition in optical satellite imagery to identify oil on the ocean surface which is highly relevant for KSAT. Together with two expert analysts from KSAT, we identified the Persian gulf as an interesting area due to knowledge of previously detected oil spills in this area with good oil detection capabilities, which could be used as training data. For the oil detection, high spatial resolution data was preferred. A finer spatial resolution gives more detailed information about an oil slick [4], and we are more likely to detect smaller oil slicks. Sentinel-2 was chosen due to the wide range of free data available and past experience with the data. To minimize the seasonal variation of the climate and weather in the dataset, we decided to mainly rely on imagery from June 2021 and June 2022. Due to the high availability of previously detected oil slicks in this area, we decided that the thesis was going to use supervised classifiers, which meant that a large portion of the work involved with this thesis was the annotation and creation of the dataset used for training data.

4.1 Image selection

Sentinel-2 images with detected oil slicks based on the Bonn agreement oil appearance code, see Sec.2.3.3, were selected in collaboration with expert analysts at KSAT. For this the open access Sentinel-HUB EO-browser [49] was used to inspect the images before collecting. A large number of images were inspected, before 12 images were included, where 10 images were used for training and testing, and two were used for prediction. These are described in Sec. 4.3.1. To avoid seasonal differences only images from June 21 and June 22 were chosen for the training dataset. All of these images contain oil slicks, and some contain look-alikes and complex areas with shallow water and calm wind. There clearly are differences in sunlight conditions and ocean color in the images. This could potentially affect the classifiers, leading to a higher variance in the dataset, but due to the nature of a monitoring service, where the sunlight conditions will change, it was necessary to get a broad specter of different sunlight conditions. It could potentially make the ML models more robust to handle different sunlight conditions.

4.2 Data collection

The images were downloaded from the European space agency's (ESA) Copernicus open access HUB [50]. This is an open access HUB, which means that anyone can download data from all the Sentinel satellites for free. Only a free-to-make user is required. When searching for the images, we had the option to download either Top of atmosphere (TOA) level-1 data or surface reflection level-2 data. The main focus of this work is about creating a dataset and training classifiers, so we decided to use the level-2 products. This meant that we did not have to perform atmospheric correction on the remotely sensed data when creating the dataset.

4.3 Data specification

4.3.1 Tiles

A Sentinel-2 product consists of granules that have a fixed size. The smallest indivisible partition of a product is known as a granule, which contains all spectral bands. Level-2A granules, also known as tiles, are ortho-images in UTM/WGS84 projection covering an area of 110x110 km^2. The Earth is divided into a predefined set of tiles using a 100 km step in UTM/WGS84 projection. However, each tile has an area of 110x110 km^2 to ensure overlap with the neigh-

boring tiles [30].

Data from the following Sentinel-2 tiles were used for training the classifiers:

- Tile 1: 2022-06-22 06:46:29 UTC T40 RDP
- Tile 2: 2022-06-22 06:46:29 UTC T40 RDQ
- Tile 3: 2022-06-07 06:46:31 UTC T40 RDP
- Tile 4: 2022-06-27 06:46:41 UTC T40 RCP
- Tile 5: 2021-06-05 06:56:21 UTC T39 RYJ
- Tile 6: 2021-06-05 06:56:21 UTC T40 RBP
- Tile 7: 2021-06-12 06:46:21 UTC T40 RDQ
- Tile 8: 2021-06-17 06:46:29 UTC T40 RDN
- Tile 9: 2021-06-17 06:46:29 UTC T40 RCP
- Tile 10: 2021-06-22 06:46:31 UTC T40 RCP

The following Sentinel-2 tiles were used for prediction in this thesis:

- Tile 2: 2022-06-22 06:46:29 UTC T40 RDQ
- Tile 11: 2022-07-02 06:46:29 UTC T40 RDQ

Naming is based on date and timestamp, as well as the grid ID for the covered areas. Figure 4.1 shows the Persian Gulf and the area where the tiles are located. An overview of the grid IDs is shown in Fig.4.2. The tiles chosen have varying sunlight conditions and ocean colors when observing the true color composites.

True color composites of tiles used for prediction

Figure 4.3 shows the true color composite of tile 2. The red rectangles marked on the figure are areas of oil slicks, and are enlarged in Fig. 4.4 to 4.6 The yellow rectangles marked on the figure are areas with complex reflectance due to weather and shallow water conditions, and are enlarged in Fig. 4.7 and

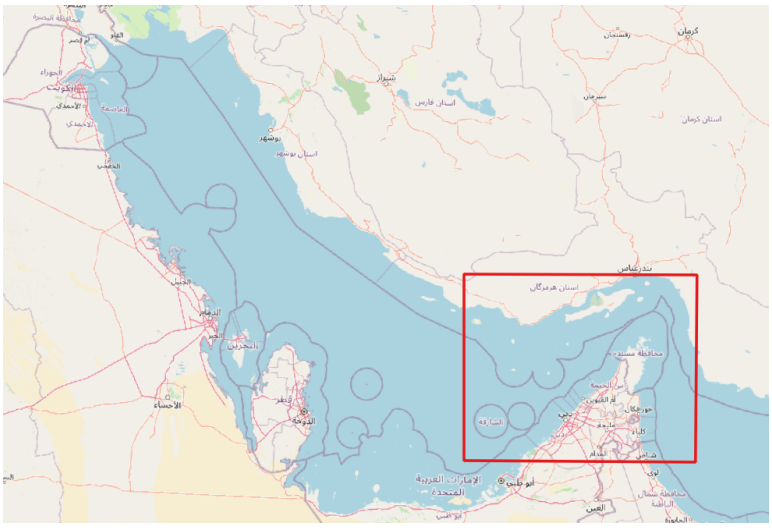


Figure 4.1: Image showing the Persian Gulf. Highlighted area (red square) shows data collection area.

Figure 4.2: Image showing the Sentinel-2 tiling grid for the data collection area.

4.8.

Figure 4.9 shows the true color composite of tile 11. The red rectangle marked on the figure shows the location of the oil slick enlarged in Fig. 4.10.

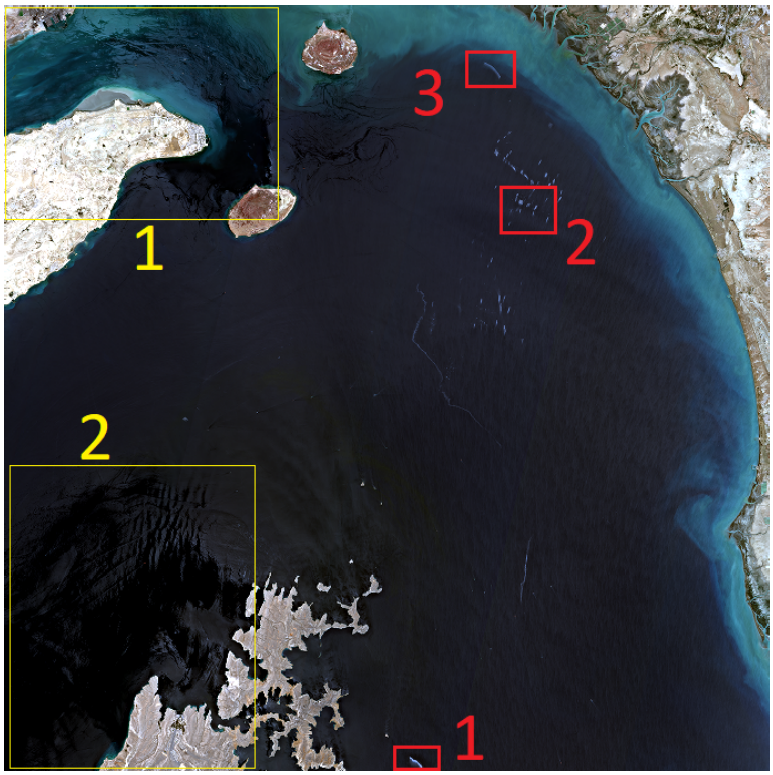


Figure 4.3: True color composite of tile 2.

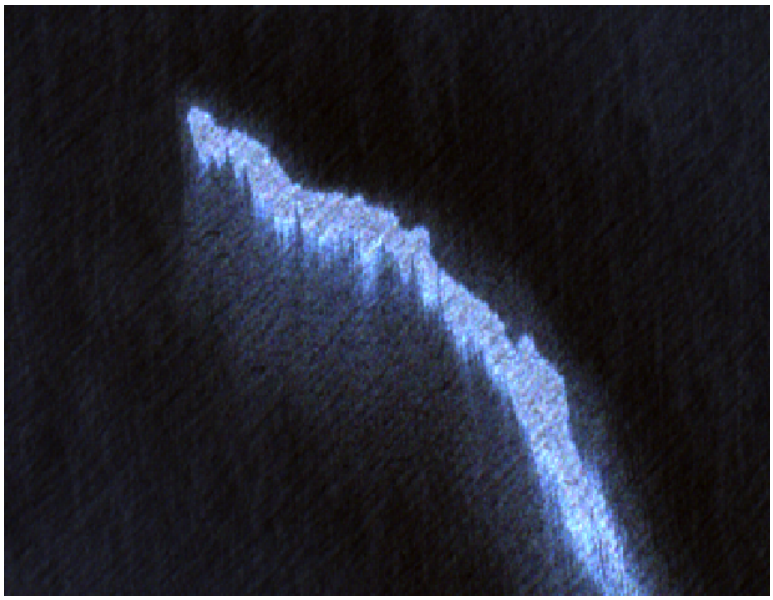


Figure 4.4: Enlarged area of oil slick 1 (red) in tile 2.

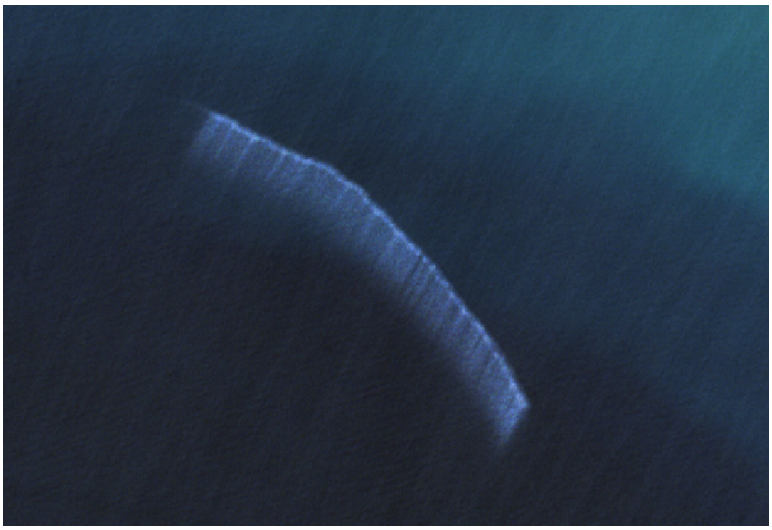


Figure 4.5: Enlarged area of oil slick 2 (red) in tile 2.

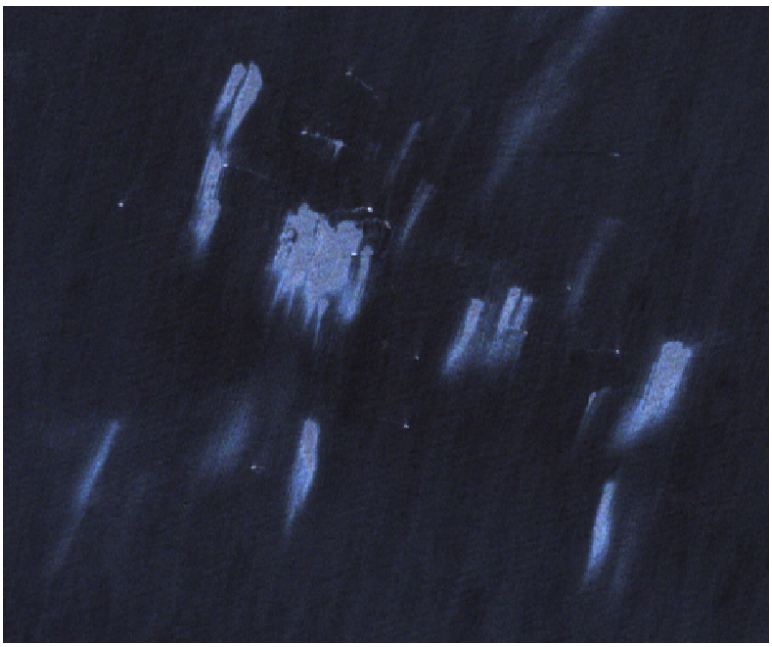


Figure 4.6: Enlarged area of oil slick 3 (red) in tile 2.

4.3.2 Bands

Sentinel-2 have 13 available bands as shown in Fig. 4.11.

For this work we used three VIS bands; band two, three and four, one Near Infrared (NIR) band; band eight, and two short wave infrared (SWIR) bands;

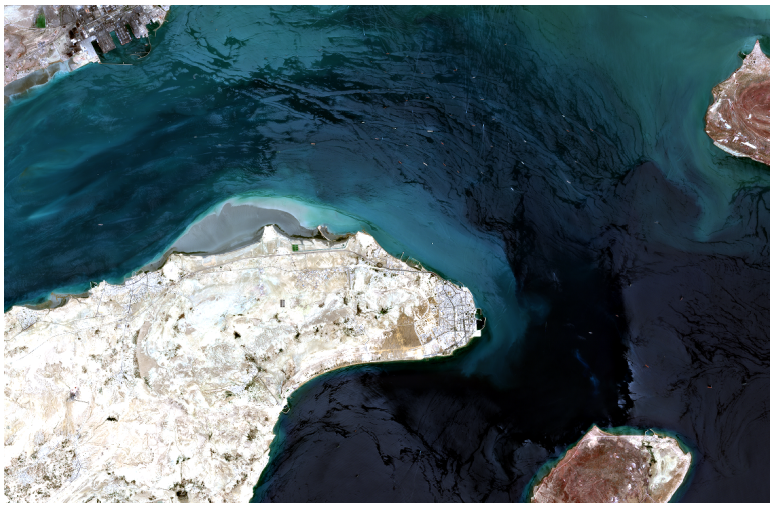


Figure 4.7: Enlarged area of complex area 1 (yellow) in tile 2.

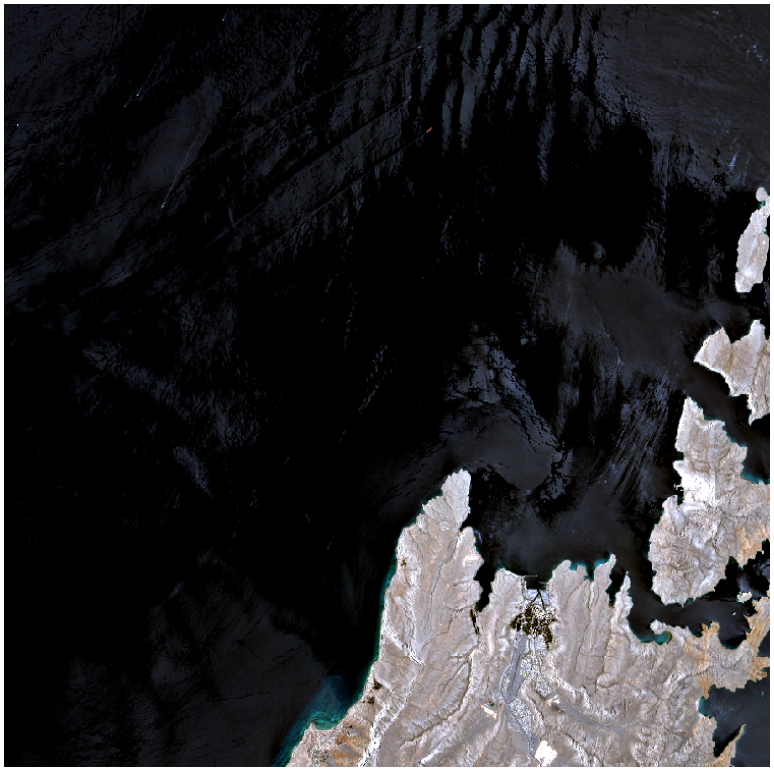


Figure 4.8: Enlarged area of complex area 2 (yellow) in tile 2.

band 11 and 12, for the training, testing and the prediction.

Figure 4.9: True color composite of tile 11.

4.3.3 Label data

The data was annotated in co-operation with experts at Kongsberg Satellite Services, using the Bonn Agreement [7] for thick and thin oil. Figure 4.12 shows the total number of observations in the dataset and the distribution on the classes.

In Fig. 4.12 we see that there were 1 883 511 observations in total. 6 % of those were thick oil with 118 737 observations and 11 % were thin oil with 211 734 observation. 60 % were ocean water with 1 120 986 observation, 22 % turbid water with 415 159 observations and 1 % vessels with 16 895 observations. We tried to include as much thick oil observations as possible from the image data, and at the same time have much more observations from the water classes to roughly imitate real world distribution between the classes. Figure 4.13 show examples of the annotation, one for each class.

Figure 4.13a shows annotation of thick oil with red outline and Fig. 4.13b shows annotation of thin oil with red outline. Figures 4.13c and 4.13d show annotation of respectively ocean water and turbid water, marked with blue outline. Figure

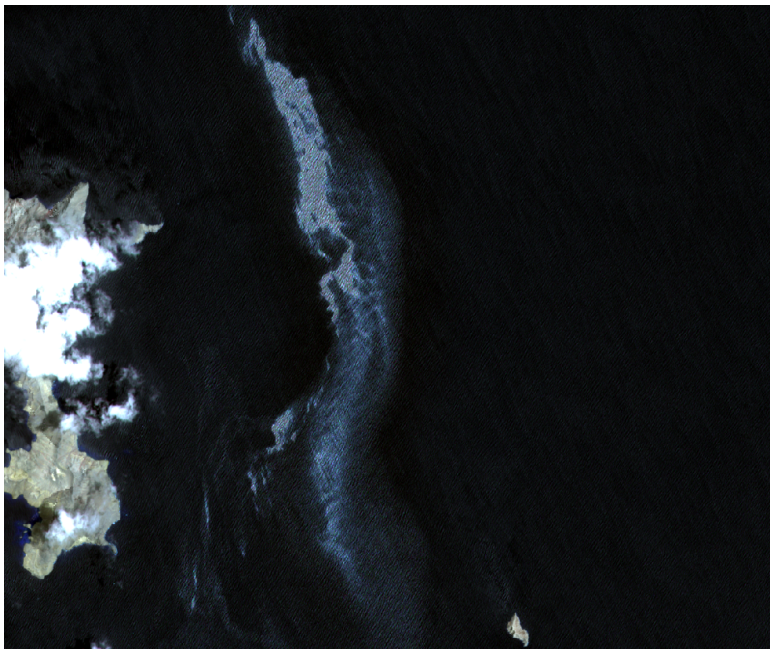


Figure 4.10: Enlarged area of oil slick in tile 11.

4.13e shows annotation of a vessel.

4.4 Spectral response

This section presents spectrum plots based on data from two of the tiles used in this works; tile 2 (Fig. 4.15 and Fig. 4.16) and tile 6 (Fig. 4.17 and Fig. 4.18). Tile 2 was chosen since it contained a larger number of oil slicks samples that could be included in the spectrum plots. Tile 6 was included since it contained great samples from thin oil and had different sunlight conditions than tile 2. The figures were created using the semi automatic classification plug-in in QGIS [51]. The plots are based on the average value of over 100 observations for each class. In most figures we also see the value range for each class. Together they show valuable information on the spectral behaviour of each class in the dataset. Fig. 4.14 shows the labels for Fig. 4.15 to Fig. 4.18.

Tile 2

Figures 4.15 and 4.16 show spectrum plots based on data from tile 2. Figure 4.15a shows the spectrum with all classes. Figure 4.15b shows the spectrum with all classes and range values. Figure 4.15c shows the spectrum without the

	S2A		S2B		
Band Number	**Central wavelength (nm)**	**Bandwidth (nm)**	**Central wavelength (nm)**	**Bandwidth (nm)**	Spatial resolution (m)
1	442.7	20	442.3	20	60
2	492.7	65	492.3	65	10
3	559.8	35	558.9	35	10
4	664.6	30	664.9	31	10
5	704.1	14	703.8	15	20
6	740.5	14	739.1	13	20
7	782.8	19	779.7	19	20
8	832.8	105	832.9	104	10
8a	864.7	21	864.0	21	20
9	945.1	19	943.2	20	60
10	1373.5	29	1376.9	29	60
11	1613.7	90	1610.4	94	20
12	2202.4	174	2185.7	184	20

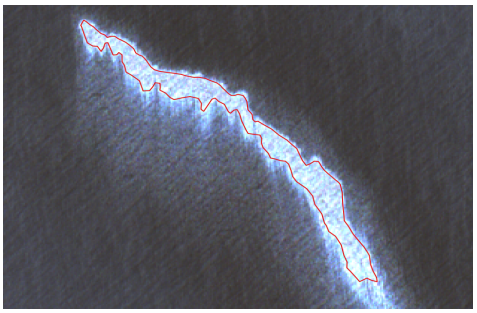
Figure 4.11: Bands in the MSI sensor for S2A and S2B [30].

	Number of observations in dataset	% of total
Total	1 883 511	100 %
Class 1: Thick oil	118 737	6 %
Class 2: Thin oil	211 734	11 %
Class 3: Ocean water	1 120 986	60 %
Class 4: Turbid water	415 159	22 %
Class 6: Vessels	16895	1 %

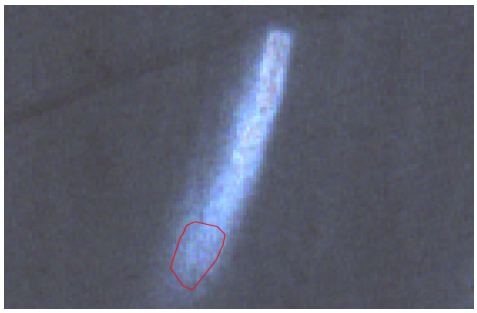
Figure 4.12: Number and distribution of observations in the classes.

vessel class. Figure 4.15d shows the spectrum without the vessel class, but with range values.

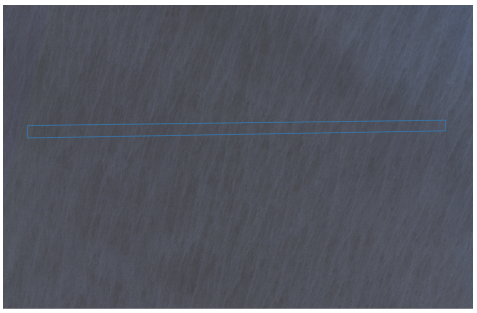
Figure 4.16a shows the spectrum with the thick and thin oil classes, including



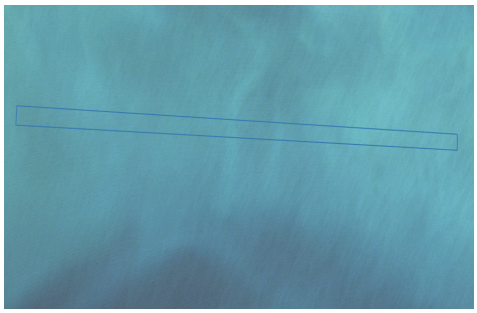
(a) Annotation of thick oil.



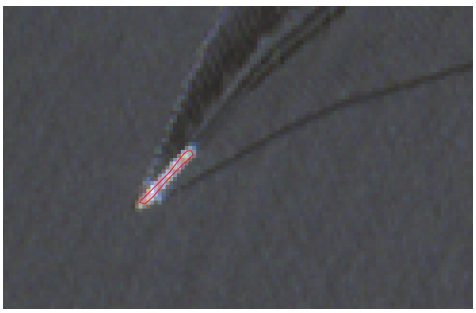
(b) Annotation of thin oil.



(c) Annotation of ocean water.



(d) Annotation of turbid water.



(e) Annotation of vessel.

Figure 4.13: Examples of annotations performed is this work for all classes.

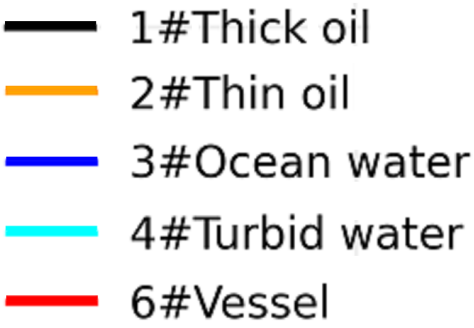


Figure 4.14: Class labels for Fig. 4.15, Fig. 4.16, Fig. 4.17 and Fig. 4.18.

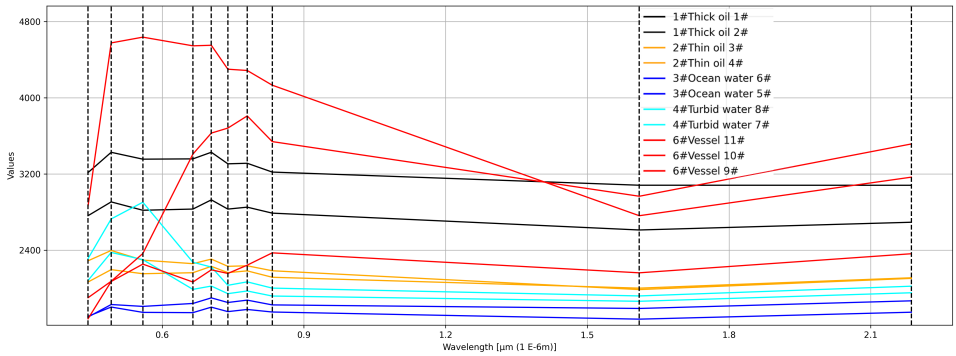
range values. Figure 4.16b shows the spectrum with the ocean and turbid water classes, including range values. Figure 4.16c shows the spectrum with the thin oil and ocean water classes, including range values. Figure 4.16d shows the spectrum with the thin oil and turbid water classes, including range values.

It can be seen in Fig. 4.15 and Fig. 4.16 that the thick oil samples have a higher reflectance than the thin oil samples and the water samples on average for all wavelenghts. However, when including the max and min range we see a large overlap between the thick oil and thin oil in the VIS spectrum. Turbid water also has a small overlap with thick oil in the VIS spectrum. Turbid water overlaps greatly with thin oil in the VIS spectrum. In the NIR region we see a small overlap between thick and thin oil. In the SWIR region there is no overlap between thick and thin oil. The vessel class have a large overlap with thick and thin oil in all wavelengths, except for in the SWIR range over 2 μm where the vessel class only overlaps with thick oil.

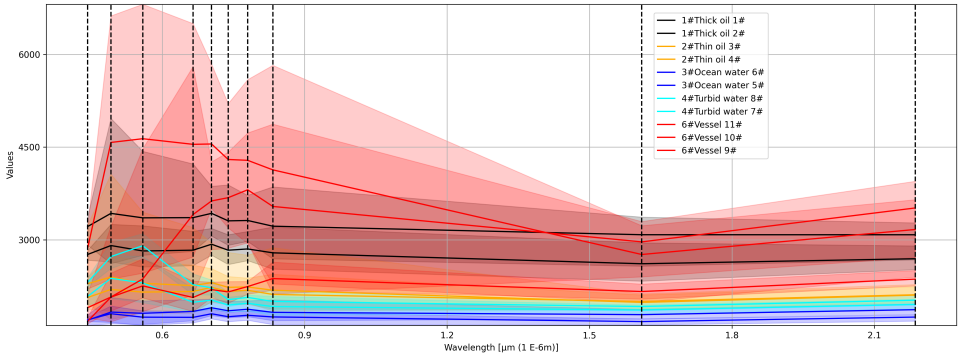
Tile 6

Figures 4.17 and 4.18 shows spectrum plots based on data from tile 6. Figure 4.17a shows the spectrum with all classes. Figure 4.17b shows the spectrum with all classes and range values. Figure 4.17c shows the spectrum without the vessel class. Figure 4.17d shows the spectrum without the vessel class, but with range values.

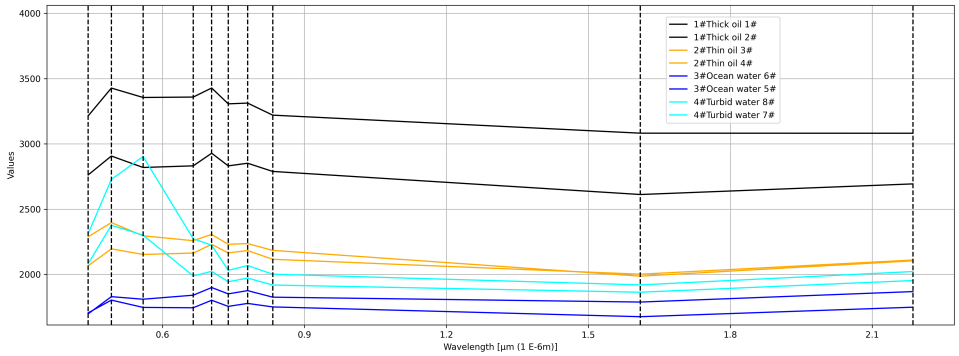
Figure 4.18a shows the spectrum with the thick and thin oil classes, including range values. Figure 4.18b shows the spectrum with the ocean and turbid water classes, including range values. Figure 4.18c shows the spectrum with the thin oil and ocean water classes, including range values. Figure 4.18d shows the spectrum with the thin oil and turbid water classes, including range val-



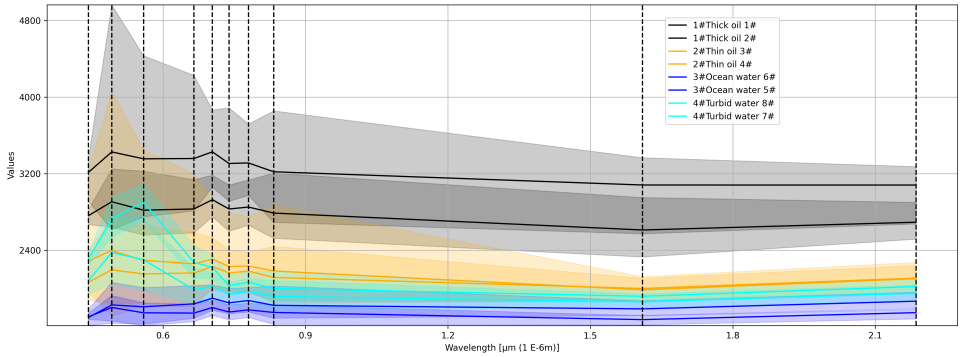
(a) Spectrum plot based on samples from Tile 2 with all classes.



(b) Spectrum plot based on samples from Tile 2 with all classes. Showing max and min range.

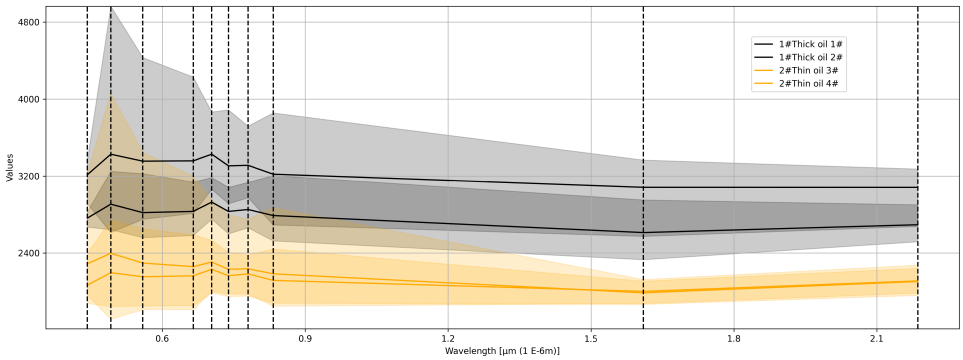


(c) Spectrum plot based on samples from Tile 2 without vessel class.

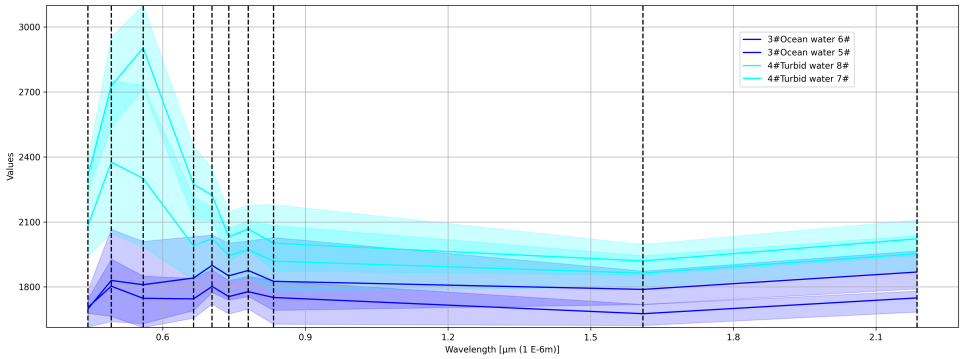


(d) Spectrum plot based on samples from Tile 2 without vessel class. Showing max and min range.

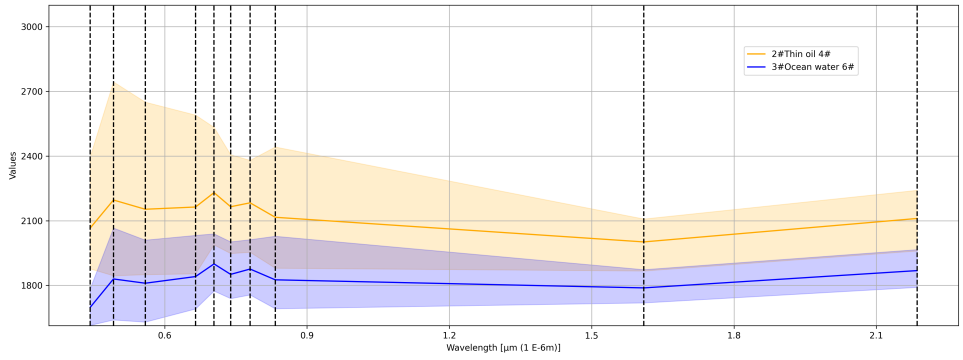
Figure 4.15: Spectrum plots based on samples from Tile 2. The spectrum plots are averages of over 100 samples.



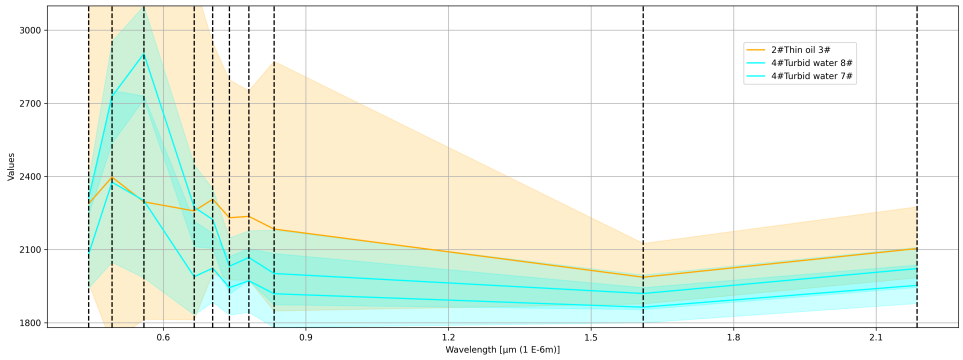
(a) Spectrum plot based on samples from Tile 2 with thin and thick oil. Showing max and min range



(b) Spectrum plot based on samples from Tile 2 with ocean and turbid water. Showing max and min range.

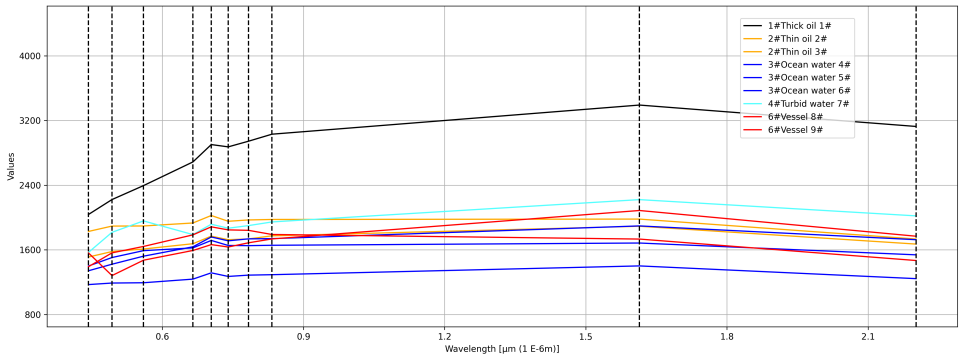


(c) Spectrum plot based on samples from Tile 2 with thin oil and ocean water.Showing max and min range

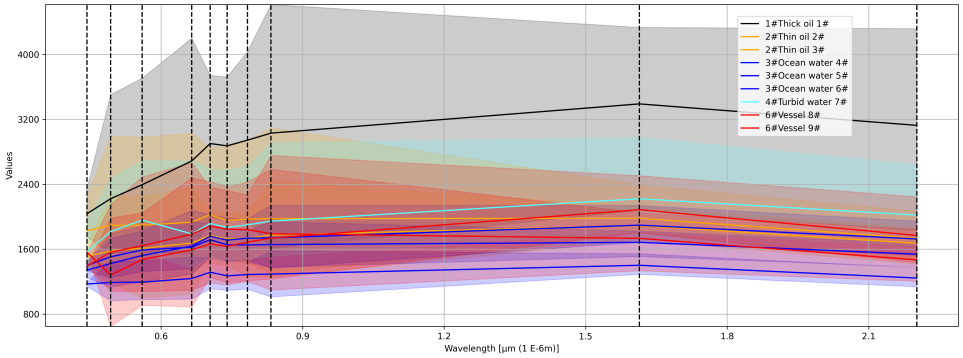


(d) Spectrum plot based on samples from Tile 2 with thin oil and turbid water. Showing max and min range.

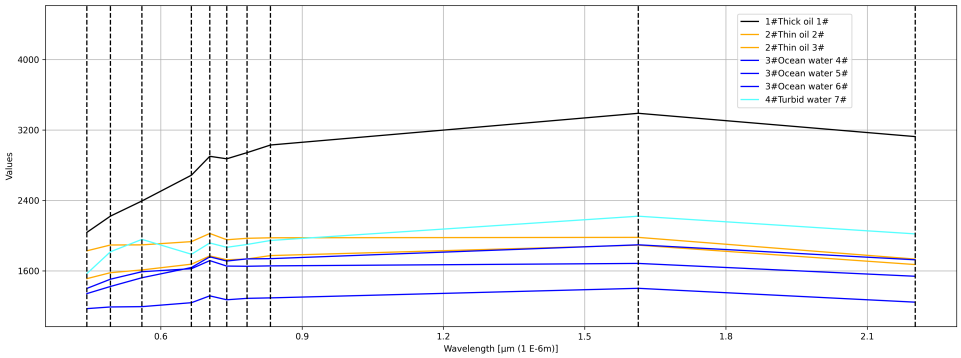
Figure 4.16: Spectrum plots from Tile 2. The spectrum plots are averages of over 100 samples.



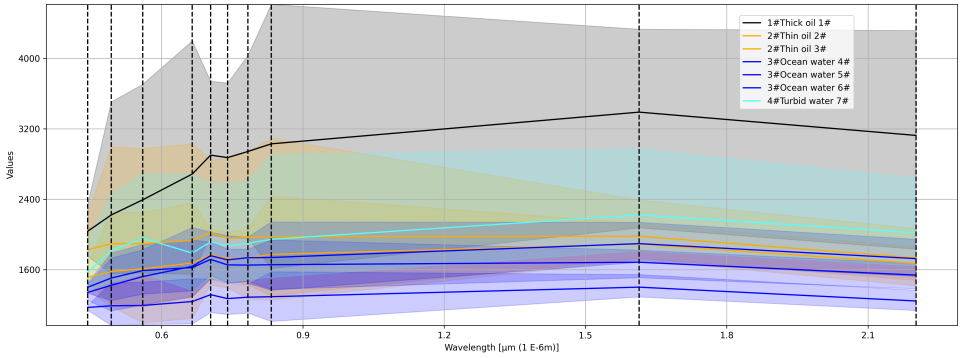
(a) Spectrum plot based on samples from Tile 6 with all classes.



(b) Spectrum plot based on samples from Tile 6 with all classes. Showing max and min range.



(c) Spectrum plot based on samples from Tile 6 without vessel class.



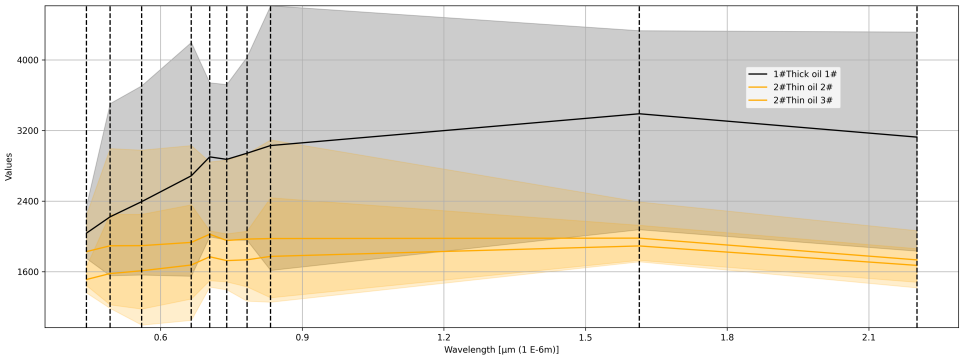
(d) Spectrum plot based on samples from Tile 6 without vessel class. Showing max and min range.

Figure 4.17: Spectrum plots based on samples from Tile 6. The spectrum plots are averages of over 100 samples.

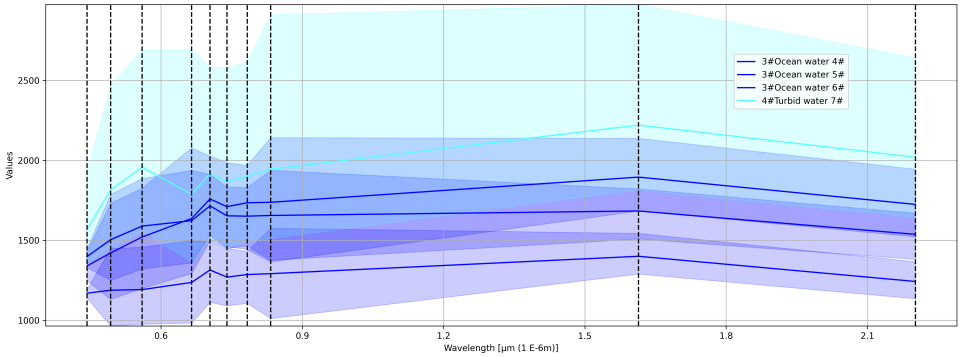
ues.

The spectrum plots in Fig. 4.15, Fig. 4.16, Fig. 4.17, and Fig. 4.18 shows a different separability for the classes in the SWIR wavelengths compared to the VIS wavelengths and indicates that the inclusion of the SWIR bands in this work can have a beneficial effect on the ML models.

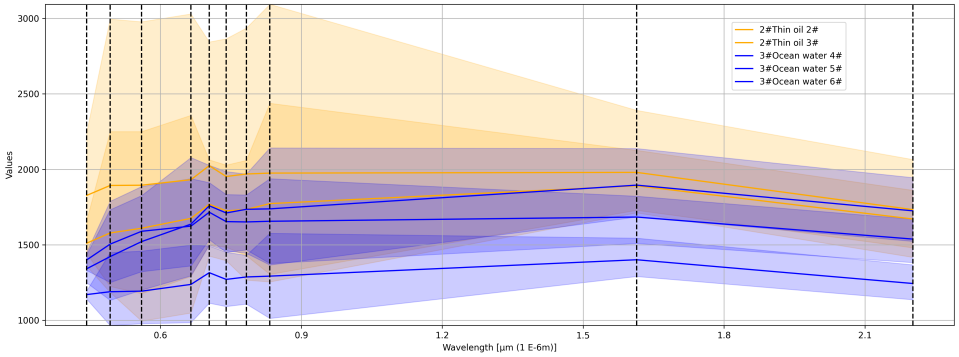
Just like in in Fig. 4.15 and Fig. 4.16, Fig. 4.17 and Fig. 4.18 show that the thick oil samples have a higher reflectance than the thin oil samples and the water samples on average for all wavelenghts. Again, there is a large overlap between the the thick oil and the thin oil from 0.4 μm to 1.2 μm when including max and min range. Both ocean water and turbid water greatly overlap with thin oil in all wavelengths. It can also bee seen that turbid water overlaps with thick oil in all wavelengths. The vessel class overlaps with all classes except with thick oil in the spectrum from 0.4 μm to 2.2 μm.



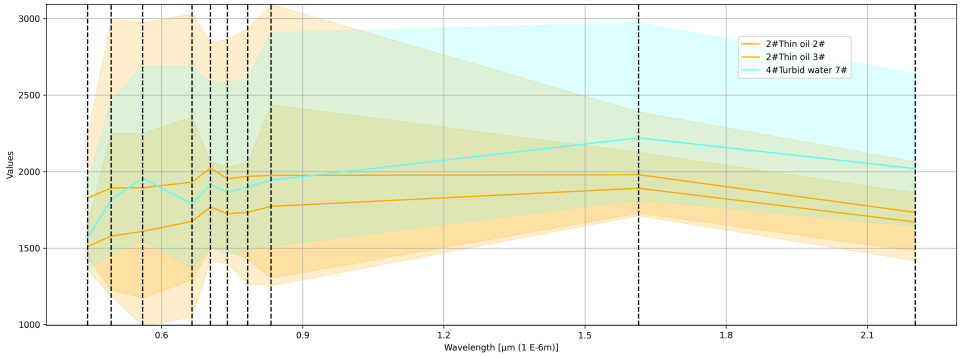
(a) Spectrum plot based on samples from Tile 6 with thin and thick oil. Showing max and min range



(b) Spectrum plot based on samples from Tile 6 with ocean and turbid water. Showing max and min range.



(c) Spectrum plot based on samples from Tile 6 with thin oil and ocean water.Showing max and min range



(d) Spectrum plot based on samples from Tile 6 with thin oil and turbid water. Showing max and min range.

Figure 4.18: Spectrum plots from Tile 6. The spectrum plots are averages of over 100 samples.

5 Methodology

This section describes the methodology used in this works. Section 5.1 to Sec. 6.4 describe the process from starting working on the labeling of the data, to the prediction applying our trained classifiers on unlabeled data. Section 5.7 to Sec. 5.8 describe the configurations for the classifiers and the software used in this thesis.

5.1 Annotation

Annotation is the manual work of assigning classes (labels) to pixels in the remotely sensed image by drawing polygons around the selected pixels. In this thesis the annotation was done on true color composite images with 10 m spatial resolution from the remotely sensed data using QGIS software, see Sec.5.8. True color composites were used so that we could utilize the Bonn agreement, see Sec. 2.3.3.

We initially chose six classes to work with because we wanted to include algae in the thesis. This needed to be changed to five because none of the selected images included an algal bloom event. The classes were as follows; thick oil (class 1), thin oil (class 2), ocean water (class 3), turbid water (class 4) and vessels (class 6), as shown in Tab. 5.1 The oil was identified in collaboration with KSAT expert analysts using the Bonn agreement oil appearance code [7]. The annotated shapefiles were exported to PNG format, and imported to Matlab.

Thick oil was defined as metallic and true colored oil (code 3, 4 and 5) and thin oil was sheen and rainbow (code 1 and 2), based on the Bonn agreement. See Sec. 2.3.3.

Table 5.1: Classes used in this thesis

Class	Class name
1	Thick oil
2	Thin oil
3	Ocean water
4	Turbid water
6	Vessel

5.2 Band selection

The Multi spectral instrument (MSI) sensor used by Sentinel-2 have 13 bands with resolutions from 10 m to 60 m depending on the band, see Fig. 4.11. Due to the varying size of the identified oil slicks we needed the highest resolution. We included band two, three, four and eight, centered at 492 nm, 559 nm, 665 nm and 833 nm due to the resolution of 10 m. Relevant literature [14] shows promising results for oil classification using band 11 and band 12 at the short wave infrared (SWIR) range with 20 m resolution. The spectral response curves shown in Fig. 4.16a and Fig. 4.18a also shows that thick and thin oil are more separable in the SWIR range. Therefore, we decided to include band 11 and 12 in the initial training phase, where we would train classifiers with four bands and with six bands. To use band 11 and band 12 in combination with the 10 m spatial resolution bands, we had to perform up-sampling. This is described in Sec.5.3. Based on the results from the first training phase, we would then proceed with the band combination showing the highest accuracy for thick oil.

5.3 Upsampling

To include SWIR bands 11 and 12 centered at 1612 nm and 2195 nm we had to up-sample the bands from 20 m to 10 m spatial resolution. This was done through the ESA Sentinel Application Platform (SNAP) tool, see Sec.5.8, using bilinear upsampling. Three different resampling techniques were considered; Nearest neighbor, Bilinear interpolation and Cubic convolution. The Nearest neighbor method was excluded due to the possible position error connected with this technique [52], and the Cubic convolution is only slightly better, but

more time consuming [52]. Due to this we chose to use Bilinear interpolation up-sampling. The Bilinear method finds the four nearest pixels to a point in the original image, and calculates the weighted average. The point does not need to be in the center of the four pixels, so the weight part is the distance, either from the middel or upper left corner, from the four pixels to the given point. The weighted average is the assigned value for the new pixel. Up-sampling took about one hour for each tile.

5.4 Dataset

When the bands where selected and up-sampled to correct resolution, QGIS [51] was used to extract the bands and merge the data into ENVI header(hdr) files. This file type was chosen because it stores metadata and it can be easily imported to programming environments such as Matlab. Matlab was used to merge the remotely sensed data with the labeled data, creating the training data for the classifiers [53].

5.5 Training and testing

This section describes the methodology concerning training and testing of the dataset. We decided to split this part of the work into three phases: phase 1, a selection phase were multiple ML models were trained and tested on parts of the dataset, phase 2, a configuration phase were the best performing ML models from phase 1 were tuned, trained and tested on parts of the dataset, and phase 3, the final phase were the best performing ML models from phase 2 were trained and tested on the whole dataset. In all training phases, 75 % of the data was used for training and 25 % was used for testing. Alternatively, the dataset could be split into three parts, training-validation-testing. However, due to the setup of this work with three phases, where phase 2 is testing different parameters for each ML model, we chose a two-split, setting a part of the data for validation.

The testing is used to evaluate the accuracy of the trained ML models. Two test methods were considered, the Holdout validation and the Cross-validation method. Holdout validation simply splits the data into a training part and a validation part. In Cross-validation, also called k-fold cross validation, data is randomly split into a k number of folds (groups). Using this method, the classifier will be trained a k number of times, each time, a unique fold will be the validation fold, and the rest will be the training folds. The output will then be the average accuracy over all folds. The Cross-validation gives a better indica-

tion on the performance of the classifier because it is trained on multiple sets of the same data [42]. However, the method is more time consuming than the Holdout validation, especially for large datasets. Due to the size of the dataset in this thesis, the Holdout validation was considered the preferred option. All training and testing was performed in Matlab using the "classification learner" app from the "statistics and machine learning" toolbox, see Sec.5.8.

5.5.1 Phase 1 - Selection

The goal for the first phase was to test multiple classifiers and evaluate their performance on a part of the dataset, about 10 %, which was chosen to be representative for the whole dataset. A second goal was to investigate the affects when including SWIR bands in the dataset. This meant that we had two datasets for phase one. One with four bands, including band two, three, four and eight (centered at 492 nm, 559 nm, 665 nm and 833 nm), and one with six bands, including band two, three, four, eight, eleven and twelve (centered at 492 nm, 559 nm, 665 nm and 833 nm, 1612 nm and 2195 nm). The list of classifiers include; KNN classifiers, NN classifiers, decision trees classifiers, naive bayes classifiers, and discriminant classifiers. The KNN and decition trees were included due to that they are simple, time-efficient and effective classifiers [41]. We included six KNN classifiers; a fine KNN (1 neighbor), a medium KNN (10 neighbors), a coarse KNN (100 neighbors), a cosine KNN (10 neighbors), a cubic KNN (10 neighbors) and a weighted KNN (10 neighbors). We included three decition trees; a fine Tree (maximum 100 splits), a medium Tree (maximum 20 splits) and a coarse Tree (maximum 4 splits).

NN are one of the most popular classifier used in oil detection today [38]. NN can greatly vary in complexity, therefore two simple and time-efficient variants were included in this works; A single-layered feed forward neural network with 25 nodes and a tri-layered feed forward neural network with 12 nodes in each layer. Two Naive bayes models were incluced; The Gaussian naive bayes due to that is simple and time-efficient, and the more time consuming Kernel naive bayes, to compare the performance with the gaussian Naive bayes. The Naive bayes are based on Bayes Theorem. Two discriminant models were included: a Quadratic discriminat model and a Subspace discriminant model. The discriminant classifiers are based on discriminant analysis and were included due that they are simple and time-efficient [41].

The performance of all ML models were evaluated and ranked based on the accuracy of classifying thick oil and the overall accuracy. We then proceeded with the most promising dataset (four bands versus six bands) and the four most promising classifiers were chosen for further evaluation. The four classifiers were fine KNN, weighted KNN, a single-layered neural network and a tri-layered

neural network. The configuration for each classifier is described in Sec.5.7. The result from phase 1 is described in Sec.6.1.

5.5.2 Phase 2 - Configuration

The first phase showed that the performance was best using all six bands. The goal for the second phase was then to test the four best classifiers from phase one on a larger portion of the six-band dataset. For this phase about 30 % of the dataset was used with different configuration for the classifiers. For the KNN classifiers this meant testing different K-values, and for the ANN classifiers this meant testing different layer sizes. Again, their performance were evaluated and ranked based on the accuracy of classifying thick oil and the overall accuracy. We proceeded with the most promising configuration for the classifiers, that would be the base for phase 3. The result from phase two is described in Sec.6.2.

5.5.3 Phase 3 - Test results

In the last phase we ran the full dataset on the four classifiers with the best modified setup from phase two. The results from phase 3 is described in Sec.6.3, including confusion matrices and overall accuracy.

5.6 Prediction

The models were used to classify two Sentinel-2 tiles for visual interpretation. The first was tile 2, which was part of the training data. The second, tile 11 was not included in the training data, and can therefore be seen as independent from the training data. The prediction has no reference values, and we can only speculate about the quality of the model observing the prediction. However, the prediction gives an overview on how the models will perform in a practical situation, for instance as aid for operators at KSAT. Finally all result, described in Sec.6 were discussed as described in Sec.7

5.7 Classifier setup

In this section we will describe the classifier setups. For all classifiers the data was normalized before training.

After phase 2 of the training, we worked with two KNN classifiers and two NN classifiers. The first was a fine KNN classifier with 2 nearest neighbors. The distance measure was euclidean and the distance weight was equal. The second was a weighted KNN classifier with 8 nearest neighbors. The distance measure was also euclidean, but the distance weight was squared inverse.

The neural network classifiers were feedforward fully connected models. For both classifiers, we used the Rectified Linear Unit(ReLU) as the activation function. The iteration limit was set to 1000, and the regularization strength was set to 0. The first was a simple single-layer NN classifier with 25 nodes. The second was a tri-layered NN classifier with 12 nodes in each layer.

5.8 Software

In this section we will explain the software used in this thesis. These applications are QGIS, ESA Snap and the Classification Learner app in Matlab.

Quantum Geographic Information System (QGIS) is free and open source GIS software, used to analyse and edit spatial information, with a wide range of applications [51]. For this thesis we used version 3.22.13. See "https://qgis.org/en/site" for more information.

The European Space Agency (ESA) Sentinel Application platform (SNAP) is a free Earth observation analysis app. It is used for processing data products from a wide range of satellite missions, including EU's Copernicus Sentinel-1, Sentinel-2, Sentinel-3 and ESA's SMOS mission. "SNAP enables people to explore, analyse and process remote-sensing data, facilitating cutting-edge scientific research, education and training activities, and the development of a wide range of operational applications" [54].

The Classification Learner is an application in Matlab used to train models for data classification. The app has a wide range of different classifiers used in supervised machine learning. Here you can train desired classifiers and validate their performance, and also export the trained models if needed [55]. See "https://se.mathworks.com/help/stats/classificationlearner-app.html" for more information about the app.

6 Results

This section presents the results of several machine learning algorithms evaluated on remotely sensed data. The performance of the methods is assessed by computing the confusion matrix, TPR and OA. Then the selected models are applied to MSI S2 images to visually evaluate their performances, and hence simulate the use of such models in operational environments. As explained in Sec. 5.5, the training and testing was done in three phases. Section 6.1 shows the results of evaluating a selection of ML models. Sec. 6.2 shows the results of evaluating different configurations of the selected ML methods based on their performances of predicting thick oil from phase 1. Section 6.3 presents the resulting final ML models. These models are further studied by computing confusion matrices and the OA. Section 6.4 shows the prediction when using the selected ML models that had the highest accuracy in thick oil classification to classify unlabeled remotely sensed data.

6.1 Phase 1

Figure 6.1 shows the results after training and testing a various well recognized ML models on about 10 % of the labeled data. This data (10 % of the total dataset) was further split into 70 % for training and 30 % for testing.

The first column shows the selected classifiers. The second column shows the TPR for thick oil using 4 bands, while the third column shows the TPR for thick

Classifier	TPR (thick oil, 4 bands)	TPR (thick oil, 6 bands)	OA (4 bands)	OA (6 bands)
Fine tree	79.4 %	79.9 %	99.3 %	99.4 %
Medium tree	79.1 %	79.4 %	99.2 %	99.3 %
Coarse tree	79.1 %	78.4 %	98.9 %	99.1 %
Fine KNN	82.7 %	86.3 %	99.0 %	99.3 %
Medium KNN	81.6 %	85.4 %	99.3 %	99.5 %
Coarse KNN	81.1 %	83.6 %	99.3 %	99.4 %
Cosine KNN	77.2 %	81.3 %	99.0 %	99.3 %
Cubic KNN	81.5 %	85.5 %	99.3 %	99.5 %
Weighted KNN	82.6 %	86.5 %	99.3 %	99.5 %
Single-layer NN	81.6 %	86.0 %	99.4 %	99.5 %
Tri-layered NN	82.2 %	86.2 %	99.4 %	99.5 %
Gaussian naive bayes	77.8 %	77.9 %	97.5 %	97.6 %
Kernel naive bayes	78.5 %	79.1 %	97.7 %	97.6 %
Quadratic discriminant	80.1 %	79.4 %	98.5 %	98.6 %
Subspace discriminant	74.6 %	69.6 %	98.6 %	98.8 %

Figure 6.1: The selected classifiers and their performance after phase 1; TPR for thick oil (class 1), and overall accuracy, using two datasets, one with 4 bands and one with 6 bands. The green background in a cell shows which of the band combinations, 4 bands or 6 bands, that have the highest accuracy for the given classifier in the same row.

oil using 6 bands. The fourth column shows the OA using 4 bands, and the fifth column shows OA using 6 bands. Green background in a cell shows which of the band combinations, 4 bands or 6 bands, that has the highest accuracy for the given classifier in the same row. Overall the figure shows that both the TPR and OA were higher when using six bands. This is why we wanted to proceed using the dataset with 6 bands. The best results were obtained by using the KNN and NN classifiers based on the TPR for thick oil. Then the top four models for TPR thick oil were as follow:

- Weighted KNN 86.5 %
- Fine KNN, 86.3 %
- Tri-layered NN 86.2 %
- Single-layer NN 86.0 %

That is why we chose to further study and compare these ML models.

6.2 Phase 2

Figure. 6.2 shows the results after training and testing the best performing ML models from phase 1 with different configurations. In this case about 30 % of the total labeled data was used for training and testing the methods.

Classifier	TPR (thick oil)	Overall Accuracy	Training time (sec)	Config
Fine KNN	88.4 %	96.1 %	21	1 nn
Fine KNN	93.8 %	96.1 %	27	2 nn
Fine KNN	91.5 %	97.1 %	19	3 nn
Fine KNN	92.7 %	97.1 %	37	4 nn
Medium KNN	91.3 %	97.2 %	15	5 nn
Medium KNN	92.0 %	97.3 %	42	6 nn
Medium KNN	91.5 %	97.3 %	50	8 nn
Medium KNN	91.4 %	97.3 %	53	10 nn
Medium KNN	91.4 %	97.3 %	52	12 nn
Medium KNN	91.2 %	97.3 %	43	14 nn
Medium KNN	91.1 %	97.3 %	44	16 nn
Weighted KNN	88.4 %	96.1 %	17	1 nn
Weighted KNN	88.4 %	96.1 %	29	2 nn
Weighted KNN	91.1 %	97.1 %	38	4 nn
Weighted KNN	91.1 %	97.2 %	43	6 nn
Weighted KNN	91.1 %	97.3 %	52	8 nn
Weighted KNN	91.1 %	97.3 %	56	10 nn
Weighted KNN	91.1 %	97.3 %	55	12 nn
Weighted KNN	91.0 %	97.3 %	45	14 nn
Weighted KNN	91.0 %	97.3 %	47	16 nn
Single-layer NN	89.4 %	97.1 %	1718	20 N
Single-layer NN	90.8 %	97.2 %	2526	25 N
Single-layer NN	89.8 %	97.2 %	2727	30 N
Tri-layered NN	91.0 %	97.3 %	2720	10 N per layer
Tri-layered NN	91.4 %	97.2 %	1984	12 N per layer
Tri-layered NN	91.2 %	97.4 %	2597	14 N per layer

Figure 6.2: The selected classifiers and their performances after phase 2; TPR for thick oil, overall accuracy, training time, and configuration (nn = Nearest Neighbor, N = Nodes). The green background in a row shows the configuration giving the best result for each of the chosen ML models.

Similar to phase 1, the data (30 % of the total labeled data) was split into 70 % for training and 30 % for testing. The first column in Fig. 6.2 shows the

classifiers. The second column shows the TPR for thick oil (here only using the 6 band combination), while the third column shows the OA. The fourth column shows the training time used for training each model and the fifth column is the model specific configuration. For the KNN models, various number of Nearest Neighbors (nn) were studied. This corresponds to the "config" column in Fig. 6.2. For the NN models the number of nodes in the network were varied and evaluated. A green background shows the configuration giving the best results for each of the chosen ML models.

In the column showing training time in Fig. 6.2, for the KNN models, the training time was short and always under 1 min. For the NN models the training time was longer, and took between 28 min and 46 min depending on the configuration. The best configurations for each ML model were:

- Fine KNN using 2 neigbors
- Weighted KNN using 8 neighbors
- Single-layer NN using 25 neurons
- Tri-layered NN using 12 neurons in each layer

That is why we chose to go forward with these configurations for the ML models for further evaluation in the rest of this thesis.

6.3 Phase 3

Figure 6.3 shows the results after training and testing the best performing ML models from phase 2 on 100 % of the labeled data.

Classifier	TPR (Thick oil)	OA	Training time (sec)	Config
Fine KNN	84.2 %	92.1 %	24	2 nn
Weighted KNN	78.3 %	94.1 %	29	8 nn
Single-layer NN	64.1 %	92.1 %	2391	25 N
Tri-layered NN	64.5 %	91.9 %	2390	12 N per layer

Figure 6.3: The selected classifiers and their performances after phase 3; TPR for thick oil, overall accuracy, training time, and configuration (nn = Nearest Neighbor, N = Nodes).

As in phase 1 and 2, the data was split into 70 % training data and 30 % testing data. Figure 6.3 has the same setup as Fig. 6.2, TPR for thick oil, OA, training

time and configuration. It can be seen in Fig. 6.3 that the TPR for thick oil have decreased for all classifiers compared to using only 30 % of the data. A reason for this can be the difference in the reflectance values in the dataset due to varying sun angle. For thick oil, the fine KNN has a TPR of 84.2 %, Weighted KNN 78.3 %, single-layer NN 64.1 % and tri-layered NN 64.5 %. The overall accuracy has also decreased. The OA for fine KNN is 92.1 %, for weighted KNN 94.1 %, single-layer NN 92.1 % and tri-layered NN 91.9 %.

Section 6.3.1 to 6.3.4 presents the confusion matrices after testing the selected four ML models.

6.3.1 Model 1 - Fine K-nearest neighbors

Figure 6.4 shows the confusion matrix for model 1, fine KNN. For class one, thick oil, we see a TPR of 84.2 %. Most incorrectly classified pixels were classified as class two or three. For class two, thin oil, we see a TPR of 79.3 %, where most incorrectly classified pixels were classified as either thick oil or ocean water. Class three and four, ocean water and turbid water, had the highest TPR with 94.4 % and 97.2 %. The vessel class only scored a TPR of 23.8 %, where almost half of the pixels where misclassified as ocean water.

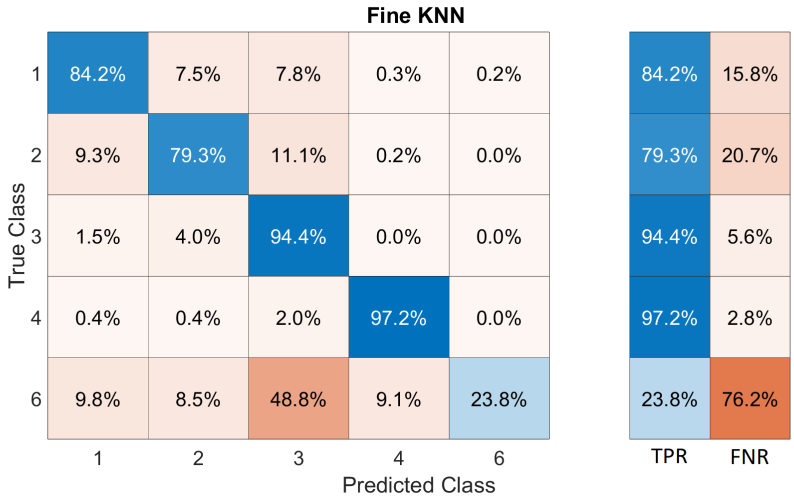


Figure 6.4: Confusion matrix for model 1, fine KNN with 2 nearest neighbours.

6.3.2 Model 2 - Weighted K-nearest neighbor

Figure6.5 shows the confusion matrix for model 2, weighted KNN. The results were similar to the fine KNN model. For class one, thick oil, the TPR is 78.3

%. Most incorrectly classified pixels were classified as class two (thin oil) or three (oceanic water). For class two, thin oil, the TPR was 79.5 %, where most incorrectly classified pixels were classified as ocean water. Class three and four, ocean water and turbid water, respectively, had the highest TPR of 98.4 % and 97.3 %, respectively. The vessel class only scored a TPR of 26.9 %, where over half of the pixels were misclassified as ocean water.

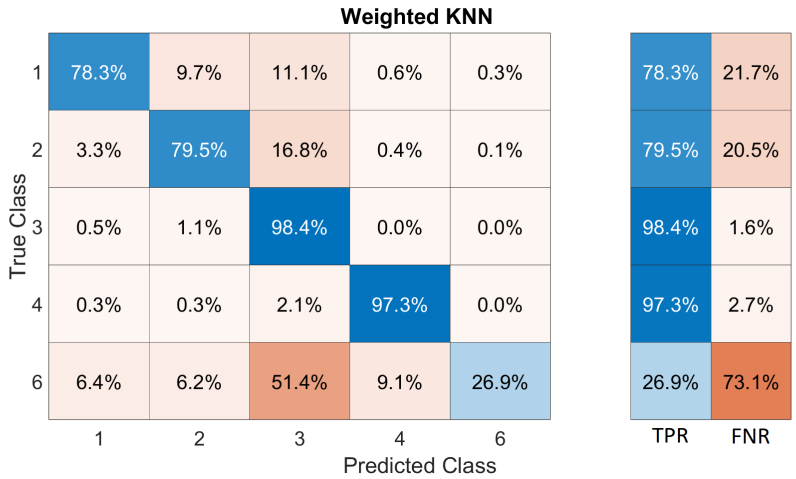


Figure 6.5: Confusion matrix for model 2, weighted KNN with 8 nearest neighbours.

6.3.3 Model 3 - Single-layer neural network

Figure 6.6 shows the confusion matrix for model 3, the single-layer NN. For class one, thick oil, we see a TPR of 64.1 %. Most incorrectly classified pixels were classified as class two or three, similarly to the first two models. For class two, thin oil, the TPR is 71.9 %, and the most incorrectly classified pixels were classified as ocean water. These results are similar to the ones obtained by using model 2. Class three and four, ocean water and turbid water, respectievly, had again the highest TPR of 98.1 % and 97.1 %. The vessel class only scored a TPR of 18.5 %, where over half of the pixels where misclassified as ocean water.

6.3.4 Model 4 - Tri-layered neural network

Figure 6.7 shows the confusion matrix for model 4, the tri-layered NN. For class one, thick oil, we see a TPR of 64.5 %. Most incorrectly classified pixels were classified as class two or three, similar to all the other models. For class two, thin oil, we see a TPR of 73.3 %, where most incorrectly classified pixels were classified as ocean water, similar to model 2 and 3. Class three and four, ocean

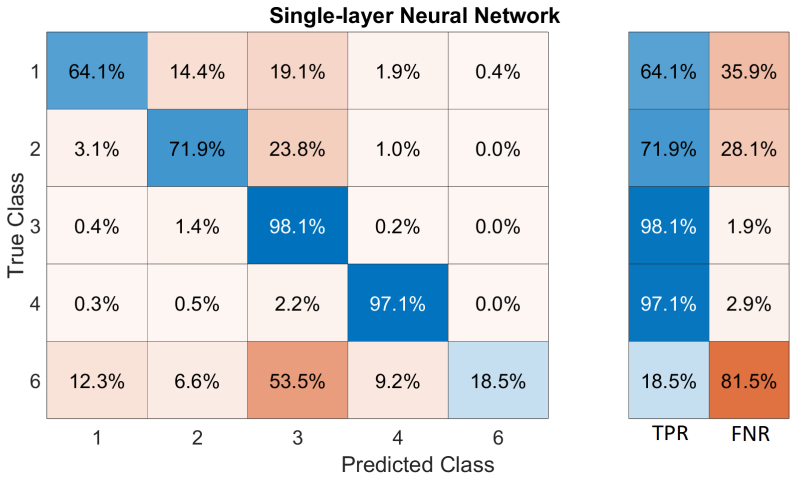


Figure 6.6: Confusion matrix for model 3, single-layer NN with 25 nodes.

water and turbid water, had again the highest TPR of 97.9 % and 96.7 %. The vessel class only scored a TPR of 4.8 %, where over half of the pixels where misclassified as ocean water just like for the other models.

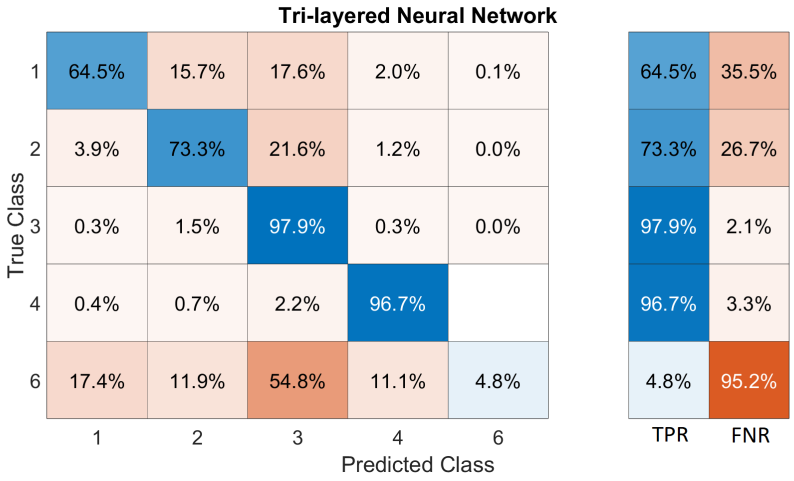


Figure 6.7: Confusion matrix for model 4, tri-layered NN with 12 nodes in each layer.

Based on the computed statistical measures for the models some similarities can be observed. When thick oil was misclassified, it was mostly classified to thin oil and/ or ocean water. Model 2, 3 and 4 misclassified mostly thin oil as ocean water, while model 1 misclassified mostly thin oil as thick oil and ocean water. All models had high TPR for ocean water and turbid water, while having a poor TPR for vessels.

6.4 Prediction

This section presents the prediction power of the classifiers in two different ways. Section 6.4.1 presents the results when the training data was included in the prediction data. Hence, the models have already seen parts of this data during training, and the prediction will not reflect the models true ability to classify unseen data. On the other hand, Sec. 6.4.2 evaluates the robustness of the classifiers. In this case, the training data was not included in the prediction, and hence we gain an unbiased estimate of it's performance. The inference time for the ML models running the prediction is shown in Fig. 6.1.

Table 6.1: Inference time for the ML models running the prediction.

ML model	Inference time
Fine KNN	1305 sec
Weighted KNN	2213 sec
Single-layer NN	183 sec
Tri-layered NN	137 sec

The inference time was significantly longer for the KNN models than for the NN models. The tri-layered NN model used on average 137 sec and the single-layer NN used on average 183 sec, while the fine KNN model used 1305 sec on average and the weighted KNN model used 2213 sec on average predicting the data. In the classified image, the white color represents the mask covering land areas extended an additional 500-1000 m into the ocean in order to avoid adjacency. Fig.6.8 shows the label colors. Class 1 is black, class 2 is yellow/golden, class 3 is blue, class 4 is teal/light blue, and class 6 is red.

(a) Black color label.

(b) Yellow/golden color label.

(c) Blue color label.

(d) Teal/light blue color label.

(e) Red color label.

Figure 6.8: Label colors.

6.4.1 Visual representation of the trained classifiers

This section presents the results using the models on tile 2, which was included in the training data and can not be considered as an independent dataset. Figure 4.3 shows that tile 2 has some complex areas due to weather conditions and shallow waters. In the south-west corner we see a large low wind area with very little sunlight reflection, and in the north-west corner we see a complex shallow area with a mixture of different colors. The prediction of tile 2 using all models are shown in Fig. 6.9. In tile 2, three different oil slicks have been identified. Enlarged areas of the prediction done on the oil slicks are shown from Fig. 6.10 to Fig. 6.12.

Prediction of tile 2

Figure 6.9 shows the result after using the ML models on tile 2. Here Fig. 6.9 clearly shows the blue class 3, ocean water and the teal class 4, turbid water, together with the white land mask. This indicates a good overall performance for these classes. In the top left corner of Fig. 6.9a and Fig. 6.9b we see a large red cluster of class 6 vessels, a black cluster of class 1 thick oil and a golden cluster of class 2 thin oil. When observing the same area in Fig.4.7, we see a mix of shallow water and calm wind, which means that a large part of this area have been misclassified by the KNN models. The NN models, Fig. 6.9c and Fig. 6.9d seem to have managed better mostly classifying this area as class 4, turbid water. However, the cluster classified as thin oil is still present. In the bottom left corner we have another large complex ocean area, seen in Fig. 4.8, with calm wind. In this area all ML models have classified lots of class 2, thin oil. The NN had the poorest performance in this area. The single-layer NN, Fig. 6.9c classified much more thin oil than the KNN classifiers. The tri-layer NN classified this area as a cluster of mostly class 1, thick oil, but also class 2 thin

oil and class 4 turbid water. Figure 4.8 indicates that this area is mostly ocean water.

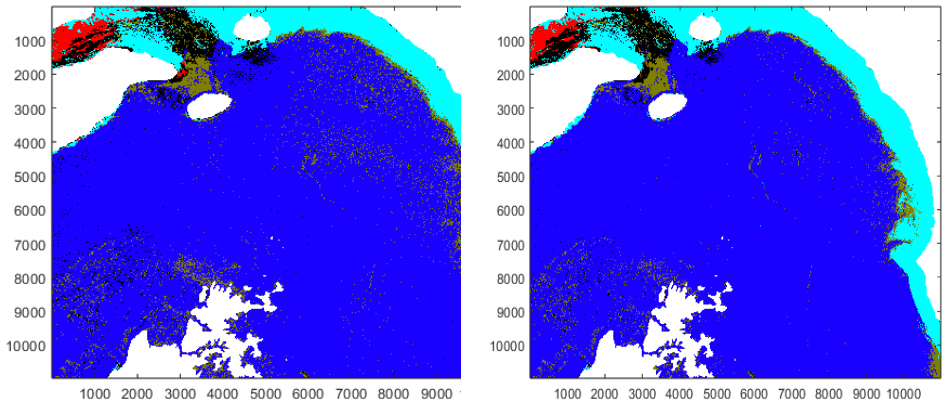
Along the border between the blue and the teal colors on the right side of the images, we observe clusters of yellow color. Figure 6.9 shows more clusters of yellow on the KNN models than on the NN models. Comparing to the true color composite of tile 2, Fig. 4.3, we see that this also is a misclassification. We also see yellow and black color spread inside the blue areas. The identified oil slicks in tile 2 are difficult to spot in Fig. 6.9, when not enlarged. Further we will look at the prediction of the oil slicks more closely.

Prediction of oil slick 1

Figure 6.10 shows an enlarged area of the prediction on the first oil slick in tile 2. Fig. 4.4 in Sec. 4.3.1 shows the equivalent true color composite image. All models managed to classify the thick oil with satisfying results. The NN models, Fig. 6.10c and Fig. 6.10d have classified thin oil differently than the KNN models, Fig. 6.10a and Fig. 6.10b. We see a much wider slick of thin oil with a discrete edge on both sides of the thick oil, for the KNN models. The discrete edge is likely due to that this data has been used for training the ML models. We therefore have an example of the affects including training data in predictions. However, this does not seem to have affected the NN models, where we can't observe any discrete edge similar to the KNN models.

Prediction of oil slick 2

Figure 6.11 shows enlarged areas of the prediction on the second oil slick in tile 2. As we can see from Fig. 4.6 in Sec. 4.3.1, the equivalent true color composite image, this oil slick is dominated by thin oil, with a narrow line of thicker oil fronting the ocean towards the north-east. In Fig. 6.11d we see that the tri-layer NN model managed to classify the oil slick and the surroundings very well, with clear clusters of thick oil, thin oil, turbid water, and ocean water. The single-layer NN model, Fig. 6.11c managed similar to the tri-layer NN model, but we observe a larger amount of data classified as turbid water mixed with thin oil. The KNN models, Fig. 6.11a and Fig. 6.11b, also managed to classify the thicker oil and the thin oil, but we see that a large portion of the ocean water and turbid water has been misclassified as thin oil. Also, in Fig. 6.11a, some data has been misclassified as thick oil instead of thin oil or turbid water.



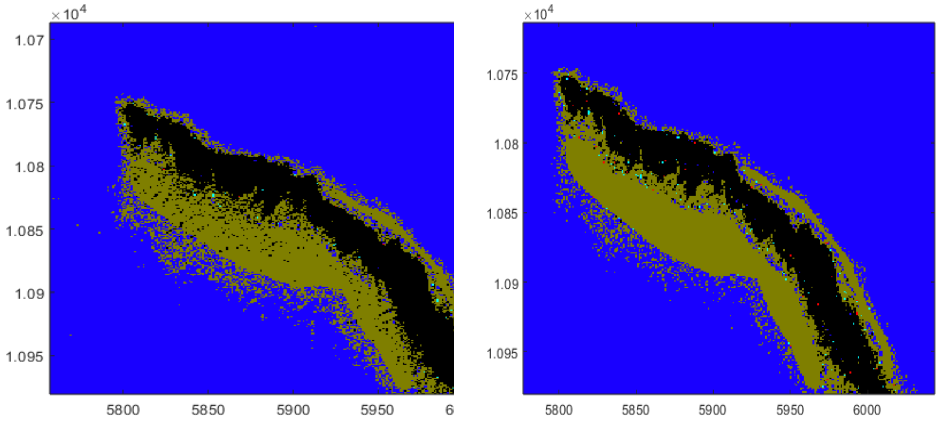
(a) Prediction of tile 2 using the Fine KNN model.

(b) Prediction of tile 2 using the weighted KNN model.

(c) Prediction of tile 2 using the single-layer NN model.

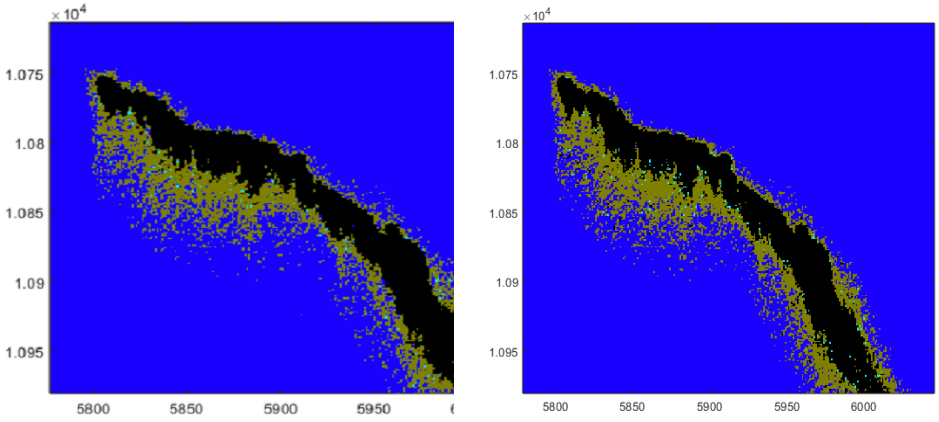
(d) Prediction of tile 2 using the tri-layer NN model.

Figure 6.9: Prediction of tile 2 using the ML models.



(a) Enlarged area of the prediction on oil slick 1 using the fine KNN model.

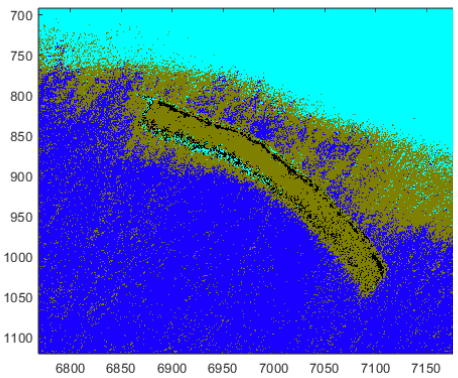
(b) Enlarged area of the prediction on oil slick 1 using the weighted KNN model.



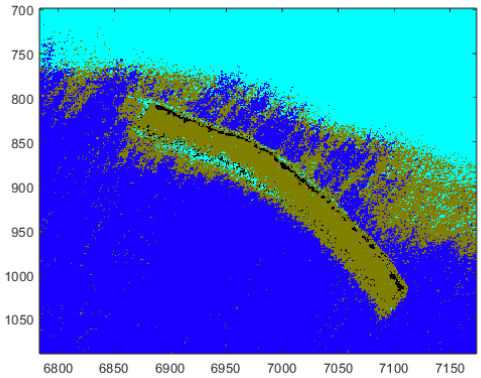
(c) Enlarged area of the prediction on oil slick 1 using the single-layer NN model.

(d) Enlarged area of the prediction on oil slick 1 using the tri-layer NN model.

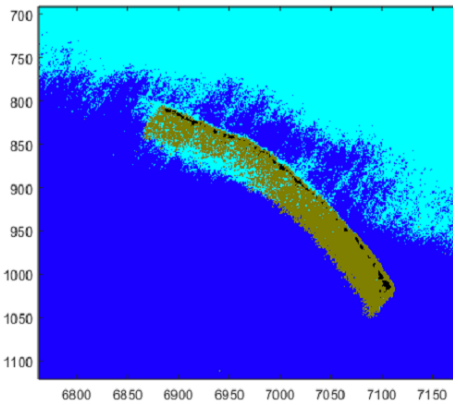
Figure 6.10: Enlarged areas of the prediction on the first oil slick in tile 2 using the trained ML models.



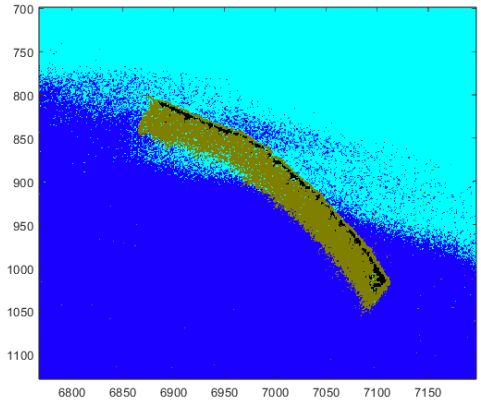
(a) Enlarged area of the prediction on oil slick 2 using the fine KNN model.



(b) Enlarged area of the prediction on oil slick 2 using the weighted KNN model.



(c) Enlarged area of the prediction on oil slick 2 using the single-layer NN model.



(d) Enlarged area of the prediction on oil slick 2 using the tri-layer NN model.

Figure 6.11: Enlarged areas of the prediction on the second oil slick in tile 2 using the trained ML models.

Prediction of oil slick 3

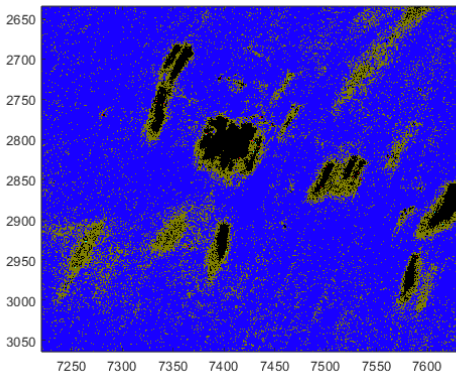
Figure 6.12 shows enlarged areas of the prediction on the third area with oil slicks in tile 2. Fig. 4.6 in Sec. 4.3.1 shows the equivalent true color composite image. These images shows multiple oil slicks compared to Fig. 6.10 and Fig. 6.11. The prediction done by the ML models are alike; all have performed well and classified the thick oil in discrete clusters. However, the NN models, Fig. 6.12c and Fig. 6.12d, and the weighted KNN model, Fig 6.12b, have also managed to classify the thin oil well. In the fine KNN model, Fig. 6.12a, lots of ocean water has been classified as thin oil spread around the image, and also the thin oil areas are "infected" with small appearances of thick oil.

6.4.2 Robustness evaluation

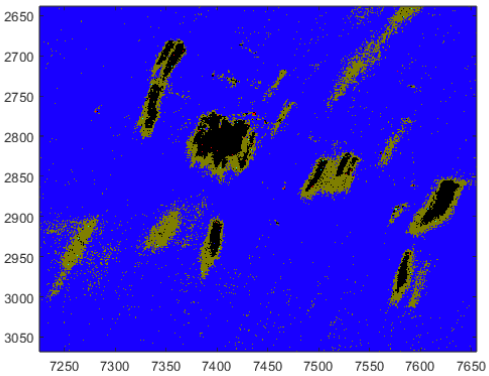
This section presents the results when applying the models on tile 11, which was not included in the training data. Hence the data is independent and suitable for robustness evaluation. As we can observe in the equivalent true color composite image, Fig.4.9, tile 11 has some complex areas due to clouds and shallow waters. In addition, the image appears darker in the ocean water areas and brighter in the turbid water areas, compared to tile 2 (Fig. 4.3). The results of the prediction on tile 11 using all models are shown in Fig. 6.13. In tile 11, one oil slick have been identified. Enlarged areas containing the oil slick are shown in Fig. 6.14.

Prediction of unlabeled data

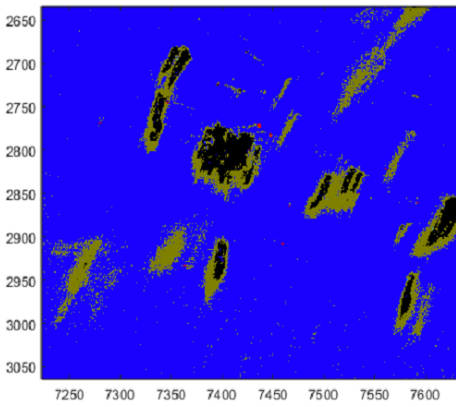
Figure 6.13 shows the result after using the ML models on tile 11. We see clearly the white landmask, the blue ocean water and the teal turbid water in all figures. This indicates a good overall performance for these classes. The prediction done by the KNN models, Fig. 6.13a and Fig. 6.13b show large clusters of red colors and smaller clusters of black color in the north-west corner and along the white landmask in the east, especially in the north-east. This is obviously a misclassification comparing to the true color composite in Fig. 4.9, which shows bright shallow water areas. This indicates a poor performance for thick oil and vessels for these models. The prediction done by the NN models, Fig. 6.13c and Fig. 6.13d do not show large red clusters. The single-layer NN, Fig. 6.13c shows large clusters of black color in the north-west and the north-east side of the image, which is also a misclassification comparing to the true color composite in Fig. 4.9. The prediction by tri-layer NN, Fig. 6.13d shows an overall best performance, with only small black clusters in the shallow water areas. The prediction by the NN models also shows less "noise" in the ocean waters, meaning less ocean water data has misclassified as thick oil spread



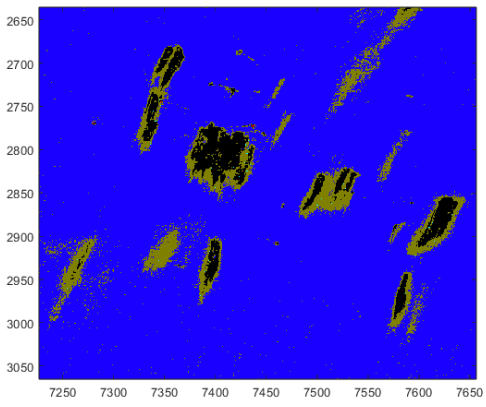
(a) Enlarged area of the prediction on oil slick 3 using the fine KNN model.



(b) Enlarged area of the prediction on oil slick 3 using the weighted KNN model.

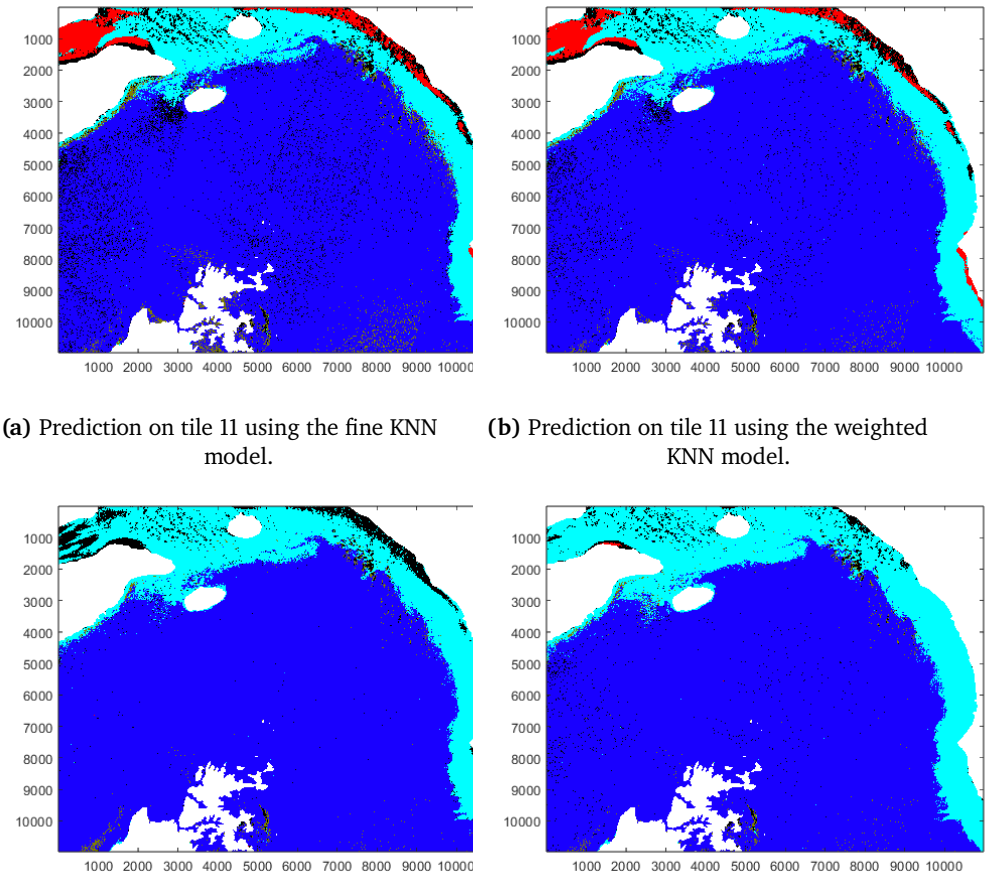


(c) Enlarged area of the prediction on oil slick 3 using the single-layer NN model.



(d) Enlarged area of the prediction on oil slick 3 using the tri-layer NN model.

Figure 6.12: Enlarged areas of the prediction on the third oil slick in tile 2 using the trained ML models.



(a) Prediction on tile 11 using the fine KNN model.

(b) Prediction on tile 11 using the weighted KNN model.

(c) Prediction on tile 11 using the single-layer NN model.

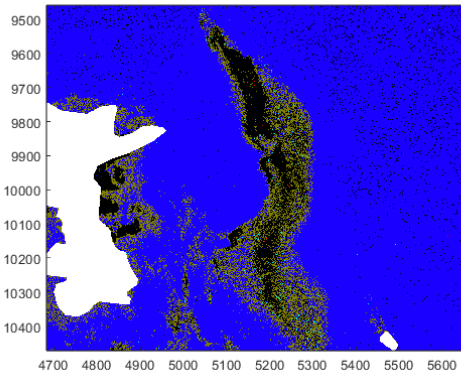
(d) Prediction on tile 11 using the tri-layer NN model.

Figure 6.13: Prediction of unlabeled data - tile 11.

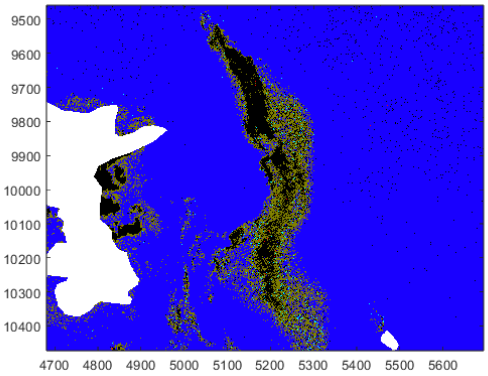
around the ocean. The identified oil slick is difficult to observe in Fig. 6.13, and will next be discussed using Fig 6.14.

Prediction of unlabeled oil slick

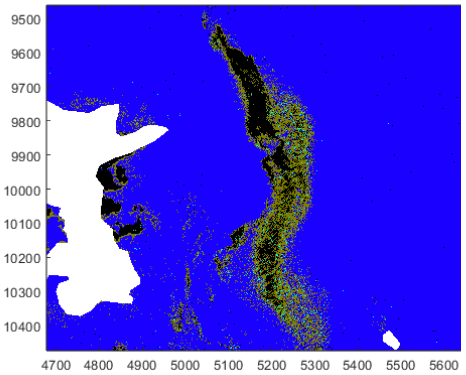
Figure 6.14 shows enlarged areas of the prediction on an oil slick identified in tile 11. Fig. 4.10 in Sec. 4.3.1 shows the equivalent true color composite image. All models successfully manage to identify the oil spill, detecting the thick oil and thin oil well. All models have also misclassified the clouds seen next to the land area, as thick oil and thin oil. This shows the need for a mask covering the clouds, to prevent clouds being classified as oil using these ML models. We can also see that a small amount of data inside the thin oil has been classified



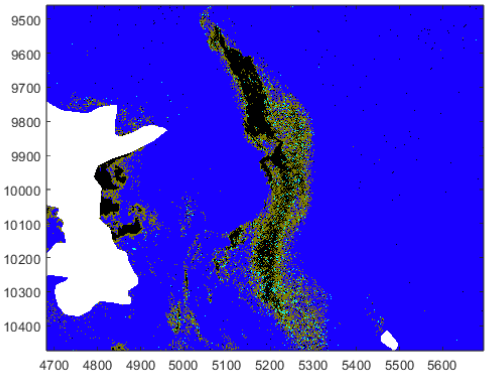
(a) Enlarged area of the prediction on an oil slick in tile 11 with a Fine KNN.



(b) Enlarged area of the prediction on an oil slick in tile 11 with a Weighted KNN.



(c) Enlarged area of the prediction on an oil slick in tile 11 with a single-layer NN.



(d) Enlarged area of the prediction on an oil slick in tile 11 with a tri-layer NN.

Figure 6.14: Prediction of unlabeled data - oil slick.

as turbid water.

7

Discussion

This project experimented with different ML models on multispectral satellite data with the goal of identifying thick oil on the ocean surface. We also included classes for thin oil, ocean water, turbid water and vessels. First we created the dataset. Second we trained a variety of different ML models. Third we tuned the parameters of the ML models showing best results. Finally we trained and applied four ML models which were found to have the highest accuracy in classifying thick oil. Two of the ML models were KNN models and two were NN models. The models were trained and validated with data from areas in the Persian Gulf, see Fig.4.1. The trained models were then used to classify two images from the same area, to visualize the performance of the classifiers.

All classifiers showed good performances based on the OA, varying from 91.9 % to 94.1 %, where the weighted KNN performed best, and the tri-layered NN had the lowest classification accuracy, see Fig. 6.3. It is important to note that the distribution of observations from each class were not equal, due to that limited amount of observation of thick oil, thin oil and vessels, compared to ocean water and turbid water. Especially the number of ocean water observations where high, with around 60 % of the total observations . See Sec.4.3.3 for the distribution of labeled data. Due to a higher amount of observations from the water classes, a high TPR for these classes (class 3 and 4) made a significant impact on the OA.

The classifiers showed varying performance classifying thick oil (class 1). The

lowest TPR was 64.1 % for the single-layer NN and the highest TPR was 84.2 % for the fine KNN. The highest accuracy (TPR) for thick oil, 84.2 %, was achieved using the fine KNN classifier with two nearest neighbors. This is in good correspondance with the results reported in [9], where a two-class classification was studied by using KNNs. The classes were oil and look-alikes, the three KNNs had 1, 2 and 3 nearest neighbors, and the achieved accuracy was 71 %, 72 % and 76 %, respectively. Another study with six classes [14], showed an accuracy of 88 % for oil using KNN. The results in the aforementioned studies ([9, 14]) show good correspondence with the fine KNN used in this thesis.

For the NN models, the highest TPR for thick oil achieved was 64.5 % with the Tri-layered NN. A smilar study [9] achieved an accuracy of 79 % using a multilayer perceptron (MLP) with 1 hidden layer using 6 nodes. This shows that our NN models have a poorer performance classifying oil compared to similar studies. A reason for this can be the simplicity of our NN models. In our most complex model we used a three layered feed forward NN with 12 nodes. This is a simple model, but yet the study [9] used only one layer with six nodes and got a better result. A possible error is over-fitting, having high variance in the training data. When using 10 % of the dataset for training we saw a TPR of 86 % for the NN classifiers, and when working with 30 % of the dataset in phase 2, we saw a TPR of over 90 % for both NN classifiers. Then we implemented the whole dataset, and saw a decrease in TPR from over 90 % to 64 %. If the ocean color reflectance was different in the last six tiles of the dataset due to for instance different sunlight conditions, this may have caused the decrease in performance.

In this work class 2, thin oil, was defined according to code 1 (Sheen) and code 2 (rainbow) from the Bonn agreement [7]. The classifiers performed acceptable for this class, with a TPR varying from 71.9 % for the Onelayered NN model to 79.5 % for the Weighted KNN model. Using the Bonn agreement to label this class was challenging due to the gradual transition between ocean water (class 3) and sheen oil (code 1), and the gradual transition between sheen/rainbow oil and thick oil (class 1). Also look-alikes features can easily be confused with thin sheen oil slicks, due to their similar characteristics in the visible true color composite images. In this work, to prevent labeling false oil slicks, we only included thin oil in connection to thick oil with visible metallic or true colored signatures.

For class 3 (ocean water) the classifiers performed very well, where the lowest TPR was 94.4 % for Fine KNN and the highest TPR was 98.4 % for weighted KNN. Class 4 (turbid water) showed similar results, with TPR varying from 96.7 % for tri-layered NN to 97.3 % for weighted KNN. In this work we mainly included observation from waters with favorable wind conditions and sun angles, while also including observations from near range and far range. This means that

we only included a limited number of observation from complex water areas in the ocean and turbid water classes, such as water areas with unfavorable wind conditions or oil look-alikes. In this way we minimized the variation in reflectance values for the ocean and turbid water class. In the predicted images we also see that the classifiers mainly have classified the ocean waters and the turbulent water correctly, but high rate of misclassification can be observed in the complex water areas, classifying water areas as oil or vessels. See Fig.6.9 and Fig. 6.13.

All classifiers performed poorly classifying vessels (class 6), with TPRs varying from 4.8 % to 26.9 %. In this case the weighted KNN performed best and tri-layered NN had the lowest accuracy as well. This was expected, since the reflectance spectra of vessels show large variations, see Fig. 4.15a. Therefore, classifying vessels is challenging, which is especially valid for the KNN classifier. KNN uses neighborhood data values in the classification process, hence large variance impacts the accuracy of the classification. In addition, this class had few observations compared to the total amount of observations. However, this work mainly focused on classifying oil. Labeling data corresponding to vessels was not the focus, it was only an additional exercise. Classifying vessels could be improved by including a larger number of observations and tuning the classifiers to handle large variance within the class.

Two similar studies, one for multispectral data [9] and one for SAR data [8] uses only two classes when classifying oil with ML; oil and look-alikes. This may be a good approach when the goal is to differentiate between oil and look-alikes. However, the goal of this work is not just to separate oil from look-alikes, but to identify oil in oceanic and turbid waters. A ML model could potentially be of aid for an operator monitoring marine areas for oil slicks or as a fully automatic service. This means that the ML model needs to classify the whole marine area in the remotely sensed image. When an oil monitoring service uses a ML model as a tool aiding an operator, in terms of error it is crucial that the classifier rather identify false positives instead of missing true positives. This is due to that an operator can make her own assessment on detected oil spills found by the ML model. Masks for land and clouds should be used to make it simper for a ML model. Ideally a ML classifier should be able to separate oil from all other classes present at the ocean surface.

A limitation of this study is that the variation of sun and view angles in the Sentinel-2 data were not considered. This can possibly create a larger variation of reflectance values in the dataset. Another limitation is that the robustness evaluation was only performed on one unlabeled Sentinel-2 image. This means that the robustness evaluation only was performed at a specific sun angle. Evaluation the robustness on different sun angles would give a broader view of the performance of the ML models. We see two ways of doing a similar study. The

first would be to only pick multi-spectral data with a certain sun and view angle for your training and test data. The second is to first do a background study on the historical sun angles in Sentinel-2 data, before including Sentinel-2 data with a variety of sun angles based on the background study. A final limitation is that the training data included limited numbers of data from complex areas (weather, look-alikes), which may impact the performance using the ML models on images where complex areas are present.

8

Conclusion

This project showed that it is possible to identify and classify oil slicks in the Persian Gulf using multispectral remote sensing data and ML methods. Several ML methods were evaluated: ANN, KNN, Decition Trees, Naive Bayes and discriminant classifiers. Two ML methods resulted in the highest accuracy and were further evaluated in dept. These two methods were the ANN and KNN. The performance of a fine KNN model and a weighted KNN model on the test data was satisfactory (Fig. 6.3). When applied to an independent dataset the KNNs were able to identify areas with thick oil (Fig. 6.10, Fig. 6.11 and Fig. 6.12). The performance of a single-layer NN model and a tri-layered NN model was poor in testing, but they were still able to identify thick oil when applied to a non independent dataset (Fig. 6.10, Fig. 6.11 and Fig. 6.12). A robustness evaluation was performed on unlabeled independent data. The NN models and the KNN models were both able to identify and classify the thick oil with good precision when compared to the equivalent true color composite image (Fig. 6.14). All models showed poorer performance classifying in areas with complex reflectance due to calm wind, clouds or shallow waters in combination with low wind.

The ANN and KNN models used in this work show great potential to aid operators in an oil spill monitoring service, but they are not good enough to work as ML models integrated in a fully automatic service, due to the large amount of false positives. Even tough the KNN models used significantly less time in the training process, the NN models used significantly less time computing the predictions, and are therefor a better option for an oil spill monitoring service

with near-real time (NRT) requirements. However, more tuning and robustness testing of the NN models are likely needed to find the most optimal NN model for an oil spill monitoring service.

This project provided valuable perspectives in both the possibilities and limitation in classifying oil slicks using ML. While NN and KNN have been used for oil spill classification in multi-spectral imagery, the ML models have to the author's knowledge not been used to classify oil slicks with multi-spectral remote sensed data using a combination of visible, NIR and SWIR bands in the Persian Gulf. The result indicates that the studied ML methods are promising for this purpose, and potentially be subject for up-scaling and implemented in a oil spill monitoring service.

Future work

Reliable ML models can be an important asset for monitoring and detection of oil spills impacting the environment and society. Further work with the ML models created in this work is needed to create satisfactory ML models for oil spill monitoring services. To achieve this, a large scale robustness evaluation on established ML models, testing the images on a large number of Sentinel-2 scenes is necessary. It might also be necessary to tune the ML mode with a greater variety of configurations than done in this work, to evaluate for better performance.

Bibliography

[1] J. Beyer, H. C. Trannum, T. Bakke, P. V. Hodson, and T. K. Collier. Environmental effects of the deepwater horizon oil spill: A review. *Marine Pollution Bulletin*, 110(1):28–51, 2016.

[2] L. Camus and M. G. D. Smit. Environmental effects of arctic oil spills and spill response technologies, introduction to a 5 year joint industry effort. *Marine Environmental Research*, 144:250–254, 2019.

[3] M. Barron, D. Vivian, R. Heintz, and U. H. Yim. Long-term ecological impacts from oil spills: Comparison of exxon valdez, hebei spirit, and deepwater horizon. *Environmental Science Technology*, XXXX, 04 2020.

[4] I. Leifer, W. J. Lehr, D. Simecek-Beatty, E. Bradley, R. Clark, P. Dennison, Y. Hu, S. Matheson, C. E. Jones, B. Holt, M. Reif, D. A. Roberts, J. Svejkovsky, G. Swayze, and J. Wozencraft. State of the art satellite and airborne marine oil spill remote sensing: Application to the bp deepwater horizon oil spill. *Remote Sensing of Environment*, 124:185–209, 2012.

[5] J. R. Jensen, E. W. Ramsey, J. M. Holmes, J. E. Michel, B. Savitsky, and B. A. Davis. Environmental sensitivity index (esi) mapping for oil spills using remote sensing and geographic information system technology. *International Journal of Geographical Information Systems*, 4(2):181–201, 1990.

[6] J. B. Campbell and R. H. Wynne. *Introduction to remote sensing*. The Guilford Press, 2011.

[7] Bonn agreement oil appearance code. `https://www.bonnagreement.org/site/assets/files/1081/special_on_volume_calculation_20160607.docx`.

[8] S. Singha, T. Bellerby, and O. Trieschmann. Satellite oil spill detection using artificial neural networks. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6:2355–2363, 12 2013.

[9] L. Corucci, F. Nardelli, and M. Cococcioni. Oil spill classification from multi-spectral satellite images: Exploring different machine learning techniques. volume 7825, page 782509, 10 2010.

[10] Z. Zhang, L. Yan, X. Jiang, J. Ding, F. Zhang, K. Jiang, and K. Shang. Exploring the potential of optical polarization remote sensing for oil spill detection: A case study of deepwater horizon. *Remote Sensing*, 14(10), 2022.

[11] S. T. Yekeen and A. L. Balogun. Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment. *Remote Sensing*, 12(20):3416, 2020.

[12] Eo sar imagery prices limit opportunity. `https://www.nsr.com/eo-sar-imagery-prices-limit-opportunity/`.

[13] J. Hedley, A. Harborne, and P. Mumby. Technical note: Simple and robust removal of sun glint for mapping shallow-water benthos. *International Journal of Remote Sensing - INT J REMOTE SENS*, 26:2107–2112, 05 2005.

[14] R. Trujillo-Acatitla, J. Tuxpan-Vargas, and C. Ovando-Vázquez. Oil spills: Detection and concentration estimation in satellite imagery, a machine learning approach. *Marine Pollution Bulletin*, 184:114132, 2022.

[15] G. E. Stillman. 21 - optoelectronics. In Wendy M. Middleton and Mac E. Van Valkenburg, editors, *Reference Data for Engineers (Ninth Edition)*, pages 21–1–21–31. Newnes, Woburn, ninth edition edition, 2002.

[16] Electromagnetic spectrum figure. `https://www.japanistry.com/electromagnetic-spectrum`.

[17] Em-wave figure. `https://www.oceanopticsbook.info/view/theory-electromagnetism/level-2/plane-wave-solutions`.

[18] C. Elachi and J. Van Zyl. *Introduction to the physics and techniques of remote sensing*. John Wiley and Sons, 2006.

[19] Ocean optics web book. `https://www.oceanopticsbook.info/`.

[20] Y. Lu, X. Li, Q. Tian, G. Zheng, S. Sun, Y. Liu, and Q. Yang. Progress in marine oil spill optical remote sensing: Detected targets, spectral response characteristics, and theories. *Marine Geodesy*, 36(3):334–346, 2013.

[21] Z. Otremba. Coefficient of light reflection foreseen for chosen oils. *Journal of KONES. Powertrain and Transport*, 21:375–379, 01 2014.

[22] W.G. Rees. *Physical Principles of Remote Sensing*. Cambridge University Press, 2012.

[23] M. Seelye. *An introduction to ocean remote sensing*. Cambridge University Press, 2014.

[24] Ksat, oil spill detection service. https://www.ksat.no/earth-observation/environmental-monitoring/oil-spill-detection-service/.

[25] V. C. Coffey. Multispectral imaging moves into the mainstream. *Opt. Photon. News*, 23(4):18–24, Apr 2012.

[26] H. Zhai, H. Zhang, L. Zhang, and P. Li. Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:235–253, 2018.

[27] A. Himes-Cornell, L. Pendleton, and P. Atiyah. Valuing ecosystem services from blue forests: A systematic review of the valuation of salt marshes, sea grass beds and mangrove forests. *Ecosystem Services*, 30:36–48, 2018.

[28] A. Dupuis and F. Ucan-Marin. *A literature review on the aquatic toxicology of petroleum oil: An overview of oil properties and effects to aquatic biota*. Canadian Science Advisory Secretariat Vancouver, BC, USA, 2015.

[29] Z. Otremba and J. Piskozub. Monte carlo radiative transfer simulation to analyze the spectral index for remote detection of oil dispersed in the southern baltic sea seawater column: The role of water surface state. *Remote Sensing*, 14:247, 01 2022.

[30] Sentinel-2 satellite, esa. https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

[31] Landsat-8 satellite, usgs. https://www.usgs.gov/landsat-missions/landsat-8.

[32] Landsat-9 satellite, usgs. https://www.usgs.gov/landsat-missions/landsat-9.

[33] Prisma satellite, asi. https://directory.eoportal.org/web/eoportal/satellite-missions/p/prisma-hyperspectral.

[34] Aqua satellite, nasa. https://aqua.nasa.gov/.

[35] Terra satellite, nasa. https://terra.nasa.gov/.

[36] Msi, esa. https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument.

[37] National Research Council (US) Committee on Oil in the Sea: Inputs Fates and Effects. *Oil in the Sea III: Inputs, Fates, and Effects*. National Academies Press (US), 2003.

[38] R. Al-Ruzouq, M. A. Gibril, A. Shanableh, A. Kais, O. Hamed, and S. A. Mansoori. Sensors, features, and machine learning for oil spill detection and monitoring: A review. *Remote Sensing*, 12:3338, 10 2020.

[39] M. Cunliffe and J. C. Murrell. The sea-surface microlayer is a gelatinous biofilm. *The ISME Journal*, 3, 2009.

[40] S. Shalew-Shwartz and S. Ben-David. *Understanding machine learning*. Cambridge university press, 2014.

[41] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, Elsevier, 2009.

[42] J. J. Salazar, L. Garland, J. Ochoa, and M. J. Pyrcz. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering*, 209:109885, 2022.

[43] E. Alpaydin. *Introduction to machine learning*. MIT Press, 2014.

[44] K. Taunk, S. De, S. Verma, and A. Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 1255–1260, 2019.

[45] S. B. Imandoust and M. Bolandraftar. Application of k-nearest neighbor (knn) approach for predicting economic events theoretical background. *Int J Eng Res Appl*, 3:605–610, 01 2013.

[46] S. Theodoridis. Chapter 18 - neural networks and deep learning. In Sergios Theodoridis, editor, *Machine Learning (Second Edition)*, pages 901–1038. Academic Press, second edition edition, 2020.

[47] Node figure. https://www.v7labs.com/blog/neural-networks-activation-functions.

[48] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11):e00938, 2018.

[49] Sentinel-hub eo browser. https://apps.sentinel-hub.com/eo-browser.

[50] Esa copernicus open access hub. https://scihub.copernicus.eu.

[51] Qgis software. https://qgis.org/en/site.

[52] S. Baboo and R. Devi. An analysis of different resampling methods in coimbatore, district. *Global Journal of Computer Science and Technology*, 10, 01 2010.

[53] Matlab software. https://se.mathworks.com/products/matlab.html.

[54] Esa snap software. https://www.eoportal.org/other-space-activities/snap-sentinel-application-platform#snap-sentinel-application-platform-toolbox.

[55] Matlab: Classification learner application. https://se.mathworks.com/help/stats/classificationlearner-app.html.