Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Guro Drange Veglo

Uncertainty in Hyperspectral Remote Sensing: An Evaluation in Forest Monitoring

Master's thesis in Cybernetics and Robotics
Supervisor: Joseph L. Garrett
June 2023

NTNU
Norwegian University of
Science and Technology

Guro Drange Veglo

Uncertainty in Hyperspectral Remote Sensing: An Evaluation in Forest Monitoring

Master's thesis in Cybernetics and Robotics
Supervisor: Joseph L. Garrett
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

0.1 Preface

The present thesis marks the culmination of my five-year Master's program in Cybernetics and Robotics at the Norwegian University of Science and Technology (NTNU). This work is a constituent of the FAUBAI initiative, backed by the European Space Agency (ESA), as a collaborative venture involving the NTNU, the University of Oslo (UiO), and S&T Norway. The objectives of this project seamlessly align with the ongoing investigations undertaken by the SmallSat team at NTNU, specifically centered on the autonomous monitoring of forests via hyperspectral Earth Observation (EO) technology.

The overarching goal of this thesis is to augment the ESA In-Orbit Demonstration (IOD) Proposal [S&T (2021)], specifically aiming toward waypoint 319 and tasks encompassing the independent verification of the Silvisense network and the evaluation of uncertainty propagation. Embarking on this thesis journey has opened my eyes to the intriguing world of hyperspectral remote sensing and Artificial Intelligence (AI). It has been exciting to navigate the intricacies and challenges of this field of study.

A big thank you goes to Joe Garrett, the architect behind the models that served as the cornerstone for my analysis. Not only did he provide the foundational tools for testing batch performance, but his support, mentorship, and advice were instrumental throughout my thesis. Joe has pushed me in challenging directions, from which I have learned so much. Additionally, I would like to thank Norwegian Institute of Bioeconomy Research (NIBIO) for providing the ground truth data and answering my questions.

I must also extend heartfelt thanks to my friends who made office breaks something to look forward to. Without your companionship and shared laughter, I might have lost my sanity. Lastly, thanks to my family and boyfriend for helping with the read-troughs. Your support and encouragement have inspired me and given me "the freedom to choose".

Trondheim, June 4, 2023

Guro Drange Veglo

Summary

Hyperspectral Imager (HSI) offer abundant data relevant to assessing Earth's surface condition, affirming their widespread application across diverse models. However, model-based uncertainties can lead to erroneous conclusions with significant real-world implications if not correctly quantified and addressed. Consequently, our conclusions must be buttressed by rigorous evidence to ensure their utility and reproducibility.

This thesis scrutinizes uncertainties affecting HSI data, specifically focusing on Support Vector Machines (SVM) and Convolutional Neural Network (CNN) within forest management. It investigates uncertainties arising from data, parameters, and model structure, subsequently enhancing our comprehension of factors causing misclassification errors.

The investigation of NIBIO's ground truth and PRISMA's HSI revealed data uncertainty and variations in the geographical distribution. The introduction of artifacts resulted in noticeable performance degradation, with *smile* distortion significantly degrading prediction accuracy, leading the classifiers to misclassify pixels in the spatial edges. When utilizing the One Dimensional (1D) CNN, *smile* distortions led to a degradation in tree species accuracy of at least 76%, within the first pixel of distortion.

The study revealed that augmenting the number of training images does not uniformly enhance prediction accuracy. Additionally, model convergence across batches revealed that the SVM model might offer greater reliability, despite the CNN exhibiting higher prediction accuracy. Intriguingly, the CNN was observed to be more sensitive to artifacts, displaying a significantly higher level of degradation under all distortions than the SVM. At the same time, the *Deciduous* class was shown to be more robust towards noise than *Spruce* and *Pine*.

The findings of this study underscore the presence of considerable model uncertainty and its impact on SVM and CNN models. They illuminate the path toward refining model performance strategies and contribute to developing reliable decision-making processes in HSI-based applications, ultimately ensuring the robust application of these models in real-world scenarios.

Sammendrag

Hyperspektral data gjør det mulig å overvåke jordens tilstand, og åpner dermed for bruksområder innen maskinlæringsmodeller. Likevel kan usikkerhet i data og modeller føre til feilaktige konklusjoner og fatale konsekvenser dersom de ikke blir fullstendig gjort rede for. Dermed er det avgjørende at studier kan vise til analyser av usikkerhet når resultater presenteres.

Denne oppgaven analyserer usikkerheter som påvirker hyperspektral data innen skogforvaltning, med fokus på Support Vector Machines (SVM) og Convolutional Neural Networks (CNN). Usikkerheter kan oppstå fra data, parametere, og modellstrukturen, og sammen få store konsekvenser. I denne analysen etterforskes hver kilde til usikkerhet separat for å forstå faktorene som forårsaker feilklassifiseringer i modeller.

Etterforskning av NIBIOs grunnsannhet og PRISMA's hyperspektrale data avslørte usikkerhet i dataen. En sammenligning av SVM og CNN viste at støy påvirker modellene likt i de fleste tilfeller. Samtidig resulterte introduksjonen av forvrengninger og uklarheter i en merkbar nedgang i prediksjonsnøyaktighet. Dette var spesielt merkverdig for *smile*, der prediksjonsnøyaktigheten for CNN ble degradert 76% innen en pixel forvrengning. Resultatene viste også at støy, forvrengninger og uklarhet påvirket CNN-modellene mer enn SVM-modellene.

I tillegg avslørte konvergens av modeller på tvers av batcher at SVM-modellen er mer pålitelig, til tross for at CNN-modellen har høyere prediksjonsnøyaktighet. Studien avslørte også at større datasett ikke nødvendigvis forbedrer prediksjonsnøyaktigheten.

Disse funnene understreker behovet for å forstå og kvantifisere usikkerheter innebygd i hyperspektral data, samt deres innvirkning på prediksjoner fra SVM- og CNN-modeller. De understreker også nødvendigheten av å analysere usikkerheten for alle deler og stadier av modeller for å kunne utvikle pålitelige og robuste modeller.

Contents

Abbreviations

1D One Dimensional ii

AI Artificial Intelligence i

ANN Artificial Neural Network 13

ASI Italian Space Agency 32

CNN Convolutional Neural Network ii

CRN National Research Council 32

dB decibel 25

DL Deep Learning 12

EO Earth Observation i

ESA European Space Agency i

Faubai FPGA Acceleration for Push-Broom Satellite Imagery 30

FC Fully Connected 21

FRM Fiducial Reference Measurements 32

GSD Ground Sampling Distance 32

HSI Hyperspectral Imager ii

IOD In-Orbit Demonstration i

ML Machine Learning 1

MS-NFI Multi-Source National Forest Inventory 36

NFI National Forest Inventory 36

Glossary

Activation function In artificial neural networks, the activation function of a node defines the output of that node given an input or set of inputs. 24

Batch size Defines the number of samples processed before the model is updated. 24

Drop-out rate The practice of disregarding specific nodes in a layer at random during training. A dropout is a regularization approach that prevents overfitting by ensuring that no units are codependent with one another. 24

Epoch Refers to the one entire passing of training data through the algorithm. It determines the process of training the machine learning model. 24

F1 score F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. 72

Ground truth The information known to be real or true, provided by direct observation and measurement. 33

Learning rate Defines the adjustment in the weights of a network concerning the loss gradient descent. It determines how fast or slow it moves toward the optimal weights. 24

Read noise Introduced during the readout process of the detector, and it is due to the electronics that amplify the signal from the detector. 27

Spectral bands Spectral bands are parts of the electromagnetic spectrum of specific wavelengths. 24

Stride Number of pixels shifts. 20

Chapter 1

Introduction

Satellite imagery has become indispensable for monitoring and managing the Earth's natural resources. Hyperspectral Imager (HSI) data, in particular, has gained attention due to its ability to capture detailed spectral information of the Earth's surface, allowing for more accurate and precise analysis. In this context, Machine Learning (ML) models have shown promising results for image classification tasks [Yang et al. (2018)]. However, the complex nature of HSI data can be affected by various uncertainties.

These uncertainties pertain to the optical data itself, which can be detrimentally impacted by various factors, including noise, blur, and specific optical distortions, notably referred to as *smile* and *keystone* effects. Concurrently, further complications arise due to inconsistency within the ground truth data and in the parameters and structures of the models used. The sum of uncertainties contributes to the difficulty of accurately detecting specific features or phenomena within the hyperspectral data. Moreover, incomplete consideration of these uncertainties can lead to false conclusions with real-world impacts and potentially damaging consequences for decision-making processes. Despite this, uncertainty consideration is incomplete within and across scientific fields [Simmonds et al. (2022)].

Uncertainty within forest management and remote sensing causes inaccurate identification and categorization of land cover classes and types, such as areas with and without forest. This data uncertainty is especially noticeable because of the complexity of HSI imagery. A study conducted by Sexton et al. (2013) highlighted the difficulties in extrapolating data between varying spatial resolutions, particularly in global tree cover datasets. Meanwhile, Coops et al. (2007) reviewed several methods for detecting forest disturbances, emphasizing the significance of carefully selecting remote sensing data, classification algorithms, and spatial resolution to reduce the likelihood of errors in classification.

For these reasons, several notable misclassification errors using remote sensing applications have occurred. Breidenbach, Ellison, Petersson, Korhonen, Henttonen, Wallerman, Fridman,

Gobakken, Astrup & Næsset (2022) highlights how the misinterpretation of remotely sensed data led to the overestimation of forest areas. They argue that advancements in satellite-based mapping, while providing more detailed data, might contribute to misclassification if not carefully assessed. The higher-resolution imagery may detect small-scale changes in the land cover, which can be incorrectly identified as harvest areas when they may result from natural disturbances or regeneration processes. These examples highlight the importance of addressing misclassification errors and developing strategies to mitigate their impact.

There have been several calls for more consideration of uncertainty. An article by Pratihast et al. (2014) highlighted the possibility of incorporating ground-based validation and local knowledge of areas to reduce misclassification in remote sensing applications for forest management. By integrating diverse data sources, they demonstrated that the accuracy and reliability of land cover classification improved.

However, there exists no universal taxonomy that captures the fully coherent picture of scientific uncertainty [Volodina & Challenor (2021), Kreps & Kriner (2020), Seibold et al. (2019), Edeling et al. (2021)]. For this reason, Simmonds et al. (2022) proposed a source framework for uncertainty consideration. This framework divides model-related uncertainty into three primary sources: data, *parameter estimates*, and *model structure*. The data component is divided into two sub-sources: the *response*, which refers to the focal variable being explained (i.e., tree classification), and the *explanatory variables*, which encompass any variables used to explain or predict the response. Consequently, this approach allows for a comprehensive assessment of four distinct sources of uncertainty in tree classification. For the specific case of forest type classification, a breakdown of the sources of uncertainty is given in Table 1.1.

Uncertainty in Remote Sensing AI: Using Forests as a Lens

Apprehending uncertainty in remote sensing AI is an intricate endeavor that necessitates the examination of systems that facilitate reliable observations and the identification of latent variables. In this context, forests emerge as an optimal subject of study, primarily attributed to two pivotal factors.

Firstly, unlike other observable entities, such as algal blooms, trees are stationary and do not move. This static nature allows for a better approximation of ground truth, providing a stable baseline against which to measure the accuracy and reliability of remote sensing AI predictions. In a world where data is continuously shifting, the permanence of forests provides a unique and invaluable opportunity for developing, testing, and refining AI systems.

Secondly, the history of scientific forestry offers critical lessons about the perils of overlooking latent variables, such as soil conditions. Forests are not just a collection of trees for eco-

nomic gain but also a complex ecosystem that includes animals, insects, and other living organisms. Past technological advancements in forestry often neglected these factors due to their long emergence time, beyond a human lifespan, leading to a misconception of their value [Scott (1998)]. Nevertheless, these overlooked variables are fundamental to the functioning and health of the forests and, by extension, to the accuracy of our AI models.

Remote sensing AI models that analyze forest conditions, including algorithms like SVMs, and CNNs, can make predictions based on large datasets. However, these models can overlook latent variables or miss data due to the inaccessible or unobservable parts of the forest, leading to model uncertainty. This is a reminder of the history of scientific forestry, where the forest was viewed solely as a collection of trees for economic gain, neglecting the complex ecosystem's essential aspects, causing unforeseen problems and incomplete forest management plans [Scott (1998)].

As we navigate the challenges of understanding uncertainties in remote sensing AI, it becomes essential to acknowledge the limitations of our models. Using them in conjunction with other data and observations allows us to leverage their capabilities while compensating for their weaknesses fully. By studying forests, we can delve into the complexities of remote sensing AI, developing methods to account for uncertainties and latent variables, which pave the way for more accurate and reliable predictions. Thus, forests serve as an ideal paradigm for understanding uncertainty in remote sensing AI, offering invaluable insights that can enhance our models and predictions.

To illustrate this further, this thesis will investigate how ML models, specifically the SVM and CNN architectures, respond to uncertainties in the classification of forest types. The study employs the framework mentioned earlier (Table 1.1) to assess each source of uncertainty, as detailed in Table 1.2. The investigation entails thoroughly examining the ground truth data, evaluating the quality of the hyperspectral data, analyzing the impact of model parameters, and comparing the performance of different models. By undergoing this comprehensive process, the study aims to highlight the challenges in HSI-based forest management applications. Ultimately, the goal is to contribute to more accurate and reliable decision-making processes in forest management.

Source	Definition	How is it a source of uncertainty?
Response	*"The variable(s) of interest, the quantity(ies) we want to explain or predict. Could be something measured or something latent (unobserved) or simulated."* The response variable is the forest type that the SVM and CNN models are attempting to classify and predict.	The forest type could be subject to • misclassification, • measurement errors, or • natural variations (age, density, etc.) in the ground truth data. This uncertainty affects the model's accuracy and generalization ability when predicting forest types.
Explanatory variables	*"Any variable that explains or predicts the response. Could be something measured, latent (unobserved), simulated, or theoretical."* The hyperspectral image data from the PRISMA satellite is used to predict or explain the forest-type classifications.	Hyperspectral data can have • measurement noise, • artifacts, • sensor limitations. These uncertainties can affect the model's ability to identify and classify forest types accurately.
Parameter estimates	*"Values given to unknown parameters in the model either through estimation, optimization, or chosen."* Values are assigned to unknown parameters in the SVM and CNN models through hyperparameters, optimization during training, and selection.	Uncertainty in parameter estimates arises from • the process of model training, and • the choice of hyperparameters. Since the resulting parameter values may not be optimal for all cases, it can lead to variability in model performance.
Model structure	*"Uncertainty in the process being investigated – the structure of the equations that link the response, explanatory variables, and parameters."* Uncertainty in the investigation process refers to the structure of the equations that link the *response*, *explanatory variables*, and *parameters variables* in the SVM and CNN models.	Different model structures have different • assumptions, • parameters, • explanatory variables, • relationship between variables. This uncertainty can impact the model's ability to classify forest types accurately.

Table 1.1: Sources of Uncertainty inspired by framework presented in Simmonds et al. (2022) with the original definitions given in *italic*.

Source	Assessment Steps
Response	1. Examine the ground truth data used for training and validation. Ensure that it is representative of the population of interest and is free from biases or artifacts. 2. Assess the ground truth data's quality, reliability, and variability, including the methods used to collect and label it.
Explanatory variables	1. Investigate the quality of the hyperspectral data from the PRISMA satellite to ensure that they do not introduce biases or artifacts. 2. Analyze the impact of noise or measurement errors on model performance by comparing the model's performance on a dataset tainted by artifacts.
Parameter estimates	1. Assess the performance and convergence of partially trained models by monitoring training progress, comparing with fully trained models, and evaluating model stability.
Model structure	1. Compare the performance of SVM and CNN models using metrics such as accuracy and F1-score to evaluate each model. 2. Investigate the assumptions made by each model and assess whether they are reasonable for the problem domain.

Table 1.2: Assessment Steps for each source of uncertainty in Table 1.1.

1.1 Motivation

The motivation for this thesis lies in the growing importance of remote sensing and, in particular, HSI data in monitoring and managing Earth's natural resources, particularly in forest management. Uncertainty in HSI data, model parameters, and model structure can lead to misclassification errors with real-world consequences and potentially destructive decision-making processes. By investigating how ML models respond to these uncertainties in forest-type classification, the thesis aims to identify challenges and propose potential solutions to reduce misclassification errors.

The choice of SVM and CNN models for this investigation is motivated by their proven effectiveness in image classification tasks, including HSI data analysis [Hasan et al. (2019)]. Through this comprehensive analysis, the thesis seeks to contribute to more accurate and reliable decision-making processes in HSI-based forest management applications, ultimately promoting better stewardship of our planet's natural resources.

1.2 Contribution

The contribution of this thesis is threefold. First, it systematically investigates the impact of various types of uncertainties on HSI data classification, explicitly focusing on SVM and CNN in forest management applications. By exploring the impact of uncertainties related to data, parameters, and model structure, the thesis comprehensively explains the factors contributing to misclassification errors and their real-world consequences.

Second, this thesis applies the proposed framework for uncertainty consideration to assess the performance of SVM and CNN models in forest-type classification tasks. By investigating the datasets used, analyzing the effect of model parameters, and comparing the performance of the models, the thesis identifies the strengths and limitations of each approach. Additionally, it provides insights into the framework's suitability for HSI-based forest management applications.

Finally, the thesis contributes to developing more accurate and reliable decision-making processes in HSI-based forest management applications by shedding light on the challenges and potential solutions for reducing misclassification errors. This, in turn, promotes better stewardship of our planet's natural resources by improving the accuracy and reliability of remote sensing tools used in forest management.

1.3 Thesis Structure

This thesis begins by laying the foundational groundwork for understanding the complex intersections of remote sensing technology and AI in chapter 2. It presents a comprehensive overview of the relevant literature surrounding these topics, providing knowledge for the study of uncertainty within the *response variable*. Additionally, it describes the training and testing procedures, clearly stating the methodologies employed. Following this theoretical groundwork, chapter 3 delves into the specifics of the dataset used in this research.

In chapter 4, the thesis offers a meticulous exploration of *explanatory variables* by investigating how different forms of uncertainty appear. This is performed by analyzing the single-image trained SVM and CNN models' response to artifacts. The chapter also includes the characterization of artifacts and the method to apply them.

From this point, chapter 5 shifts focus onto a broader perspective, evaluating the *explanatory variables* by evaluating the models trained on an extensive training set. This analysis comprehensively explains the model's performance when generalized to a larger dataset. The response of the ML models to variability within the labeled tree species is investigated by assessing how geographical locations and tree age affect their performance. Additionally, the *model parameters* are investigated.

In chapter 6, the thesis transitions into a critical discussion, including the assessment of *model structure* uncertainty. Additionally, the findings from the various stages of this research are synthesized and evaluated using the framework presented in Table 1.2, and summarized in Table 6.1. The conclusion in chapter 7 summarises the insights from this research.

Chapter 2

Background and Theory

This chapter offers essential background knowledge for comprehending the usefulness of remote sensing and AI in forest monitoring. Additionally, it outlines the common errors and uncertainties in hyperspectral data that can impact the accuracy of forest monitoring predictions.

2.1 Remote Sensing

Remote sensing refers to any measurement technique that gathers information at a distance by using devices that are not in physical or intimate contact with the object (typically from satellite or aircraft) [Campbell & Wynne (2011)]. Sensor data of physical objects that scientists wish to examine is collected by recording electromagnetic radiation emitted or reflected from the landscape. Effective use of sensor data requires analysis and interpretation to convert data to information that can be used to address practical problems. The interpretations create extracted information that contains transformations of sensor data designed to reveal specific kinds of information. Finally, applications use the analyzed remote sensing data and other data to address a practical problem [Campbell & Wynne (2011)].

In the 1990s, satellite systems were developed to gather remotely sensed data on the Earth's surface. These images possess unique properties that enable us to detect patterns, monitor changes over time, and measure sizes, areas, depths, and heights. To this day, they are used to study changes in land cover, forest density, coastal morphology, and the biodiversity of remote islands and reefs.

This thesis focuses on a particular type of remote sensing, which involves detecting and monitoring forestry by measuring reflected and emitted radiation. Different materials reflect and emit radiation differently based on their wavelengths [Bakken et al. (2019)]. Therefore, targets can be identified by their spectral signatures in remotely sensed images. An example of such an image, provided by the Hypso-1 satellite, is shown in Figure 2.1.

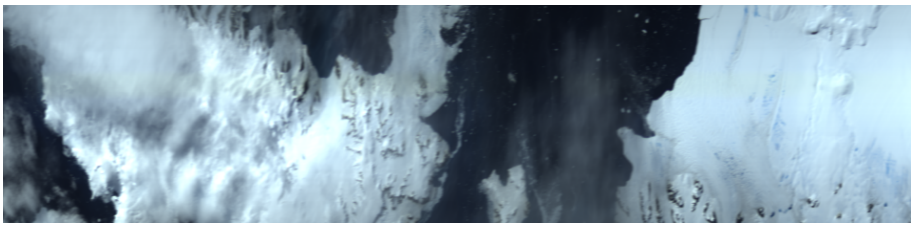
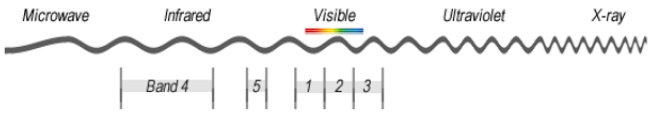


Figure 2.1: Hypso-1 image capturing a part of Antarctica.

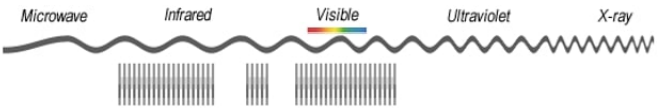
2.1.1 Hyperspectral Remote Sensing

The original form of remote sensing involved aerial photography in the visible portion of the electromagnetic wavelength. Technological developments have enabled information acquisition from satellites over many wavelength bands. This is called multispectral or hyperspectral data, including near-infrared, thermal infrared, and microwave [Kairu (1982)]. Specifically, during the 1980s, scientists at the Jet Propulsion Laboratory began to develop instruments that could collect 200 or more precisely defined spectral regions [Campbell & Wynne (2011)]. From that point, the field of hyperspectral sensing has advanced remote sensing's analytical powers to new levels.

Spectral resolution defines the number and width of spectral bands of a sensor. With a higher resolution, the wavelength range for a given channel and band becomes narrower. Typically, multispectral imagery has less than ten bands, while hyperspectral imagery can have hundreds of continuous narrower bands, as seen in Figure 2.2.



(a) Multispectral imagery



(b) Hyperspectral imagery

Figure 2.2: Visualization of spectral resolution [Geography (n.d.)].

A HSI is formed by a sensor that acquires a spectral vector with hundreds or thousands of elements from every pixel of a given scene, such as the one illustrated in Figure 2.4. The stack of images representing the radiance in the respective band gives an equivalent interpretation of HSI. For this reason, HSIs are called three-dimensional (3D) image cubes. The first two dimensions represent the x and y coordinates of the image pixels and are called spatial information. The third dimension represents the different spectral bands or wavelengths in the image cube,

also called the spectral information of a scene (seen in Figure 2.3) [Bioucas-Dias et al. (2013)].

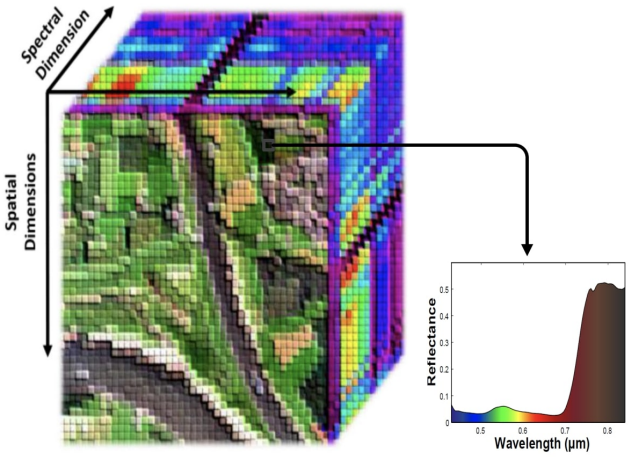


Figure 2.3: Left: Hyperspectral image cube. Right: The reflectance of the material within a pixel. Illustration from [Rasti et al. (2018)].

HSI is a powerful remote sensing and forest monitoring tool due to its unique ability to capture and analyze data in numerous narrow and contiguous spectral bands. Capturing data in hundreds of narrow and contiguous bands allows for more detailed and accurate forest health, composition, and structure analyses. Additionally, it contributes to fields like precision agriculture, water quality assessment, and climate change research, amongst others [Guarini et al. (2017), Loizzo et al. (2018), Bakken et al. (2019)].

HSI is particularly useful in forest monitoring because it enables the identification of specific tree species and can detect subtle changes in forest health before they become visible to the naked eye [Adão et al. (2017)]. It also allows for detecting forest disturbances, such as insect outbreaks or disease, which can be crucial for timely management interventions [Zhao et al. (2022)]. Additionally, HSI can aid in the estimation of forest biomass, carbon storage, and other forest ecosystem services, which are important for climate change mitigation and sustainable forest management [Fernandes et al. (2020)]. Overall, the benefits of HSI in remote sensing and forest monitoring include higher accuracy, improved detection of subtle changes, identification of specific tree species, and better estimates of ecosystem services.

2.1.2 Hyperspectral Satellites

Hyperspectral satellites are space-based imaging systems that capture images of the Earth's surface in hundreds of narrow, contiguous spectral bands. Many countries have developed hyperspectral sensors as the next generation of advanced EO satellites. EnMap (German), HyspIRI (USA), HISUI (Japan), PRISMA (Italy), and HYPSO-1 are some of them [Galeazzi et al. (2008*a*), Sang et al. (2008), Chien et al. (2009), Ohgi & A. (2010), Grøtte et al. (2021)]. This thesis uses data

from the PRISMA satellite.

Over the past decade, extensive research and development have been carried out in hyperspectral remote sensing. However, the benefits of the technology have yet to be realized entirely due to the low temporal resolution caused by the time needed for a large satellite with many users to revisit sites. For real-time applications, this time frame is unacceptable. Thus, the monolithic satellites can not provide a monitoring service, only proof of the technology.

2.1.3 Push Broom Scanning

Push-broom scanning in hyperspectral imaging offers advantages such as high spatial resolution due to continuous data capture, efficient coverage over large areas, reduced spectral mixing due to continuous line scanning, and a high Signal to Noise Ratio (SNR) ratio.

As illustrated in Figure 2.4, the push broom scanner continuously collects spectrograms that form a three-dimensional (3D) data cube while moving over its target surface [Yokoya et al. (2010)]. While the satellite moves along-track, denoted as y in the picture, the image is collected one line at a time. Thus, the dimensions of the image cube are defined by the chosen number of pixels for the cross-track (swath width), the number of frames captured, and the number of pixels used for the bands [Netteland (2022)]. Cross-track is the direction perpendicular to the along-track, typically associated with the image's horizontal dimension (x).

At each of the pixels, a continuous spectral profile is obtained. Though sampling several bands, it is possible to visualize and analyze spectral characteristics of a specific spatial area (x,y), as shown in the bottom right in Figure 2.4. The spectral dimension is denoted as λ.

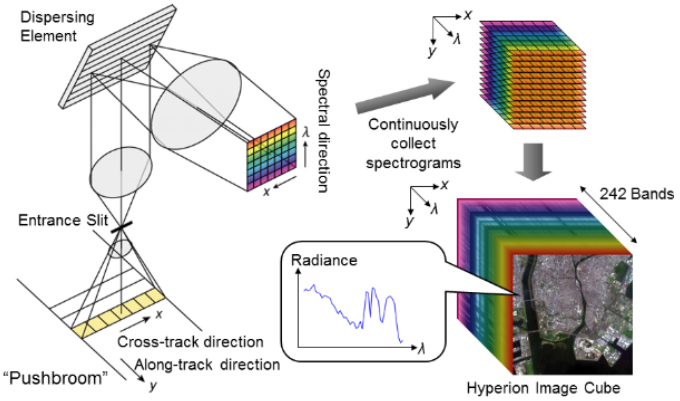


Figure 2.4: Illustration of push-broom scanning [Yokoya et al. (2010)], where samples with their respective bands are captured. It contains cross-track (x), along-track (y), and spectrum (λ) dimensions.

The imaging systems utilized in this study are based on the push broom design, which has demonstrated a sufficient signal-to-noise ratio for remote sensing applications, whether from

airborne or orbiting platforms. Acquiring high-quality spectroscopic data can be challenging, as artifacts that obscure spectral and spatial features may affect the resulting spectra. Despite these limitations, push broom imaging requires a more straightforward pointing system than the alternative whiskbroom scanning and has a less complex design than snapshot hyperspectral imaging systems.

The utilization of push broom scanners within the domain of remote sensing yields a substantial volume of HSI data, necessitating the application of advanced methodologies such as ML and AI for comprehensive analysis.

2.2 Artificial intelligence

Artificial Intelligence (AI) is a rapidly growing field of computer science that can perform tasks that typically require human intelligence. The main idea is to create algorithms and models that allow computers to reason as humans. It is becoming an increasingly important tool in remote sensing and forest monitoring applications. With the large amount of data collected through remote sensing, AI can help to efficiently analyze and process the data to identify patterns, detect changes, and classify land cover. In forest monitoring, AI can help to identify and map forest cover, assess forest health, and detect forest disturbances such as deforestation or forest fires.

In recent years, the field of AI has increased drastically and formed multiple disciplines within. A simple representation of relevant technologies for this thesis and their relationship is described in Figure 2.5. In simple terms, AI contains the subset of ML which, in turn, includes the subset of Deep Learning (DL).

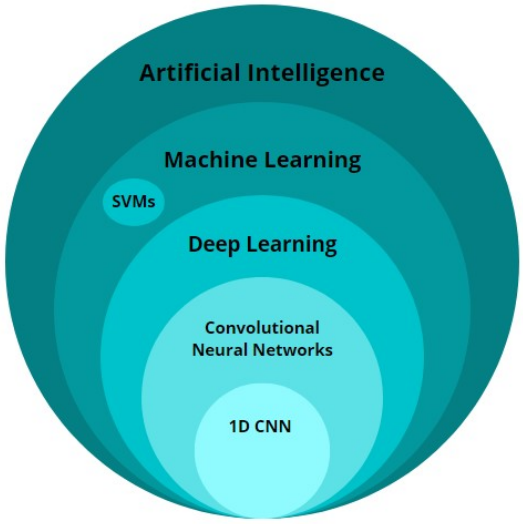


Figure 2.5: An illustration of how the field of AI is structured.

ML is one of the most common forms of AI, which involves the machine's ability to keep improving its performance without humans having to explain exactly how to accomplish all the tasks it's given [Brynjolfsson & Mcafee (2017)]. As excellent learners, ML algorithms achieve superhuman performance in recognizing patterns and predicting within various fields.

DL is a sub-field of ML, taking the concept a step further by attempting to mimic the human brain's functionality. This enables the processing of unstructured data such as documents, images, and text [Wolfewicz (2022)]. To achieve this, deep learning applications use a layered structure of algorithms called an Artificial Neural Network (ANN).

2.2.1 Support Vector Machines

SVMs are a popular and widely used class of supervised ML algorithms that have been applied to various remote sensing and forest monitoring tasks. This thesis uses a linear SVM with Stochastic Gradient Descent (SGD) optimization.

To understand the basic idea behind linear SVMs, one needs to grasp the three fundamental concepts: *separating hyperplane*, *maximum-margin hyperplane*, *soft margin* [Noble (2006)].

- *Separating hyperplane*: SVMs aims to find a hyperplane (a multidimensional plane) that separates the different classes of data as well as possible. This hyperplane is called the separating hyperplane. Specifically, the separating hyperplane tries to maximize a margin, the distance between the hyperplane, and the closest points from each class.

- *Maximum-margin hyperplane*: The hyperplane that maximizes the margin, found by the SVM algorithm. Maximizing the margin allows the SVM algorithm to find the best separation boundary between the data classes.

- *Soft margin*: In practice, finding a hyperplane that perfectly separates the data is not always possible. This is solved by using a soft margin to find a hyperplane that balances between maximizing the margin and minimizing the misclassification errors.

SVMs offers several advantages over other methods. One advantage is that they are easy to visualize and understand. SVMs aims to find the best hyperplane that separates the data into two or more classes. This is achieved by maximizing a particular mathematical function concerning a given data collection. The resulting hyperplane can be visualized as a line or plane in the feature space, as seen in Figure 2.6, making it easy to understand how the SVM is making its classification decisions. This ease of visualization can help users better understand and interpret their analyses' results, even for users without a strong background in ML. Additionally, SVMs can be implemented using a variety of software packages and programming languages, making them accessible to a wide range of users.

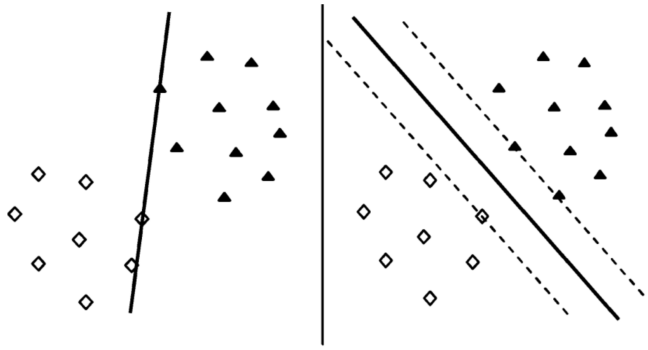


Figure 2.6: Illustration of a Support Vector Machine. (**left**) Separating hyperplane exhibiting zero margins, compared to (**right**) the maximal margin separating hyperplane, for the same classes of training samples presented in feature space [Mavroforakis & Theodoridis (2006)].

Another advantage of SVMs is their flexibility and ability to handle different data types. SVMs can be applied to a wide range of data types, including continuous, categorical, and ordinal data, making them versatile tools for many different types of analyses. Additionally, they can easily handle high-dimensional data, allowing for the analysis of large datasets with many variables. This means that they can be used in HSI.

Because SVMs are a well-established and widely used method, it is well recognized in the ML community. This means that they are often used as a benchmark for evaluating the performance of other machine learning methods, making them useful tools for comparing the performance of different algorithms. In this project, a comparison of a SVM and a CNN is made.

In summary, SVMs offer several advantages, including ease of visualization, flexibility, simplicity, and standardization. These features make SVMs a convenient and powerful tool for a wide range of ML applications. It can be used for land cover classification, image segmentation, and feature extraction in remote sensing [Shi & Yang (2015), Roli & Fumera (2001)]. Additionally, it can be trained on labeled data to classify different land cover types, such as forests, agricultural fields, or urban areas. This project classifies forest types using SVMs.

Stochastic Gradient Decent

The optimization problem in SVMs can be solved using different algorithms. Quadratic programming, sequential minimal optimization, and gradient descent are some of them, in which the latter is used in this thesis.

SGD is a variant of gradient descent that allows for specifying the maximum number of iterations, the stopping criterion, the learning rate schedule, and the initial learning rate. It is particularly well-suited for large datasets, where the training examples are processed one at a time. This is because the SGD algorithm updates the model parameters after processing each

training example, allowing the model to learn from each example without storing the entire dataset in memory. While the HSI data sets used in remote sensing can be extensive, and the computational cost of training a model on them can be very high, the SGD algorithms allow for efficient optimization of the model parameters even in high-dimensional feature spaces.

Furthermore, SGD is well-suited for online learning scenarios, where the model is updated as new data becomes available. In remote sensing applications, this is particularly useful for monitoring and tracking changes in land use, vegetation cover, and other environmental factors over time.

In this thesis, the Python library, Scikit-learn (sklearn) [scikit learn (2023)], has been extensively employed. It offers a comprehensive suite of tools for ML and statistical modeling, which includes efficient implementations of both SVMs and SGDs. These algorithms are highly customizable within sklearn, allowing users to define and tailor their models to their needs. The library's user-friendly interface, strong documentation, and active community support make it a crucial resource for scholars and practitioners in remote sensing and similar fields.

2.2.2 Artificial Neural Networks

Artificial Neural Networks (ANN) are state-of-the-art ML models that can learn to extract features from remote sensing data to make predictions or classifications. These models can be trained on large datasets to identify patterns and relationships within the data and can be used to make accurate predictions about forest cover or forest health. However, ANNs can be less intuitive than simpler models like SVMs, making it harder to understand what is happening under the surface. This is where SVMs can be useful, as they can explain uncertainty in simple and complicated cases. Nonetheless, the power and accuracy of ANNs make them invaluable tools for advanced remote sensing and forest monitoring applications.

The main element of ANNs is the various connected nodes or units - technically known as artificial neurons. The nodes form an interlayered system in which each layer is responsible for inputting, processing, and outputting data to the deeper layers. As illustrated in Figure 2.7, the input layer is the leftmost layer, and the rightmost layer is the output layer. The middle layers describe the network's depth and are called the hidden layer. The network of neurons works similarly to the human brain, allowing it to identify hidden correlations and patterns in raw data, classify them, and continuously improve [Memon (2022)].

As illustrated in Figure 2.7, each neuron in the network produces a single output that can be sent to other neurons out of its input; the input can be images or documents, or outputs of other neurons. A neuron takes the weighted sum of all the inputs to provide the correct output and adds a bias term to this sum [Dawson & Wilby (1998)]. The output of the final neurons in the network is the recognition of patterns to solve a given task.

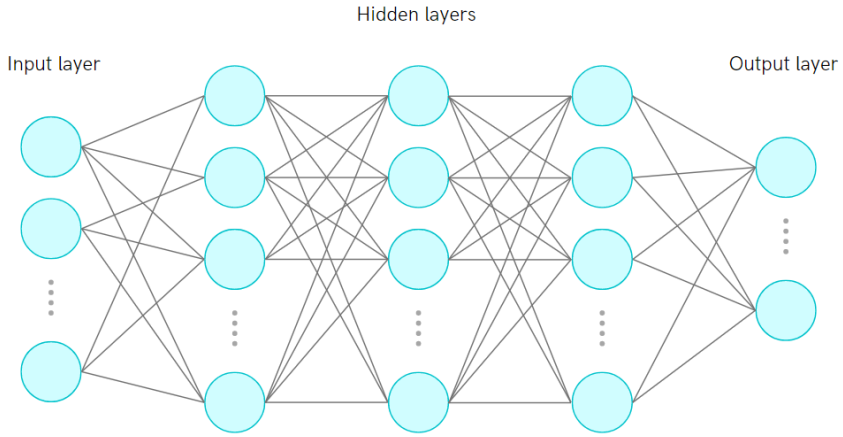


Figure 2.7: A simple Artificial Neural Network with inputs and outputs. Each blue circle corresponds to a neuron, and each grey line corresponds to a connection.

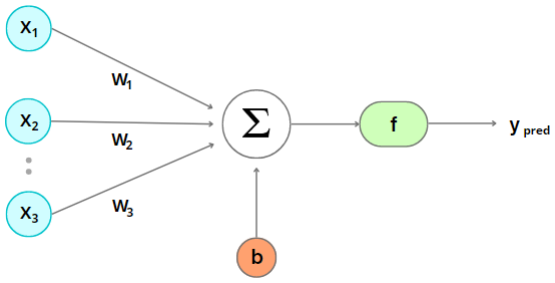


Figure 2.8: An artificial neuron used in ANNs, in which x denotes the input, w the weights, b the bias, f the activation function and y the output.

Activation Function

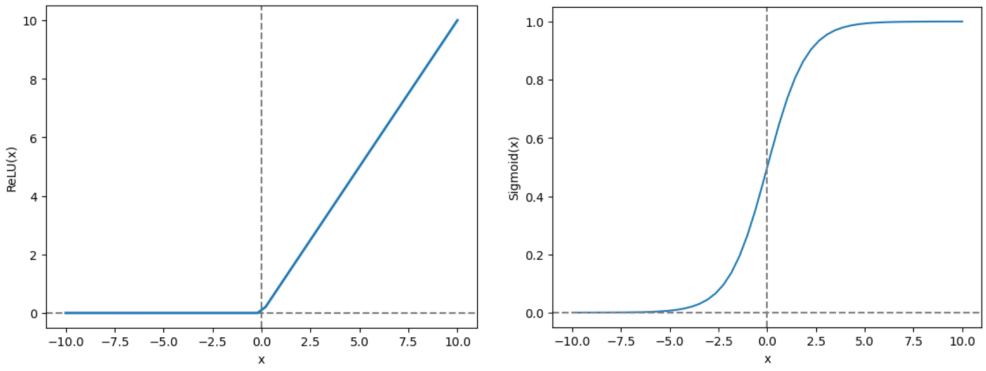
Activation functions are nonlinear functions that decide whether a neuron should be activated. It will activate if the network's input is essential in the prediction process. To this day, multiple commonly used activation functions exist, with Rectifier Linear Unit (ReLU) and Sigmoid (Figure 2.9) being two of them.

The ReLU function is described in Equation 2.1. It will output the input directly if it is positive. Otherwise, it will output zero. The main advantage is that it does not activate all the neurons simultaneously.

$$f(X) = \begin{cases} x & \text{if } a > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

The Sigmoid function is a non-linear function of the weighted sum of inputs, as seen in Equation 2.2. It is commonly used for binary classification in ANNs. When using it, the output will always be between 0 and 1.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$



(a) ReLU (b) Sigmoid

Figure 2.9: Activation functions.

2.2.3 Convolutional Neural Network

CNN is analogous to traditional ANN networks, as seen in Figure 2.5. The only notable difference between CNNs and traditional ANNs is that CNNs are primarily used in image and video recognition, recommendation systems, and image analysis and classification [Dawson & Wilby (1998), O'Shea & Nash (2015)]. Specifically, it can automate the image analysis process in remote sensing. With a large quantity of data in spectral, temporal, and spatial domains, CNN

techniques have shown to be successfully implemented for intelligent onboard processing, advanced database interrogation, and the automated analysis of multispectral imagery [Oche et al. (2021)]. CNNs have high accuracy in image classification and object detection, efficient processing of large amounts of image data, and robustness to environmental factors. Transfer learning can adapt pre-trained models to specific remote sensing tasks.

CNNs consist of multiple layers, including *convolutional layer*, *pooling layer*, and *fully-connected layer*, in which the data is passed through to extract features. [Albawi et al. (2017)]. With each layer, it becomes more capable of comprehending the image's intricacies. This implies that the initial layers concentrate on analyzing basic elements in the image, such as its edges and hues. As the image travels through subsequent layers, the network becomes proficient in identifying more intricate attributes, such as object forms.

Convolution

The convolution layers are the most significant in CNN, indicating the locations and strengths of a detected feature in an input. It utilizes the mathematical operation convolution, allowing for the merging of two sets of information.

CNNs applies convolutions to input data to extract features from image matrices using kernels. However, even though the name implies that convolution is made, most CNNs use the related cross-correlation function, calculated according to the formula given by:

$$net(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i+m, k+n)K(m,n). \tag{2.3}$$

In Equation 2.3, $net(i,j)$ is the output of the next layer, I is the input image, $*$ is the convolution operation, and K is the kernel, also called the feature detector. The kernel goes through the input image and performs element-wise matrix multiplication for it to be summed together before passing through the respective activation function. It is the activation function that defines the resulting value to be stored in the feature map.

Figure 2.10 shows how the convolution works. As illustrated, the element-wise product of the input and the kernel, given by Equation 2.4 is aggregated. The result represents the corresponding point in the next layer [Albawi et al. (2017)].

$$o_{11} = f(k_{11}a_{11} + k_{12}a_{12} + k_{13}a_{13} + ... + k_{31}a_{31}) \tag{2.4}$$

One of the drawbacks of the convolution step is the loss of information that might exist on the border of the image [Albawi et al. (2017)]. The kernel is a filter that moves across the image, scanning each pixel and converting the data into smaller, or sometimes larger formats. In order to assist the kernel with processing the image, padding is applied to allow for more space for the

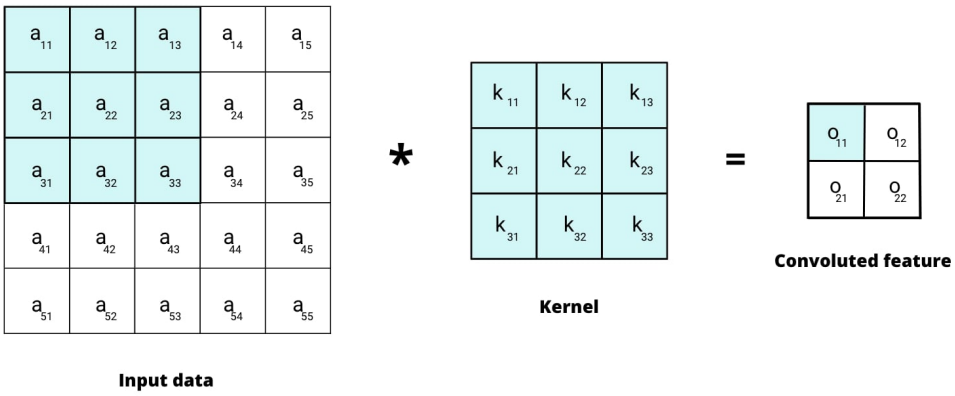


Figure 2.10: Details on convolution process. Kernel sliding over input data performing cross-correlation with a stride of 1.

kernel to cover the image.

Padding is specifically helpful for convolutional layers that need to maintain the spatial dimensions of their inputs, as it has a tendency to reduce dimensions depending on the kernel size and stride. A very simple, yet efficient method to resolve the issue, is to use zero-padding as seen in Figure 2.11. It essentially places zeros around the image cube and prevents network output size from shrinking with depth.

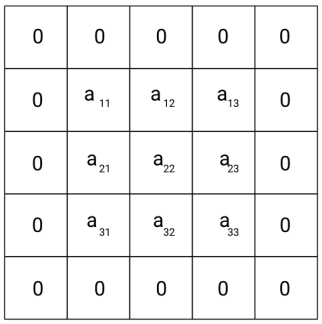


Figure 2.11: Zero padding applied to input in order to maintain spatial dimensions in a convolutional layer.

Max Pooling

CNNs almost always incorporates a pooling layer to summarize and downscale features from previous convolutional layers to reduce the dimensions of the feature maps. This reduces the number of parameters to learn and computations performed in the network. In addition, it makes the model more robust to variations in the positions of the features in the input image.

There are multiple types of pooling operations, where max-, average-, and global pooling are the most common. Max pooling operates on the hidden layers of the CNN and reduces their size by the integer multiplicative factor, $alpha = 2$ [Graham (2014)]. As illustrated in Figure 2.12 described in Equation 2.5, it selects the maximum element from the region of the feature map covered by the filter, resulting in a feature map containing the most prominent features of the previous feature map. This is the only pooling type used for the U-Net network used in this thesis.

$$o_{11} = max(a_{11}, a_{12}, a_{13}, a_{31}) \tag{2.5}$$

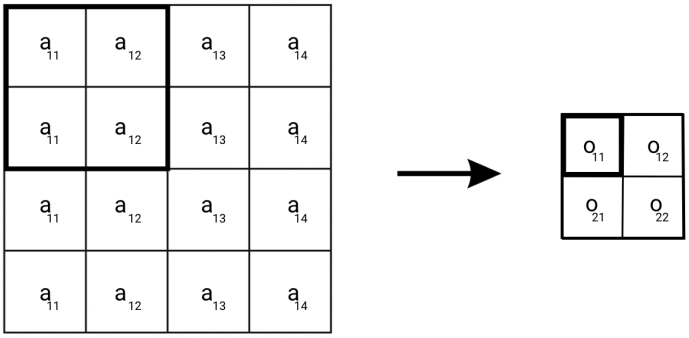


Figure 2.12: Max pooling demonstrated with a 2x2 filter and a Stride of 2. [Ronneberger et al. (2015)].

2.2.4 1D CNNs

CNNs are often used for processing two-dimensional data, such as images, where the convolutional layers of the Neural Networks (NN) learn spatial features from the input data. However, CNNs can also be used for 1D data, learning spectral or temporal features from the input sequence. An HSI cube typically comprises N_x x N_y pixels, where each pixel contains spectral data in N bands. To prepare the spectral data for analysis, a 1D input vector is created for each pixel by extracting the spectral data contained within that pixel [Hsieh & Kiang (2020)].

The goal of HSI classification is to assign each pixel in an image to a predefined class label based on its spectral signature. In this context, 1D CNNs can be used to learn spectral features from the pixel-wise spectral data, which can then be used for classification. The 1D CNN is trained using input vectors from a selected set of training pixels, as shown in Figure 2.13. Following training, the performance of the 1D CNN is evaluated using input vectors from a separate set of testing pixels.

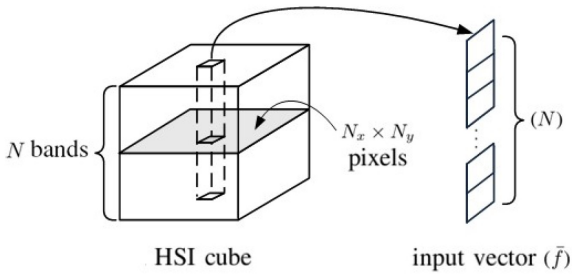


Figure 2.13: 1D input vector retrieved from a HSI cube [Hsieh & Kiang (2020)].

The ability of 1D CNNs to learn feature representations directly from hyperspectral data and capture spatial context can be particularly useful in applications such as land cover classification, where the spatial distribution of classes is essential. Moreover, the ability of 1D CNNs to extract local features from hyperspectral data can also be necessary for applications such as land-use classification, where temporal variations in vegetation or land cover can indicate a specific class. Compared to traditional machine learning algorithms like SVMs, which require manual feature extraction, 1D CNNs can automate this process, saving time and effort. This makes them well-suited for hyperspectral data classification tasks where feature extraction is a crucial step. Therefore, combining the spatial and temporal context captured by 1D CNNs can provide a powerful tool for hyperspectral data classification in a wide range of applications.

LucasCoordConv 1D CNN

While 1D CNN have been considered less effective for hyperspectral classification tasks, Riese & Keller (2019) demonstrated their potential on the LUCAS dataset [Centre (2023)]. They studied soil texture classification based on hyperspectral data and implemented three networks: LucasResNet, LucasCoordConv, and LucasCNN. This thesis employed the LucasCoordConv architecture, which uses pixel-wise spectral data.

As illustrated in Figure 2.14, the LucasCNN architecture comprises four convolutional layers, each followed by a max-pooling layer with a kernel size of two. The first two convolutional layers use 32 filters with a kernel size of 3 and a ReLU activation function. The last two layers use 64 filters with a kernel size of 3 and a ReLU activation function. The output of the last pooling layer is flattened and passed through three fully connected layers with 120, 160, and 6 neurons when adjusted for the Faubai dataset, respectively. Additionally, a "softmax" activation function in the last layer produces the output probabilities for the four classes.

In the development of LucasResNet, an identity block is incorporated into the foundational LucasCNN structure. This allows the input vector to circumvent the four convolutional layers and instead be concatenated directly to the activation output from the final convolutional layer before being processed by the initial Fully Connected (FC) layers. Unique to LucasCoordConv

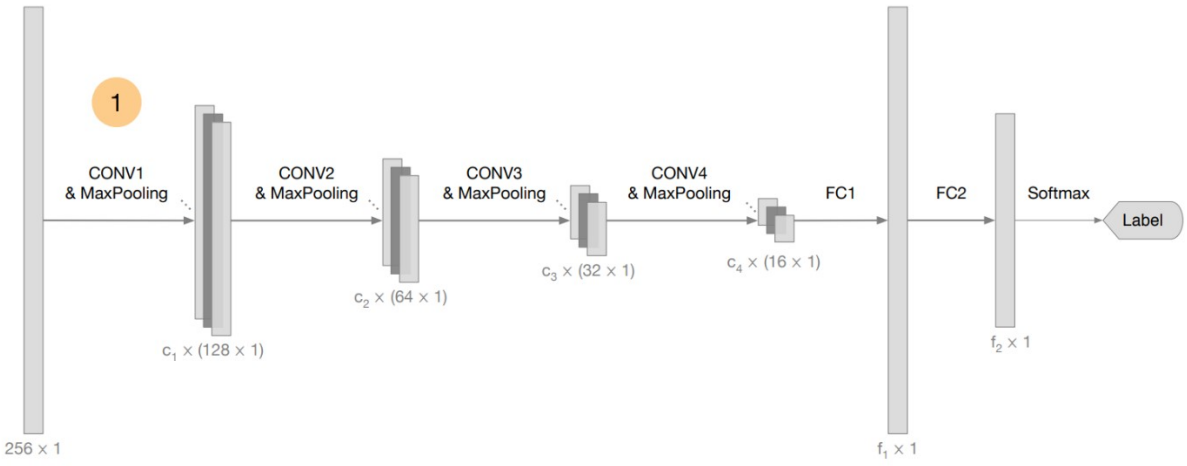


Figure 2.14: Flowchart of the LucasCNN. The network consists of convolutional, Fully Connected (FC) layers and max-pooling layers. The i-th CONV layer consists of c_i filters, and the j-th FC layer consists of f_j units. At the end of each network, a softmax layer provides the classification output as a 4-dimensional vector. For the LucasCoordConv, a coordinates layer is inserted before the first CONV layer at marker 1. [Riese & Keller (2019)].

is the positioning of a coordinate layer strategically placed preceding the initial convolutional layer within the LucasCNN framework. Specifically, it was found that adding a coordinates layer in the LucasCoordConv improves the classification performance significantly[Riese & Keller (2019)]. While the other classifiers tend to misclassify more than 70% of class T as class L, it is worth noting that the LucasCoordConv model performs better, misclassifying only around 48% of class T as class L, as depicted in Figure 2.15. Consequently, this thesis utilizes the LucasCoordConv model for further analysis and experimentation.

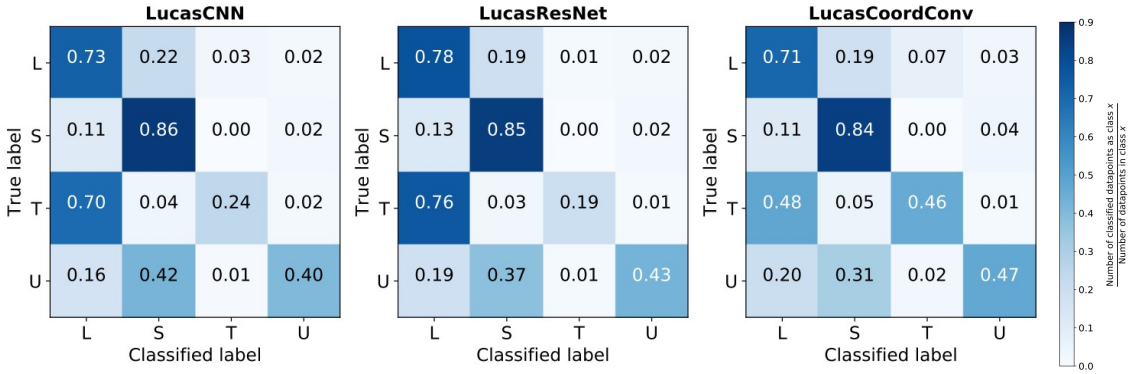


Figure 2.15: Normalized confusion matrices [Riese & Keller (2019)].

2.3 Uncertainty

Remotely sensed data and HSI have limitations. One of the key challenges in interpreting and using this data is dealing with uncertainty that arises from various sources. Understanding and quantifying uncertainty are critical for using remote sensing data effectively and making informed decisions based on the results.

2.3.1 Uncertainty in Artificial Intelligence Models

Uncertainty is a significant challenge in AI models. As the thesis initially described, Simmonds et al. (2022) divided the model-related uncertainties into data, parameters, and model structure. The data component is divided into two sub-sources: the 'response', which refers to the focal variable being explained (i.e., tree classification), and the 'explanatory variables', which encompass any variables used to explain or predict the response.

These factors can significantly affect the model's performance and make obtaining accurate and reliable predictions challenging. Moreover, incomplete consideration of model-based uncertainties can result in erroneous conclusions with consequences in the real world, eroding public trust in science. Such incomplete consideration of uncertainty is common within and across various scientific fields, despite the potential harm it can cause.

Response Uncertainty

'Response' uncertainty relates to the uncertainty in the output of the model. This is a fundamental concern in AI, particularly in ML applications, where the models are designed to make predictions or classifications based on input data. The output generated by these models is inherently uncertain due to various factors, including incomplete or noisy data, limitations in the modeling techniques used, and the complexity of the modeled real-world phenomena.

In particular, labeling data accurately can be challenging, as labels may come from multiple sources with varying degrees of reliability, or the labeling process itself may be imperfect. As a result, AI models may not always produce consistent or accurate output, leading to uncertainty in their predictions and classifications.

Within forest monitoring, output uncertainty is a significant concern as the complexity of the forest's content makes it difficult to label data accurately. The labels may come from various sources, and they can be discrepancies due to differences in interpretation or measurement techniques. As a result, the labels may be uncertain or even incorrect, leading to challenges in training machine learning models. Additionally, forests are challenging to characterize in detail, as they contain many different features and characteristics that can be difficult to capture fully. This means that some data may be lost or hidden due to the AI model choosing one label over another, even if there is a slightly higher chance of that label being correct.

Addressing output uncertainty requires careful consideration of the AI model's limitations, the input data's quality and reliability, the labeling process, and the AI models used to ensure that decisions are made based on accurate and reliable information. By understanding and addressing output uncertainty, practitioners can improve the accuracy and reliability of their AI models and ensure that trustworthy and reliable data inform their decisions.

Explanatory Variable Uncertainty

The 'explanatory variables' are connected to the input uncertainty. It refers to noise or errors in the data fed into a model, which can significantly impact its performance.

Additionally, the selection of inputs can also introduce uncertainty, as some variables may be more important or relevant to the model than others. ML practitioners must consider input uncertainty carefully, and take measures to minimize its impact on their models. The challenge of choosing the right combination of input features and output labels is crucial for achieving accurate predictions and classification results. This can be difficult due to a lack of understanding of the data distribution and its characteristics.

In HSIs, each pixel in a spatial image is represented by a vector of intensity values across many spectral bands. This high-dimensional form of data can present a challenge for input-output selection, such as classification or regression, as it can be difficult to determine which bands are most relevant for a given task. Moreover, different tasks may require different input-output pairs, adding a layer of complexity to the input-output selection process.

The choice of input features can be further complicated by noise or irrelevant features that can negatively impact the model's performance [Simmonds et al. (2022)]. However, by addressing input uncertainty in models, practitioners can improve the corresponding accuracy and reliability, and ensure that their models make well-informed decisions based on high-quality data.

Parameter Estimates Uncertainty

'Parameter estimates' uncertainty plays a crucial role in determining the performance of models. Tuning the model parameters to the right values can significantly improve the model's accuracy and robustness. However, determining the optimal parameter values can be challenging and often requires trial and error. Multiple local minima in the parameter space can also complicate the optimization process.

In HSI, selecting model parameters is important in developing effective models for classification or segmentation tasks. Some common model parameters include the number of Spectral bands, Learning rate, Drop-out rate, number of filters, Batch size, number of Epochs, and Activation function. Choosing appropriate values for these parameters is important for improving model efficiency and accuracy. Overall, the optimal values depend on the specific task and properties of the HSI data, and experimentation may be necessary to determine the best values.

Model Structure Uncertainty

Deciding which model to use can be challenging, and it often requires careful analysis of the data and problem domain. Different models have different strengths and weaknesses, and their suitability for a particular task can depend on various factors, such as the size and complexity of the data, the number of classes, and the availability of labeled data [Brownlee (2019)]. Some models may be better suited to high-dimensional data than others, and the performance of a given model can vary depending on the number and nature of the input features.

Traditional machine learning models designed for lower-dimensional data, such as SVMs and decision trees, may not be well-suited for high-dimensional HSI data in all cases. However, SVMs are excellent for high-dimensional data with fewer samples, focusing on finding hyperplanes in a multidimensional space that distinctly classifies the data points. They are less prone to overfitting and can handle a large feature space, like in HSI data. However, SVMs can struggle with large datasets or complex spatial structures inherent in the data, as they primarily work with vector inputs.

On the other hand, DL models such as CNNs are highly effective in image and video processing tasks as they can learn the spatial relationships between pixels. CNNs can capture local and global patterns in the data and thus, theoretically, should perform better with HSI data, which often requires analyzing spatial and spectral features simultaneously. However, understanding when and where a CNNs may fail can be challenging due to its deep hierarchical structure and nonlinear transformations.

In practice, the performance of these methods can vary significantly and would depend on factors like the dataset's size, the data's complexity, the way features are extracted and utilized, the amount of training, and the degree of tuning the models receive. Therefore, choosing the right algorithm for a particular task requires a careful analysis of the data and the problem domain and an understanding of the strengths and limitations of different algorithms.

2.3.2 Signal to Noise Ratio

In many ML applications, the input data may be corrupted by noise or other sources of variability that can impact the quality and reliability of the model's output. In such cases, the Signal to Noise Ratio (SNR) is one way to characterize the amount of noise.

The SNR measures the relative strength of the signal compared to the noise. When noise is applied to an image, a desired SNR can control the noise level in the resulting signal. It is common to express the SNR in decibel (dB) to provide a more intuitive way to interpret the signal's noise level. SNR is defined as the ratio of the power of the signal to the power of the noise:

$$SNR_{dB} = 10 * log10(SNR) \tag{2.6}$$

where SNR is the ratio of the signal power to the noise power [Johnson (2006)]. The "log10" function compresses the broad range of values into a smaller range, making it easier to analyze.

A positive SNR (dB) means that the signal power is higher than the noise power. The high positive values indicate a higher quality signal, with a stronger signal than the noise. Conversely, a negative SNR (dB) means the noise power is higher than the signal power. In this case, the noise can dominate the signal, making it difficult to detect or interpret. A low negative value indicates a low-quality signal, in which the noise is stronger than the signal. This can result in significant uncertainty in the model's output, as the signal may be overwhelmed by noise, making it difficult to distinguish between different classes or to make accurate predictions.

In summary, a higher desired SNR in dB means less noise will be added to the signal, while a lower desired SNR in dB means more noise will be added to the signal. Practitioners must carefully consider the SNR when developing ML models, as it can significantly impact the accuracy and reliability of the model's output.

2.3.3 Uncertainty in Hyperspectral Data

While HSI can provide a wealth of information, it can be degraded by different sources, leading to reduced accuracy and quality of the data. In recent years, the uncertainties present in HSI data have led to erroneous classifications [Glenn et al. (2005)]. For this reason, it is essential and challenging to characterize the model response caused by the uncertainty in HSI data. Specifically, various artifacts are present in HSI data, including *gaussian-, poisson-* and *salt and pepper* noise. While there are other sources of uncertainty, such as biases due to atmospheric conditions, sensor limitations, and pre-processing steps, this section concentrates on artifacts' significant impact on the accuracy and quality of HSI data.

The artifacts that affect HSI often come in mixtures, and can broadly be partitioned into signal-dependent and signal-independent noise. Signal-independent noise in HSI is uncorrelated in the spectral domain. It is not related to the signal intensity and is present regardless of the signal. *gaussian-, salt and pepper, stripe* noise, and *defocusing* are examples of signal-independent noise. On the other hand, signal-dependent noise is related to the intensity of the signal and can be modeled as a function of the signal. *Poisson* noise is an example of signal-dependent noise since it is related to the intensity of the signal being detected.

Smile and *keystone* are not examples of signal-dependent or signal-independent noise, as they are not related to the intensity of the signal. They are optical aberrations that can affect the geometry and curvature of the hyperspectral data.

Pixel Uncertainty

Different types of noise are the most common errors that affect each pixel in a HSI, including *gaussian-*, *poisson-*, and *salt and pepper* noise. Understanding their properties is crucial for modeling the noise to improve the accuracy and quality of the data.

Gaussian noise is a type of signal-independent noise that follows from the Gaussian distribution:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} \tag{2.7}$$

where μ is the mean of the average value of z and σ is its standard deviation. The standard deviation squared, σ^2, is called the variance of z [Yi & Choi (2010)]. The noise level added to an image depends on the mean and standard deviation of the gaussian distribution.

Various sources, such as sensor noise, variations in illumination, and atmospheric noise, can cause it. The atmospheric noise is caused by the interaction of electromagnetic radiation with the Earth's atmosphere. It can degrade the quality of HSI by introducing absorption, scattering, and reflection of the radiation [Hong-Xia et al. (2022)].

Modeling using *gaussian* noise considerably simplifies the analysis and the noise variance estimation. The types of noise that are considered to be spectrally uncorrelated can be modeled as signal-independent *gaussian* noise [Kerekes & Baum (2003), Landgrebe & Malaret (1986)]. Thus, the noise affects the whole image similarly, as seen in Figure 2.16. An example is Read noise, which is introduced during the readout process of the detector. It is caused due to the electronics that amplify the signal from the sensor.

Poisson noise, also called photon or shot noise is a type of noise that arises due to the random nature of photon arrival at the detector [Hasinoff (2014)]. It is modeled by the Poisson distribution given by:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{2.8}$$

where *x!* denotes factorial and the parameter λ determines the shape of the distribution.

It is proportional to the square root of the signal intensity, for which the noise variance is signal dependent. The image in Figure 2.16 shows that the lighter parts are noisier than the dark parts. Thus, *poisson* noise is particularly relevant in low-light conditions where the number of photons detected is small, making it harder to identify spectral features accurately.

Salt and Pepper noise noise is a type of impulse noise that presents itself as sparsely occurring black and white pixels randomly distributed across the image. It can be caused by sharp and sudden disturbances in the imaging signal or data corruption during transmission or storage [Rasti et al. (2018)]. It can affect the accuracy of hyperspectral data by introducing spikes or discontinuities in the image. For impulse noise, only a few pixels are modified and replaced by black or white pixels as seen in Figure 2.16.

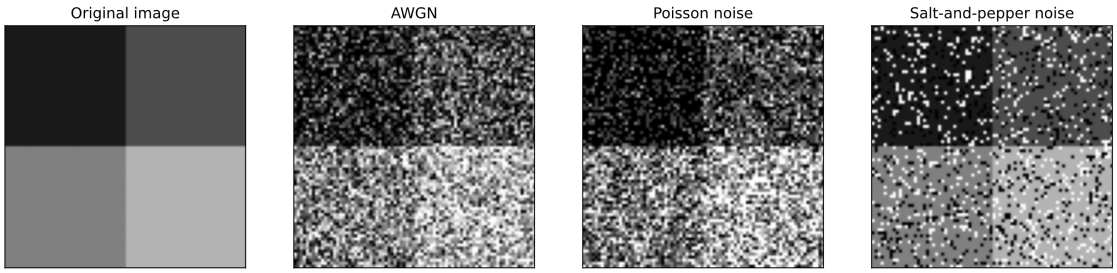


Figure 2.16: Example of different types of pixel noise, with the original image, Additive white Gaussian noise, Poisson noise, and Salt and Pepper noise [Mazet (2023)].

Spectral Uncertainty

The spectral dimension of HSI often suffers from pattern noise, such as ***Stripe noise***, in addition to spectral and spatial misregistrations. *Stripe noise* is pattern noise in HSI as regular or irregular horizontal or vertical stripes of varying intensity across the spectrogram. Because stripe noise introduces repeated spectral distortions, it can be detrimental to ML algorithms. For this reason, stripe noise reduction for push-broom scanning techniques has been widely studied in the remote sensing literature, particularly for HSI [Rasti et al. (2018)].

It can be caused by various factors, such as manufacturing defects in the image sensor, variations in the manufacturing process, electrical interference, or sensor degradation over time. Additionally, it can be caused by non-uniform illumination or exposure during image acquisition. This occurs when the illumination source is not uniform across the scene, or when the sensor is not exposed equally to all parts of the scene. The variations in the signal level across different bands lead to stripe noise.

Spectral and spatial misregistration in HSI arise due to aberrations and misalignments in the optical system. They can cause the errors seen in Figure 2.17, describing the geometric distortions ***Smile*** and ***Keystone***. The letters B, G, and R denote the spectral range's short, middle, and long wavelengths, respectively [Mouroulis et al. (2000)]. The circles represent the Point Spread Function (PSF), describing how a point source of light appears after it passes through an optical system. It represents the blurring or spreading of a point light source due to various factors, such as diffraction, aberrations, and other imperfections in the optical system [Rossmann (1969)]. Ideally, the size and shape of the PSF should keep the same size and shape for each wavelength. Thus, the lines in the figure should ideally be straight and follow the grid.

Smile is a curvature distortion of the spectral line, illustrated as the blue curved circles in Figure 2.17. In a spectrogram, a smile distortion refers to a change in the apparent wavelength of spectral lines as a function of their position along the height of the slit [Henriksen et al. (2019)].

This is typically caused by a misalignment of the slit with the optical axis, causing different wavelengths to be focused at slightly different positions along the detector. In HSI, the spectral lines appear curved, which can lead to errors in the measurement of the spectral lines' wavelengths.

Keystone distortion refers to the geometric distortion of the focal plane rectangle into a trapezoid shape due to a misalignment of the slit concerning the detector's axis [Henriksen et al. (2019)]. This can happen when the slit is not perpendicular to the detector surface, causing the different wavelengths to be dispersed along different angles. In HSI, *keystone* distortion makes the spectral lines appear tilted, leading to errors in measuring the spectral lines' intensities.

Both *smile* and *keystone* lead to reduced accuracy in detection and classification algorithms. They distort spectral profiles and thus degrade classification results [Yokoya et al. (2010)]. Small misregistrations can give significant errors, making it a crucial factor in the quality of HSI. However, due to aberrations, distortions, and diffraction, perfect co-registration may not be possible in practical optical design. Therefore, characterizing these errors and implementing corrections is important to compensate for their effect.

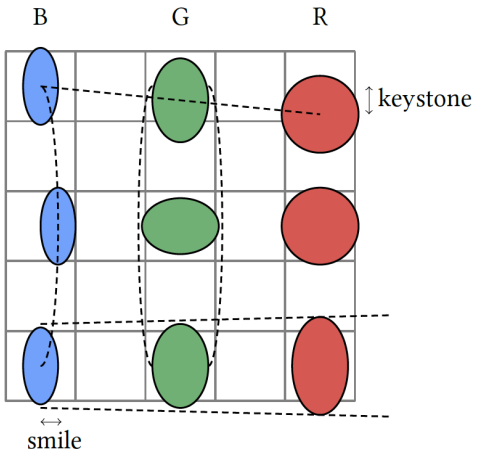


Figure 2.17: Schematic of errors in the spectrum of a nonideal push-broom imaging spectrometer. Illustration by Henriksen (2019), inspired by Mouroulis et al. (2000).

Spatial Uncertainty

Wavelength Dependent Defocusing (WDD) is a type of optical aberration that can arise in HSI. Dispersion causes the focal lengths of refractive and diffractive optical elements, like focal length, to vary with wavelength [Hillenbrand et al. (2012)]. Thus, the different point spread functions are subject to WDD, resulting in a loss of spatial resolution and decreased overall image quality.

In HSI, the spatial and spectral information in the image is tightly coupled, meaning that a loss of spatial resolution can directly impact the accuracy of the spectral information. This can

make it more difficult to accurately identify and distinguish between different spectral features in the image. For this reason, specialized lenses or optical elements, or post-processing algorithms are common approaches to mitigate the effect.

Yiwei et al. (2022) evaluated the performance of the HSI presented by Sigernes et al. (2018), and simulated the optical design. The performance showed significant chromatic aberration, which resulted in image quality degradation away from the central 550nm wavelength. This was due to the use of non-achromatic lenses. By substituting the primary lens with an achromatic lens of the same focal length, the image quality improved and was more consistent across the detector array's field of view. This is illustrated by the spot diagrams in Figure 2.18.

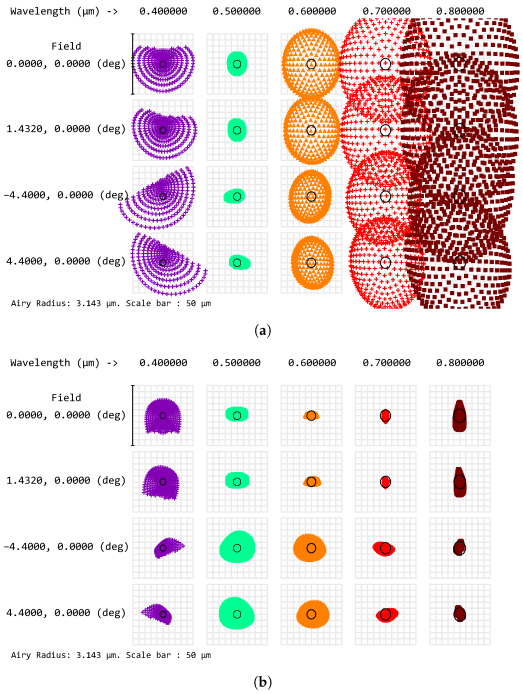


Figure 2.18: Spot diagrams. (**a**) Sigernes et al. (2018) original design, and (**b**) Yiwei et al. (2022) optimized design. The image is taken from [Yiwei et al. (2022)]

2.4 Forest Classification (FauBai Project)

This thesis is based on research conducted with the FPGA Acceleration for Push-Broom Satellite Imagery (Faubai) project. The FAUBAI project is a collaborative effort between NTNU, the University of Oslo (UiO), and S&T Norway, supported by the ESA. S&T specializes in autonomous forest monitoring using EO technology, while the SmallSat team at NTNU contributes expertise in HSI data analysis.

The objective of the Faubai project is to develop a satellite system that can perform semantic segmentation of HSI to identify forest regions and their respective tree types. NTNU and S&T Norway are developing a CNN to run onboard hyperspectral imaging satellites like NTNU's HYPSO-1 or ESA's soon-to-launch CHIME. It is a U-Net designed to classify hyperspectral images in real time, called the Faubai CNN.

Although this model was not available at the time this thesis was written, it is imperative to emphasize the significance of evaluating uncertainty propagation once it is finalized. In this context, the Faubai dataset, comprising PRISMA imagery, SVMS, and 1D CNNs, serves as a valuable resource for investigating uncertainty and risk associated with HSI in the domains of remote sensing and forest management.

2.5 The PRISMA Mission

The Italian Space Agency launched the Hyperspectral Precursor of the Application Mission (PRISMA) satellite on March 22nd, 2019, with a total mass of 830 kilograms. According to Galeazzi et al. (2008*b*), the PRISMA satellite performs EO with the following main objectives:

- In-orbit demonstration and qualification of a state-of-the-art hyperspectral camera;
- Implementation of a pre-operative mission;
- Validation of end-to-end data processing to develop a new application to manage environmental risks and EO based on high spectral resolution images.

The main characteristics of the instruments on board the PRISMA satellite are listed in Table 2.1. It is equipped with a HSI, capable of capturing images across 239 contiguous spectral bands ranging from 400 to 2500 nm. This includes 66 bands within the Visible and Near-Infrared (VNIR) spectrum and 173 bands within the Short-Wave Infrared (SWIR) spectrum, with a spectral resolution of fewer than 12 nm and a spatial resolution of 30 meters. The image scanning system is of the push broom type, as described in subsection 2.1.3 [ASI (2020)].

It can acquire images of areas on demand spanning from 180° West to 180° East longitude and 70° North to 70° South latitude, based on specific locations requested by users, in both a "standard" mode and a "strip" mode. The standard mode results in a 30 x 30 km scene, while the strip mode generates an image with a width of 30 km and a maximum length of 1800 km. Additionally, the platform includes a Payload Data Handling and Transmission Subsystem (PDHT), which provides memory for temporarily storing images and ancillary data and manages data transmission to the dedicated ground segment station.

Parameter	Value
Orbital Altitude Baseline	702 Km
Inclination	98.19°
Orbital period	5927 seconds (about 99 minutes)
Average eclipse	33.9 minutes
Lifetime	5 years
Swath / FOV	30 Km / 2.45°
Ground Sampling Distance (GSD)	Hyperspectral: 30 m, PAN: 5 m
Spatial Pixels	Hyperspectral: 1000, PAN: 6000
Spectral Range	VNIR: 400 – 1010 nm, SWIR: 920 – 2505 nm
Spectral Resolution	≤ 10 nm

Table 2.1: Instrument main characteristics of the PRISMA satellite

All hyperspectral images gathered by the PRISMA satellite are available online [Giardino & Brando (2023)] and can be divided into three different products [Guarini et al. (2018)]:

- Level 0 (L0): Raw data stream including instrument and satellite ancillary data;
- Level 1 (L1): Top Of Atmosphere (TOA) radiometrically and geometrically calibrated hyperspectral and panchromatic radiance images, the one used in this thesis;
- Level 2 (L2): Geolocated and geocoded atmospherically corrected hyperspectral and panchromatic images, atmospheric constituents maps (aerosols, water vapor, thin cloud optical thickness).

In support of the calibration and validation activities for the PRISMA hyperspectral mission, the Italian Space Agency (ASI) and the National Research Council (CRN) initiated the Scientific CAL/VAL of PRISMA mission (PRISCAV) project, aimed at scientifically assessing the performance and durability of the PRISMA payload in various operational scenarios [Genesio et al. (2022)]. The primary objective of PRISCAV is to comprehensively characterize the capabilities of the PRISMA instrument in orbit and verify its sustained performance over time. By 2022, they established a network of instrumented sites representing diverse land use and surface conditions, including snow, sea, inland, coastal water, forests, and croplands. These sites were strategically chosen to enable the collection of independent and traceable in-situ and airborne Fiducial Reference Measurements (FRM) simultaneously with PRISMA acquisitions.

To investigate the quality of the hyperspectral data from the PRISMA satellite, factors such as sensor limitations, atmospheric conditions, noise level, resolution parameters, and validation ensure that the data do not introduce biases or artifacts.

Quality of PRISMA data

PRISMA undergoes several critical processes to ensure data quality and integrity. Sensor calibration is performed using international test sites due to the absence of calibration sites in Italy. Atmospheric correction is applied at Level 2 to compensate for atmospheric effects, with notable improvements in reduced artifacts, although challenges remain at specific wavelengths.

Coherent noise has been observed in PRISMA imagery, requiring further analysis for robust correction algorithms [Genesio et al. (2022)]. In the VNIR range, the SNR is generally more significant in the visible part of the spectrum compared to the Near-Infrared (NIR) region. This is because there is typically more available light in the visible part of the spectrum, which results in a higher signal level[Ewald et al. (2016)].

Radiometric, spectral, spatial, and temporal resolution parameters significantly assess PRISMA's performance, as demonstrated in various studies [Cogliati et al. (2021), Guanter et al. (2021), Romaniello et al. (2020)]. The spectral, radiometric, and spatial performance of PRISMA has shown good agreement with field/airborne spectroscopy, while the capability of detecting and quantifying methane emissions has been established [Cogliati et al. (2021), Romaniello et al. (2020)]. However, some studies have highlighted biases in radiometric performance, indicating the need for continuous monitoring and calibration validation initiatives [Guanter et al. (2021), Romaniello et al. (2020), Genesio et al. (2022)]. Pre-processing steps are crucial in generating accurate and high-quality PRISMA products, addressing challenges such as bad pixels and geometric distortions [ASI (2020)].

Validation experiments comparing PRISMA data with Ground truth measurements have shown good agreement, validating the accuracy of PRISMA's measurements. In remote sensing, ground truths are usually represented as a two-dimensional image that describes the class of a spatial location in the image data. It relates the direct field observations of ground surface phenomena to corresponding features in remotely-sensed data for accuracy verification [Jablonski et al. (2016)]. Notably, comparisons with other sensors have demonstrated reliability in calibration processes and spectral reflectance measurements [Genesio et al. (2022), Heller Pearlshtien et al. (2021), Czapla-Myers et al. (2016)].

2.6 Ground Truth Data

The validation in this thesis is performed by comparing the PRISMA images with data derived from publicly available remote sensing datasets. Norwegian scenes utilize the Skogressurskart (SR16) produced by NIBIO[NIBIO (2022)]. Swedish scenes employ the SLU Forest Map, consisting of a variety of raster maps created by co-processing field inventories from Sweden's National Forest Inventory (SLU), surface models from Lantmäteriet's stereo-matched aerial photographs, and satellite images from Sentinel-2 [SLU (2023)]. The Finnish scenes are based on the Multi-Source National Forest Inventory (MS-NFI) of Finland 2019 from the Natural Resources Institute Finland [Luke (2021)], while Canadian scenes rely on the Tree Species 2019 (TS-19) map by the Canadian Forest Service [nfis (2020)]. They were compiled as part of the Faubai project at NTNU.

Multiple studies have been performed to verify that the ground truths are sufficient to use for further analysis. This is important to validate the model results. This section elaborates on research conducted for the Norwegian, Swedish, Finnish, and Canadian ground truth datasets utilized in this thesis.

2.6.1 Norwegian SR16

The Norwegian Skogressurskart (SR16) is produced by NIBIO [NIBIO (2022)]. The information from the Norwegian Elevation Model [WindPRO (2023)] and Sentinel 2 [ESA (2023)] data are important data sources for production. As illustrated in Figure 2.19, it provides an overview of the distribution and characteristics of Norway's forest resources. It is divided into SR16R, a raster map, and SR16V, a vector map. The dataset is produced through automated processes that combine existing maps (AR5), terrain models, 3D remote sensing data (photogrammetry and laser), and field measurements from the National Forest Inventory [NIBIO (2022)]. Volume, biomass, mean height, tree species, and site index have been estimated using an area-based method.

As earlier explained, the ground truth from the National Forest Inventory is related to the *response* variable in Table 1.1. The method estimates the forest attributes by modeling the relationship between field-based measurements and remote sensing data for each forest attribute. The estimates of forest attributes are then extrapolated to the whole forest area.

Previous research has investigated the reliability of the SR16 map. One study evaluated the accuracy of SR16's tree species predictions in Sør-Troms, comparing over 760 forest management plan stands with SR16 predictions. The study found an overall classification accuracy of 91% at the stand level, with misclassifications occurring more frequently in younger spruce stands, mixed forests, and stands without full crown coverage [Rahlf et al. (2021)]. Another study conducted in Asker, Alver, and Elverum measured 628 control plots in 55 stands and found that SR16 had a relatively low mean error for volume, height, and tree species predictions. How-

Figure 2.19: Illustration showing the SR16 vector (left)- and raster map (right) [NIBIO (2022)].

ever, systematic errors were found in SR16's predictions for stands with high tree densities and in areas with mixed deciduous and coniferous forests. The study also highlighted the potential for significant errors in using photo interpretation for volume estimation in forest management plans. While the accuracy of SR16 may vary locally, these studies suggest that it can be used as a basis for forest management planning with some caution and should be verified through field measurements [Breidenbach, Rahlf, Räty, Hauglin & Bergseng (2022)].

2.6.2 Swedish SLU

Swedish scenes employ the SLU Forest Map consists of a variety of raster maps created by co-processing field inventories from SLU, surface models from Lantmäteriet's stereo-matched aerial photographs, and satellite images from Sentinel-2 [SLU (2023)]. The volume maps contain the timber resources categorized by pine, spruce, beech, oak, birch, and other deciduous trees. Additionally, they provide estimates of basal area-weighted mean tree height, basal area-weighted mean stem diameter, stand age, as well as species-specific stem volume for the years 2000, 2005, 2010, and 2015 [Wallerman et al. (2021)]. Specifically, the biomass volume data is present for multiple tree types.

The accuracy of the SLU Forest Map has been extensively verified, focusing on forest variables such as tree canopy height, diameter at breast height, basal area, and stem volume. Results have shown that Lorey's height was the most accurately estimated variable, followed by diameter and stem volume, while basal area was the least accurate. Stand age was excluded from the 2015 map due to inadequate quality for most applications. Accuracies were similar for the years 2000-2010 but improved considerably for the year 2015 when tree canopy height data was incorporated, indicating the predictive power of these data. Errors were significant in southern regions compared to the north due to the more dense and large forest in the south. Adding canopy height data eliminated systematic errors for forest stand age classes, significantly improving most applications [Wallerman et al. (2021)].

The SLU Forest Map is the only publicly available map providing time series of continuous estimates of forest variables in Sweden, including species-specific stem volume. The quality of the latest map (the year 2015) is superior to earlier maps in terms of RMSE as well as systematic errors, making it a valuable resource for a wide range of applications [Wallerman et al. (2021)].

2.6.3 Finish MS-NFI

The Finnish scenes are based on the Multi-Source National Forest Inventory (MS-NFI) of Finland 2019 from the Natural Resources Institute Finland, which uses remote sensing and numerical map data to predict forest variables, such as the volume of trees, in Finland. Additionally, MS-NFI incorporates a diverse range of tree species, including but not limited to spruce, pine, and various broad-leaved trees such as birch and other unspecified broad-leaved species[Luke (2021)].

The main source of HSI data is medium-resolution satellite images, e.g., data from the Sentinel-2A/B satellites of the European Space Agency (ESA) and the Landsat 8 satellite of the United States Geological Survey (USGS). The data for training these predictions come from field plots in the Finnish National Forest Inventory.

Katila (2004) primarily targets evaluating and minimizing errors inherent in the MS-NFI and the k-nearest neighbor method (k-NN). The core sources of error are recognized as the field sample's representativeness regarding the estimation problem, the restricted dynamic range of spectral channel values on forestry land in high-resolution optical satellite data, the comparatively small size of NFI field plots relative to the pixel size in image data, and local inaccuracies within both image and field plot data.

Application of parameter selection criteria yielded comparable parameters across the four distinct study areas, each representing different geographical regions of Finland. However, it should be noted that significant systematic errors in small-area estimates of particular subregions were not eliminated solely by adjusting estimation parameters [Katila (2004)].

2.6.4 Canadian TS-19

Canadian scenes rely on the Tree Species 2019 (TS-19) map by the Canadian Forest Service, which consists of the tree species' presence in Canada's forested ecosystems. Products include the leading tree species and the class membership probabilities for 37 tree species [nfis (2020)].

The map layers were produced using remote sensing (Landsat imagery) and spatial modeling, in which the common leading tree species nationally were black spruce, trembling aspen, and lodgepole pine. Accuracy assessment was conducted using independent validation data also drawn from the National Forest Inventory (NFI)) indicating an overall accuracy of

93.1%Hermosilla et al. (2022). The predictor variables informing on geographic, climatic, and topographic conditions were most important to the classification models.

2.7 SVM Design and Training Approach

The SVM training for the forest classification was performed in two ways, using a single image and the full Faubai data set. Both follow an iterative approach and initializes an SVM classifier using the SGD optimizer algorithm described in section 2.7. The SGD optimizer updates the coefficients incrementally. This allows for specifying the maximum number of iterations, the stopping criterion, the learning rate schedule, and the initial learning rate.

The SVM training process was initiated by loading images alongside their corresponding labels. Following this, the VNIR component was extracted and subsequently reshaped, which entailed the removal of the first and last three bands. Incremental learning was achieved by applying the 'partial fit' function. This procedure ensures that the model adapts continuously as new data is ingested, enabling subtle modifications to the model's coefficients in response to new data over time.

2.7.1 SVM Single (SVM-S)

The training and validation of the single image based SVM (SVM-S) model use the image shown in Figure 3.1 as input. The image was split into two parts, in which half of the image was used for training, while the entire image was used for validation to assess the classifier's performance. During training, the classifier's hyperparameters were adjusted.

A stable equilibrium was never found, even though different convergence rates were tested. During the training phase, the classifier achieved an accuracy of 62.5%, the highest observed during the experiment. Subsequently, the coefficients of this model were retained for future reference and analysis. With this achieved accuracy, the model can be utilized to analyze and characterize the impact of noise on its performance. The model's label, predictions, and errors are illustrated in Figure 2.21b along with the corresponding confusion matrix given in Figure 2.20a.

2.7.2 SVM Full Faubai (SVM-F)

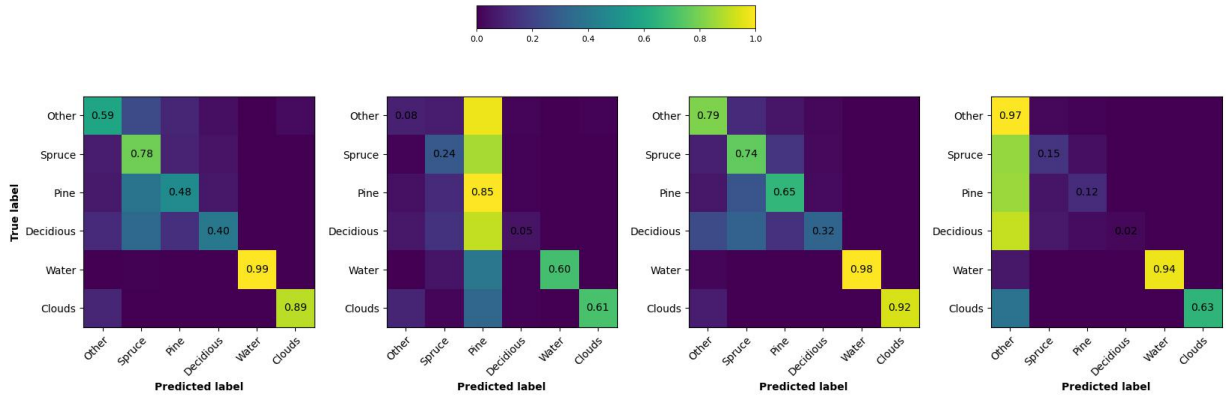
The model prediction for the SVM when training on the full Faubai dataset (SVM-F) can be seen in Figure 2.21c along with the confusion matrix given in Figure 2.20b. Specifically, the classification performance declines relative to the SVM-S when evaluated on that image with a classification accuracy of 32%. Thus, incorporating the full Faubai dataset into the analysis revealed a significant degradation in classification accuracy by 48.8%.

The model was trained on multiple batches of images. The initial training procedure for the first batch was conducted sequentially, beginning with a coarse training phase that involved a batch size of five images. This was followed by a fine-tuning phase, which expanded the batch size to fifteen images.

Subsequent models were trained with an incremented batch size, consistently set at fifteen images. For each batch, VNIR data and labels are extracted. Subsequently, the training proceeded by iterating over subsets of the data. After each iteration, the model is fit to the current subset. Once the training is completed for all images in the batch, the process moves on to the next batch, continuing until the epoch is complete. The model's performance is assessed periodically to measure its progress and effectiveness in classifying the data.

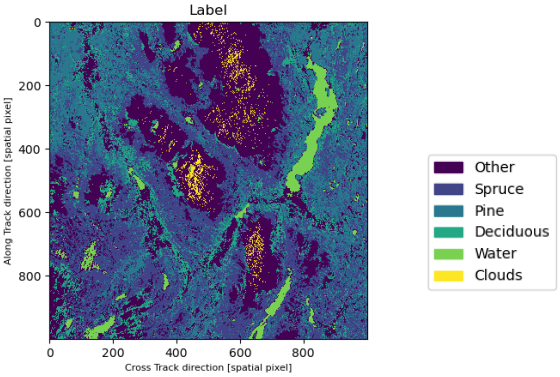
As earlier mentioned, SGD optimizer is used to update the coefficients incrementally. This quality is notably beneficial when dealing with continuously generated data, as it circumvents the necessity of simultaneously loading the entire dataset into memory.

Such an approach is especially advantageous when the classifier is tasked with training on extensive datasets, such as the complete FauBai dataset. By employing SGD, the process can efficiently handle substantial data volumes, reducing memory requirements while maintaining computational effectiveness.

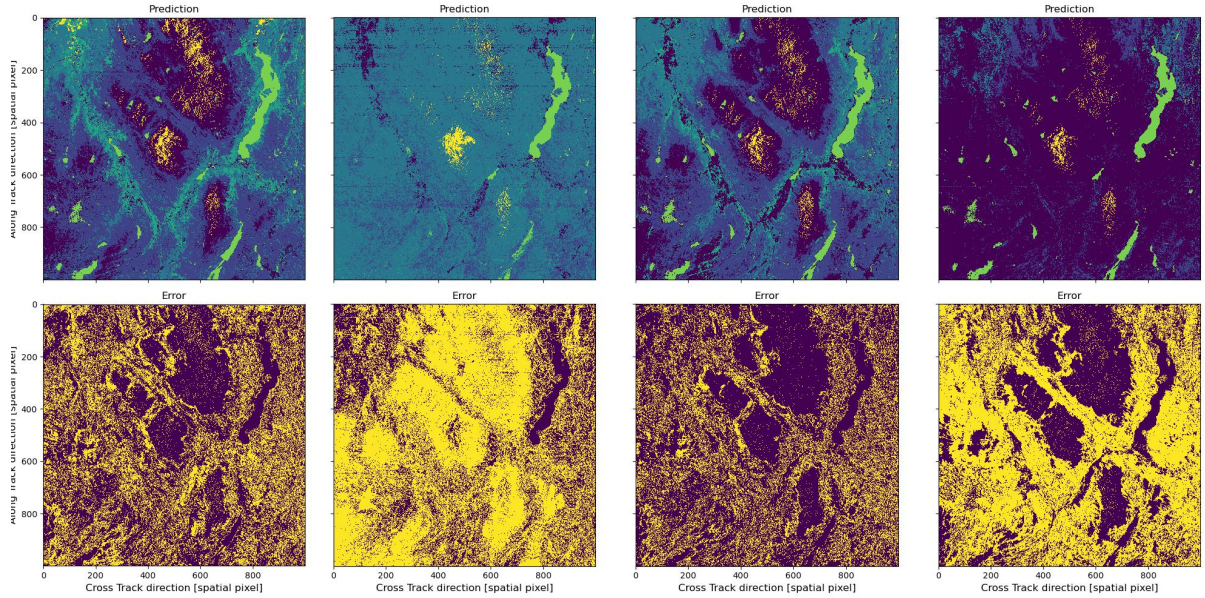


(a) SVM-S (b) SVM-F (c) CNN-S (d) CNN-F

Figure 2.20: Normalized confusion matrices of SVM and CNN of HSI image given in Figure 2.21, for single image.



(a) Label



(b) SVM-S (c) SVM-F (d) CNN-S (e) CNN-F

Figure 2.21: SVM and CNN classification by models for the image used to train the SVM-S and CNN-S. For the error plot, the yellow indicates misclassified pixels.

2.8 CNN Design and Training Approach

The CNN utilized in this thesis is derived from the 1D CNN architecture described in Section 2.2.4. The decision to utilize the 1D CNN was motivated by the fact that the Faubai CNN, initially intended for use, was not yet complete. Nonetheless, it was anticipated that the LucasCoord-Conv architecture (presented in section 2.2.4) would be a suitable alternative, sensitive to the same sources of uncertainties as a 2D CNN.

The training process begins by preparing the environment and data. A CNN model is then created using a custom function designed for multiclass classification, as indicated by the six categories. The model is compiled with the categorical cross-entropy loss function, the Adam optimizer, and the categorical accuracy metric. The training settings include using the early stopping callback to halt training when improvement ceases, thereby preventing overfitting. These callbacks monitor the validation loss during training.

2.8.1 CNN Single (CNN-S)

Similarly to the single image trained SVM, the single image trained CNN (CNN-S) is trained using half of the image and validated using the whole image. The results can be seen in Figure 2.21d and 2.20c, giving an accuracy of 71%.

2.8.2 CNN Full Faubai (CNN-F)

The 1D CNN model trained on the full Faubai set is called the CNN-F. Similarly to the SVM, the CNN trained on the full training set gives a degraded prediction accuracy for the given image. The results show the corresponding prediction, error, and confusion matrix in Figure 2.21e and 2.20d. The CNN trained on the full Faubai dataset have an overall accuracy of 41%, meaning that the accuracy is degraded by 42.2% from the single image trained model.

The model is trained using the fit method on the training data and labels. The training files are processed in a shuffled order over multiple epochs. The training data is divided into batches for each epoch containing multiple images. Each batch is processed separately. The process involves loading the image data and labels and conducting multiple training iterations on subsets of this data. After each batch, the model is evaluated on a validation set. This method exposes the model to a wider variety of images in each batch and epoch, potentially helping the model to better generalize to new data. The process continues until all epochs are complete.

Furthermore, the same data is used for validation which could be improved in a later adjustment. The model is then utilized to predict the classes of the test data.

Chapter 3

Dataset

This chapter provides a comprehensive overview of the Faubai dataset used for training and testing the four distinctive AI models presented in subsection 2.2.1 and 2.8. It offers a detailed exposition of the hyperspectral dataset and a comprehensive presentation of the ground truth labels.

3.1 Characterization of Complete Dataset

The data used in this thesis have multiple formats. For each PRISMA capture, the available data to be analyzed are as follows:

- A .h5 file with the HSI PRISMA imagery
- A .csv file comprising the labels assigned to each pixel
- A .jpg file depicting the labels as color overlays on a grayscale image
- A .npy file containing the latitude and longitude coordinates for each pixel
- A .tif file describing the tree age, from SR16 dataset.

All files have a naming scheme that adheres to the original titles of the HSI images. Furthermore, the names incorporate a single character indicating their originating dataset ('S' for Sweden, 'C' for Canada, 'F' for Finland, and no additional character for Norway).

3.1.1 Classes

The FAUBAI dataset consists of labels of different tree types georeferenced to align with the pixels in several hyperspectral remote sensing images. It is used to train and test models that classify the pixels in images of the same kind as those in the dataset as one of a selection of the labels:

(1) *Spruce*, (2) *Pine*, (3)*Deciduous*, (4) *Water*, and (5) *Clouds*, also incorporating snow. Figure 3.1 shows the labels as color overlays on a monochrome image. In addition, a class that indicates unlabelled data, named (0) *Other*, includes farms, fields, roads, cities, and barren areas, among other minor categories. The distribution of labels in training- and testing datasets can be seen in Figure 3.2 and 3.3.

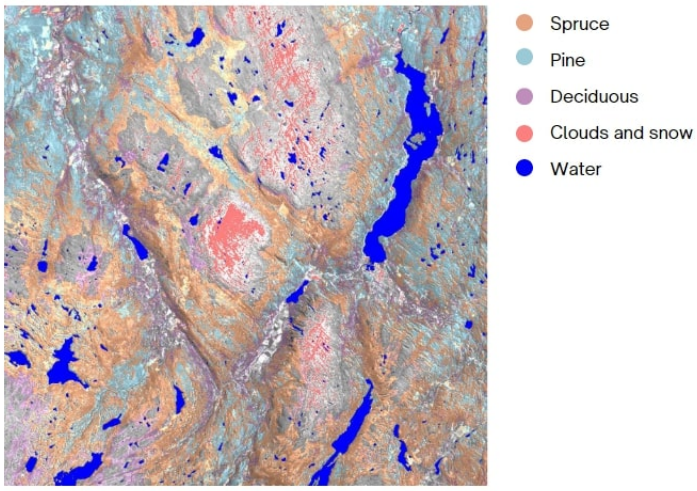


Figure 3.1: Illustration showing the labels as color overlays on a monochrome image.

3.2 Hyperspectral Data Set

An overview of the hyperspectral data used in this thesis is presented Figure 3.4. All data is collected from the PRISMA satellite and divided between spring, summer, and fall. The intended data to be labeled are the TOA L1 PRISMA data product, as outlined in section 2.5.

3.2.1 Composition of HSI dataset

The composition of HSI data was chosen for several reasons. Hyperspectral images were collected by the PRISMA hyperspectral imaging satellite due to their usability. Additionally, the images were selected to have similar distributions across dimensions.

PRISMA data

First, there is a web portal where users can request data to be collected at specific locations. This has enabled a large and representative amount of training data to be collected. The PRISMA satellite can provide high spatial-resolution images of the Earth's surface, allowing for detecting and analyzing materials and objects based on their unique spectral signatures, such as forest

monitoring. Additionally, it closely approximates the upcoming ESAs CHIME mission launch. The similarity makes the collected data valuable for testing the robustness since the PRISMA data have shown a relatively high number of artifacts (outlined in section 2.5).

Spectral Ranges

HSI PRISMA data consists of two wavelength regions, the VNIR and SWIR. As presented in section 2.5, the SNR of the VNIR range is generally larger in the visible part of the spectrum due to more light. This higher SNR makes differentiating signal and noise easier. Thus, the SVM analysis concentrates solely on the VNIR spectrum. This approach simplifies both the classification tasks and the addition of camera artifacts.

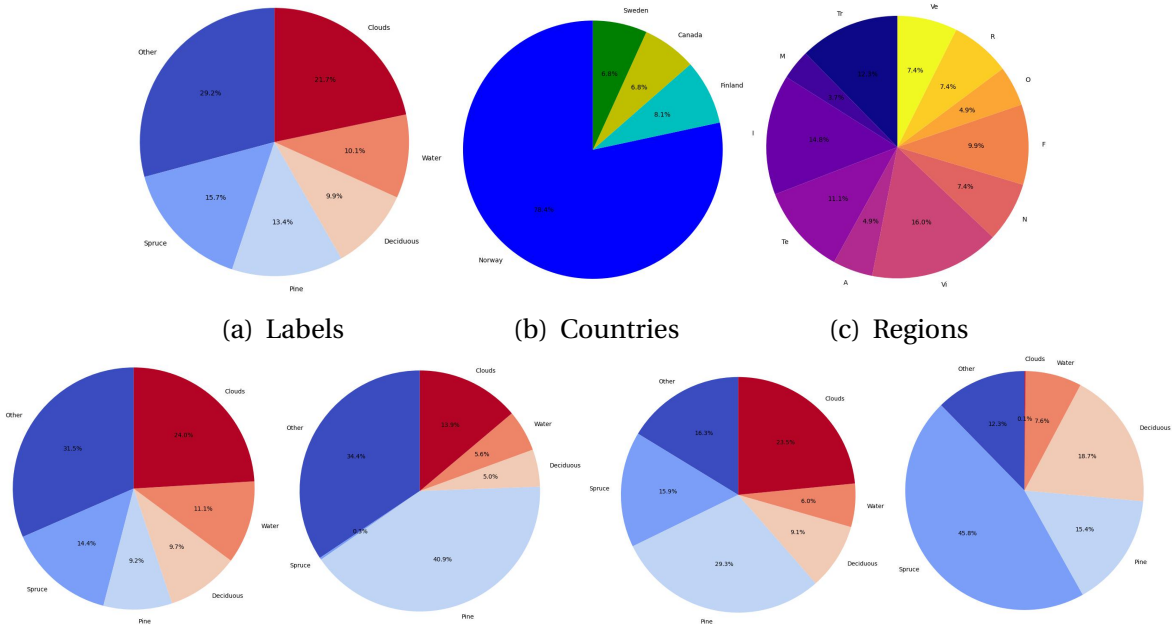
Dataset Selection

As mentioned in Table 1.2, the dataset distribution must be assessed to examine the model *response*. Attaining uniform distributions across labels, countries, and regions is imperative for mitigating potential sources of error in the model. One such error is overfitting, where the model demonstrates superior performance on particular data yet underperforms when presented with novel or unseen samples. A uniform distribution aids in circumventing this risk, thus enhancing the model's robustness and generalizability.

In addition, it fosters fairness in model predictions, advocating for consistent performance across various categories and geographical areas. This is paramount in preventing potential distortions or biases in the model that could lead to erroneous conclusions.

Furthermore, the Faubai datasets were meticulously chosen to guarantee similar distributions across several dimensions, such as labels, countries, and regions. The map of PRISMA captures are presented in Figure 3.5 along with the corresponding pie charts illustrating the distributions in Figure 3.3 and 3.2. These show that all classes are well represented in both the training and testing set, with the smallest class being *Decidious* (9.7%) in the training set and *Water* (7.9%) in the test set. The regions are specifically denoted for the Norwegian images, encompassing the following areas: Trøndelag (Tr), Møre and Romsdal (M), Rogaland (R), Finnmark (F), Innlandet (I), Nordland (N), Oslo (O), Vestland (Ve), Viken (Vi), Telemark (Te), and Agder (A). Due to the small set of test images, not all regions are present.

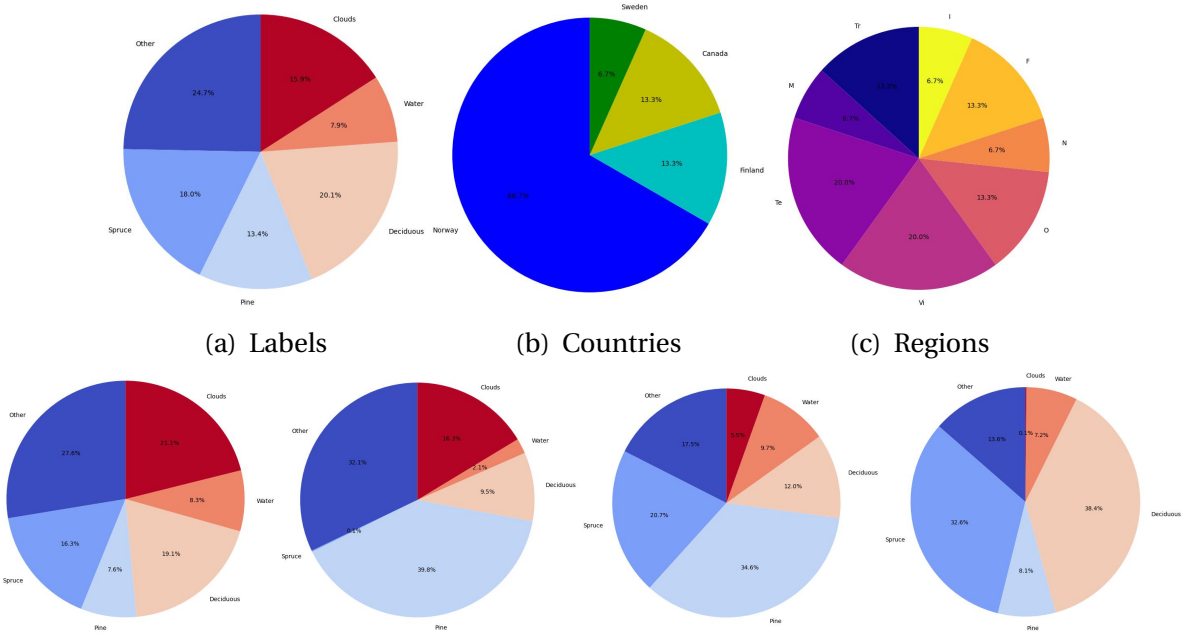
The label distribution within countries in Figure 3.3 and 3.2 *(d-g)* shows that the distribution of labels within each country is not homogenous. For instance, in the case of Canada, there are markedly fewer instances of the label *Clouds* compared to other countries.



(a) Labels (b) Countries (c) Regions

(d) Norway (e) Sweden (f) Finland (g) Canada

Figure 3.2: Distribution of training dataset.



(a) Labels (b) Countries (c) Regions

(d) Norway (e) Sweden (f) Finland (g) Canada

Figure 3.3: Distribution of testing dataset.

		TRAINING SET	TEST SET
PRISMA	NORWAY	58	9
	SWEDEN	6	2
	CANADA	5	2
	FINLAND	5	2
	TOTAL	74	15

Figure 3.4: Description of the FAUBAI data set.

(a) Norwegian, Swedish and Finish captures

(b) Canadian captures

Figure 3.5: Map of PRISMA captures, visualized in QGIS [QGIS (2023)].

3.3 Ground Truth Dataset

In the Faubai dataset, the ground truth labels are re-sampled to match the resolution and coverage of their hyperspectral image. To automate the georeferencing procedure, the Scale-Invariant Feature Transform (SIFT) algorithm is applied in conjunction with a reference map, roughly following the procedure of [Long et al. (2016)]. Each pixel's latitude and longitude coordinates are employed to obtain ground class labels from the respective national datasets. This technique ensures the accurate geolocation of each pixel, facilitating the reliable and precise classification of ground features present in the hyperspectral images.

Thus, the tree type can be returned to the image grid by summing up the corresponding labels from the database. This allows for creating a ground truth map at the same spatial resolution as the hyperspectral image, dividing the forest into (1) *Spruce*, (2) *Pine*, and (3) *Deciduous*.

The (4) *Cloud* and (5) *Water* labels, on the other hand, are generated through pixel value thresholds and are intended to be used only as markers to prevent these objects from accidentally being labeled as a tree type.

The datasets used to generate labels for HSI data are collected from Norway, Sweden, Finland, and Canada. These were thoroughly introduced in section 2.6. The re-sampling of ground truth labels of the different sources can be summarised by:

- **Norwegian SR16:** Tree species layers corresponding to three classes specified directly by SR-16.
- **Swedish SLU and Finish MS-NFI:** Ground labels are inferred according to which tree type has the greatest volume within a given pixel.
- **Canadian TS-19:** The tree species were grouped into categories approximating those in the other maps *(Spruce, Pine, and Deciduous)*.

Chapter 4

Imaging Artifacts and Model Response

While the conventional scientific practice typically involves the elimination of artifacts, the focus of this chapter diverges by intentionally incorporating such elements. Our aim is to examine and understand how different forms of uncertainty appear. The chapter initiates by detailing the characteristics of artifacts and elucidating the methodology employed to incorporate them. Moreover, it serves as a part of the assessment of the *explanatory variables*, expounded in Table 1.1, which suggested "analyzing the impact of noise or measurement errors on model performance by comparing the model's performance on a dataset tainted by artifacts."

4.1 Method for Adding Artifacts

The method used to apply the artifacts to each HSI image is visualized in Figure 4.1, and can be summarized to:

1. Acquire image cube;
2. Add artifact;
    - Spatial dimension: Add noise (*Gaussian*, *Poisson*, *Salt and Pepper*, *Stripe*) or distortions (*Smile*, *Keystone*), setting the SNR/pixel-shift.
    - Spectral dimension: Add blur (*WDD*), setting the maximum pixel blur.
    - Classify pixels with ML model
3. Construct degraded image cube.
4. Classify pixels with ML model.

This experimental framework facilitates an evaluation of the model's capacity to classify hyperspectral images affected by a diverse range of artifacts.

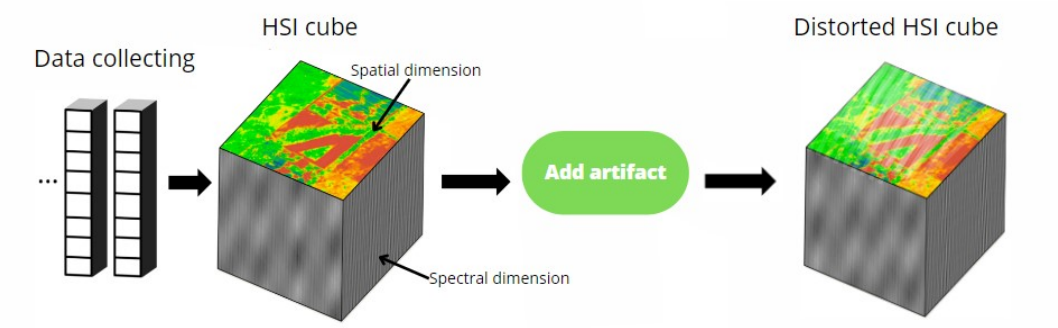


Figure 4.1: Illustrated procedure of adding artifacts to HSI.

4.2 Artifact Characterization

The task of modeling and adding noise, distortion, or blur to a HSI image can be challenging due to the complex nature of the data and the underlying physics of the HSI system. However, accurate modeling and method for applying artifacts can be crucial for improving the robustness and accuracy of ML models using HSI data.

As earlier mentioned, the hyperspectral data used are the Level 1 (L1) products from the PRISMA satellite's VNIR hyperspectral sensor, described in Table 2.5. This product contains calibrated and georeferenced hyperspectral data in radiance units. For the models used in this analysis, the noise is directly applied to the radiance version of the data.

SNR, as described in subsection 2.3.2, maximum pixel-shift (a measure of spatial distortion), and maximum blur (measured in pixels) are used as measurements to control the level of artifacts applied to the dataset.

4.2.1 Pixel Uncertainty

Pixel uncertainty includes *gaussian-*, *poisson-*, and *salt and pepper* noise. The best way to model and apply noise depends on the specific application and the type of image being processed. The approach used in this thesis involves explicitly modeling the different noise types before adding them as a noise layer to the HSI data cube. An illustration of the original VNIR data and the noisy VNIR data is given in Figure 4.2.

The *gaussian-* and *poisson* noise look similar in appearance because both types of noise as random variations of the image. From theory, it is known that *poisson* noise is correlated with the intensity of each pixel. In contrast, *gaussian* noise is independent of the original intensities in the image [Talbot et al. (2009)]. If closely studied, multiple things can be observed in Figure 4.2 and 4.2. The *gaussian* noise has a continuous intensity variation around the original signal since the whole image is affected similarly. *Poisson noise* is often seen as a "grainy" or

"speckled" pattern of random variations around the original signal. Additionally, it is seen to be more pronounced in areas of the image with high signal intensity, as the magnitude of *poisson* noise is proportional to the square root of the signal intensity. In contrast, *gaussian* noise has a constant magnitude and is not affected by the signal intensity.

On the other hand, the impulsive *salt and pepper* noise appears as light or dark pixels randomly scattered throughout the image. This is generally caused by a defective camera sensor or software- or hardware failure in image capturing or transmission. Due to this, only a portion of the image pixels is corrupted, whereas other pixels are non-noisy [van Beers & Kleijnen (2003)]. As seen in Figure 4.2, it is evident that the added light pixels are visually more discernible than the other noise types. This increased visibility can be attributed to the contrast between these light pixels and their surrounding environment.

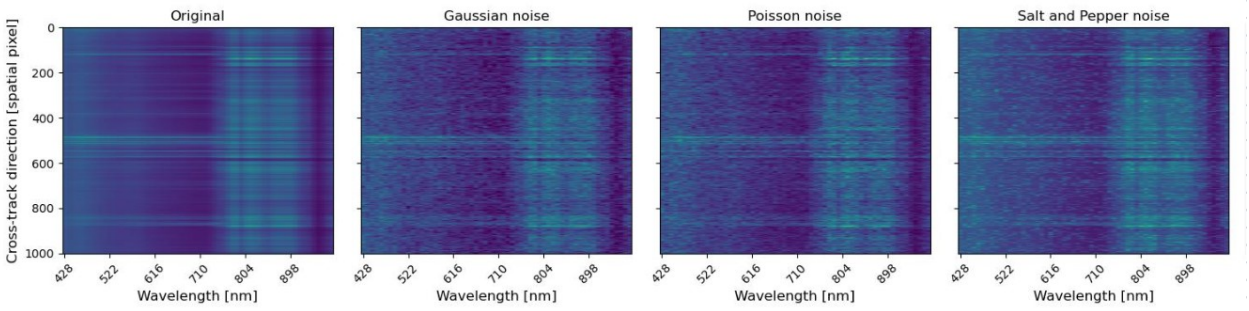


Figure 4.2: Pixel noise, illustrated for band one in the spectrogram.

The true color image of the *gaussian-*, *poisson-*, and *salt and pepper* noise can be seen in Figure 4.6. The reason why *gaussian* noise appears to be more prominent than other noise types could be due to its nature and how it affects the image. It uses normal distribution, adding intensity values to the image distributed around a mean value with a certain standard deviation. This distribution could result in more noticeable changes in the image than other noise types like the signal-dependent *poisson* noise.

4.2.2 Spectral Uncertainty

The phenomenon of patterned *stripe* noise is a prevalent occurrence within HSI. This noise can be simulated by introducing a constant pixel error, such as *gaussian* noise, to distinct rows or columns within the image. In the context of this thesis, stripe noise is emulated through applying uniform *gaussian* noise to the along-track dimension within a HSI framework.

When the *gaussian* noise is maintained as a constant, the mean and standard deviation of the noise remain identical across all pixels in the along-track dimension. This consistency suggests that the noise exhibits a uniform pattern, influencing each row or column of the image

similarly. Consequently, stripes are formed in the spatial spectra, as depicted in Figure 4.4.

Conversely, *smile* and *keystone* are not modeled as additional filters. *Smile* refers to the spatial variation in the center wavelength of the spectral bands across the field of view. In contrast, *keystone* describes the misalignment of the spectral bands in the across-track direction. Thus, this distortion requires shifting the pixels in a specific way to simulate the respective geometrical distortion.

To model *smile*, the pixels along the spectral bands are shifted to follow a curve along the horizontal and vertical directions. The complete code is given in section A.1. First, the desired maximum pixel-shift in the x-axis is given as an input parameter, and mesh grids for the image's x and y coordinates are obtained. Furthermore, a distortion function combines the smile effect in the vertical direction and a frown effect in the horizontal direction before it is applied to the mesh grid. The new coordinates are clipped to ensure they fall within the bounds of the original image dimensions. The pixel-distorted image is finally created by applying the calculated x and y coordinate mappings to the input image, using cubic interpolation to produce the output image with the *smile* distortion effect. The result can be seen in Figure 4.3.

When introducing *keystone* distortion effect, the goal is to achieve a trapezoidal shape caused by the misalignment of pixels across different spectral bands. Similarly to modeling *smile*, an input coefficient that controls the strength of the *keystone* is used as input, and a mesh grid are obtained. A distortion function represents the linear relationship between an input coordinate and the *keystone* effect. Furthermore, the *keystone* distorted mesh grid is calculated by applying the distortion function to the difference between the mesh grid's x and y coordinates and the center of the image. The final stages of obtaining the image are identical to *smile*, giving the result seen in Figure 4.3. The code for applying *keystone* to a HSI image cube is given in section A.2.

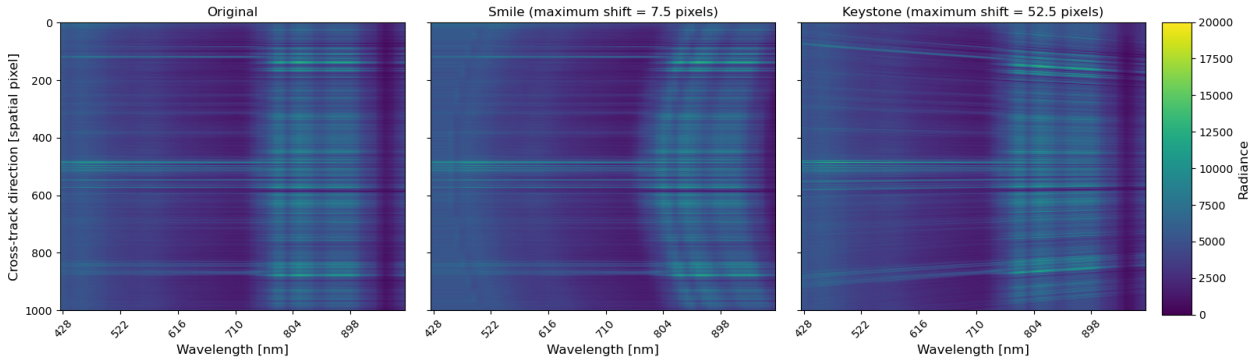


Figure 4.3: Original (left); *smile*- (center); and *keystone* distortion (right), for band zero in the spectrogram.

4.2.3 Spatial Uncertainty

WDD is modeled by taking an HSI image cube and a maximum blur value (in pixels) as inputs. It applies a gaussian filter with a standard deviation that varies for each spectral band. It has a more significant standard deviation to the ends of the spectral range to achieve a stronger blur effect.

Having the most blurring at the ends of the spectral range is reasonable since cameras in practical applications often exhibit such behavior. In general, both second-order and sigmoid functions can be used to control the amount of blur for each band, and the choice between the two depends on the desired shape of the blur effect.

A study by Torkildsen & Skauli (2018) showed that the second-order polynomials should model the distortion. Thus, blur is added as shown in Figure 4.5, showing a stronger intensity for the wavelength extremities in the data cube.

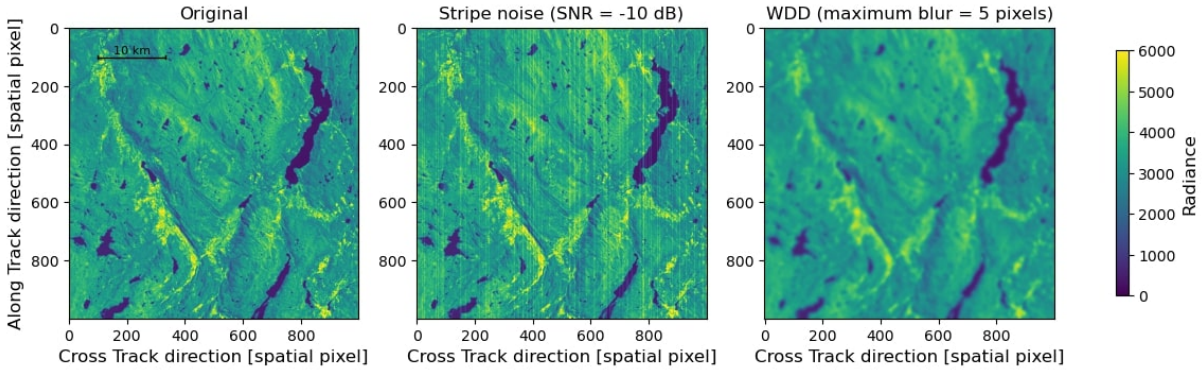


Figure 4.4: Illustration of (left) Original; (center) Stripe noise; (right) Wavelength dependent defocusing (WDD), for band zero in the spatial domain.

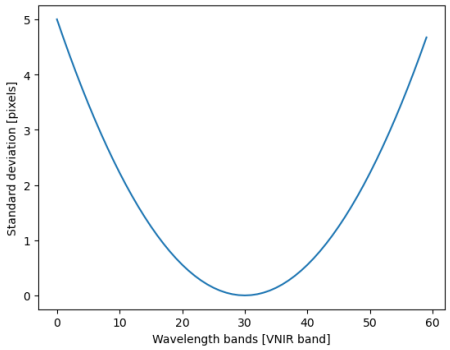


Figure 4.5: Relationship between the standard deviations and wavelength bands when applying Wavelength-dependent defocusing (Maximum blur = 5 pixels) to a PRISMA image.

4.3 Model Response to Image Artifacts

Classification accuracy can be used to evaluate the performance of the classifiers trained on a single image (SVM-S, CNN-S) for different noise levels. Figure 4.7, 4.9 and 4.11a show the SVM-S accuracy, while Figure 4.8, 4.10 and 4.11b show the CNN-S accuracy of the different classes in different colors. The overall prediction accuracy is presented in black.

An example of the prediction and error by the classifiers is presented earlier, in Figure 2.21, corresponding to the normalized confusion matrices in Figure 2.20. The normalized confusion matrix provides the proportion of correct and incorrect classifications for each class, giving a realistic view of the classifier's performance and allowing for easier comparison across different classes. Looking at both Figure 2.20a and 2.20c, the proportion of correct classification has the highest results for *Water* and *Clouds*. There is a higher accuracy for *Spruce* and *Pine* than *Decidious*. The CNN-S has better results for classifying *Pine*, while the SVM-S performs better for *Decidious*.

The normalized class accuracies are used for the following analysis. Due to the different kinds of noise, distortions, and blur, the accuracy is compared to the corresponding artifact measurement. The noise types are compared towards varying SNR, distortion towards increasing maximum pixel-shift, and the WDD in response to elevated levels of pixel defocusing.

4.3.1 Model Response to Noise

Figure 4.7 and 4.8 depicts the accuracy of a SVM-S and CNN-S classifier under the influence of four different types of noise: *gaussian, poisson, salt and pepper, and stripe*. The SNR range was chosen to span the full accuracy range.

As evident from the Figure 4.7 and 4.8, the noise types exhibit similar trends where the SVM-S and CNN-S accuracy increases with an increase in SNR. This observation is attributed to the fact that noisy input data can lead to misclassification and hinder the SVM's ability to locate a suitable hyperplane. Thus, when the SNR is low, the SVM-S may struggle to distinguish between different classes, resulting in lower accuracy.

Additionally, the precision of the *stripe noise* yielded fragmented outcomes, potentially attributable to its recurring configurations and distinct noise features. The linear patterning and striations generated by the noise may induce discontinuities in classification outcomes, thereby challenging the classifier's capacity to differentiate between actual class boundaries and the artificial boundaries introduced by the noise.

A common characteristic across all noise types in the models is that the accuracy within the SNR range of 10 dB to -10 dB has the most profound decline in accuracy. As previously highlighted, both *Water* and *Clouds* exhibit high prediction precision across all forms of noise, particularly

when the SNR is greater than -10 dB. This suggests that the spectral characteristics and patterns associated with *Water* and *Clouds* are distinct and consistent, even in the presence of noise.

On the other hand, both models misclassify classes when the SNR decreases. The SNR seems to classify species as *Clouds*, shown by its high accuracy in Figure 4.7. However, this is not as profound as the similar misclassification in the CNN-S, in which the accuracy of *Other* increases.

Additionally, it is evident that the accuracy associated with tree-type classification significantly declines when the SNR descends below zero. This highlights the sensitivity of these classes to noise interference. Trees have more diverse spectral signatures, and their classification relies on capturing fine details and subtle variations in reflectance patterns. As the noise level increases, it becomes more challenging to accurately distinguish these subtle differences, resulting in decreased classification accuracy.

On the other hand, the classification of *Deciduous* trees seems less susceptible to noise for the CNN-S, only leading to misclassification as *Other* at negative extremes. This disparity in performance can be due to the *Descidious* class being easier to predict in noise conditions, considering the diversity of tree species it encapsulates. This arises from the variability in form, size, and other morphological characteristics among different deciduous species, which significantly broadens the feature space the classification model can look for. In contrast, species such as spruce and pine, which belong to the coniferous category, exhibit more uniform characteristics. They get noticeably degraded at a SNR equal to zero.

The marginally superior performance of the SVM-S in handling noise can be attributed to its ability to define a decision boundary based on support vectors, which allows for better handling of variations and noise in the input. In contrast, although the CNN-S possesses the advantage of learning hierarchical and spatial features directly from the data, it may struggle in scenarios with noisy data due to potential overfitting or misinterpretation of noise as significant features. The emphasis on capturing local patterns and spatial dependencies in the input can render the CNN-S more susceptible to noise-related variations, resulting in a relatively inferior performance compared to the SVM-S in the presence of noise.

4.3.2 Model Response to Distortions

The prediction accuracy when applying *smile* and *keystone* distortion is illustrated as a function of maximum pixel-shift in Figure 4.9 and 4.10.

The findings suggest that with the augmentation of pixel-shift, there is an observable near-linear decrement in prediction accuracy for all classes under the influence of the *keystone* distortion, in which the CNN-S demonstrates a diminished resilience against the distortion, as evidenced by the comparatively more pronounced reduction in accuracy relative to the SVM-S. This trend is uniformly evident across all classes, except for the *Clouds* class when analyzed using the SVM-S and the *Other* class under the CNN-S model. The accurate predictions of *Clouds* can be attributed to the inherent nature of clouds, which are amorphous and lack well-defined shapes or rigid structures. However, these exceptions also hint towards the persistent misclassification into the *Clouds* and *Other* classes upon introducing distortion. Specifically, the behavior of the *Other* label in the CNN-S might be due to the classifier struggling to detect the simple features corresponding to the different tree labels, predicting it to be something else.

In contrast, the CNN-S is very sensitive to *smile* distortion, with a varied response across different classes. For tree types, the prediction accuracy rapidly declines, reaching zero within a pixel-shift of five. Specifically, *Spruce*, *Pine*, and *Deciduous* accuracy is degraded 84%, 98%, and 76% within a pixel.

The rapid decline is also the case for the SVM-S but in a smaller amount than for the CNN-S. The accuracy of *Spruce* and *Pine* is degraded 26%, and 100%, respectively, within the first pixel-shift. Interestingly, the accuracy for the *Deciduous* class appears to increase for small and large pixel amounts. For the CNN-S, this happens in a smaller amount in the *Spruce* class for higher pixel-shift amounts. These counter-intuitive behaviors are most plausibly due to erroneous classifications rather than an improvement in model performance.

The algorithms' rapid decline in tree species may be due to their struggle to recognize their complex characteristics when subjected to shifts of narrow spectral features. The subsequent increase in accuracy for *Deciduous* and *Spruce* could be a result of the algorithm incorrectly identifying other objects as *Deciduous* trees when the pixel-shift is high.

The significantly worse results for the *smile* distortion, as compared to the *keystone* distortion, can be attributed to the unique characteristics of the *smile* distortion and its effects on the images being analyzed. The *smile* distortion introduces radial or bending distortions to the image, which can cause images to appear warped or curved. This distortion can substantially alter the spatial relationships between different image regions and lead to a more severe loss of spatial information than the geometric transformations introduced by the *keystone* distortion. As a result, the SVM-S classifier might struggle to find the optimal hyperplane, leading to decreased

prediction accuracy, while the CNN-S might struggle to find patterns.

4.3.3 Model Response to Defocusing

Figure 4.11 presents the accuracy of the SVM-S and CNN-S classifiers as a function of the maximum defocus (in pixels) applied to a wavelength in a HSI datacube. The first two pixel decreases are the most important for accuracy in both classifiers.

The results indicate that the classification accuracy for the *Water* and *Spruce* classes significantly decreases as the defocusing increases in the SVM-S. Additionally, the accuracy for the *Pine, Deciduous* classes slightly increases with increasing defocusing. This may result from the inherent overlap in the spectral features of these classes. Defocusing can introduce some level of spectral smoothing, which may, in some cases, reduce the overlap between the spectral features of these classes and make them more distinguishable in the feature space. Consequently, the SVM-S classifier might become better at separating the *Pine* and *Deciduous* classes, leading to the observed improvement in classification accuracy. Furthermore, the *Cloud* class in the SVM-S classifier and the *Other* class in the CNN-S classifier exhibit increased accuracy. This increase can be attributed to the misclassification observed for the other artifacts.

A noteworthy observation is a significant degradation in the accuracy of the *Water* class, considering its demonstrated resilience towards other types of artifacts. This could be because the *Water* class possesses distinct and recognizable spectral features sensitive to defocusing. As defocusing increases, the spectral features may become less distinguishable, making it more challenging for the classifiers to accurately separate these classes in the feature space.

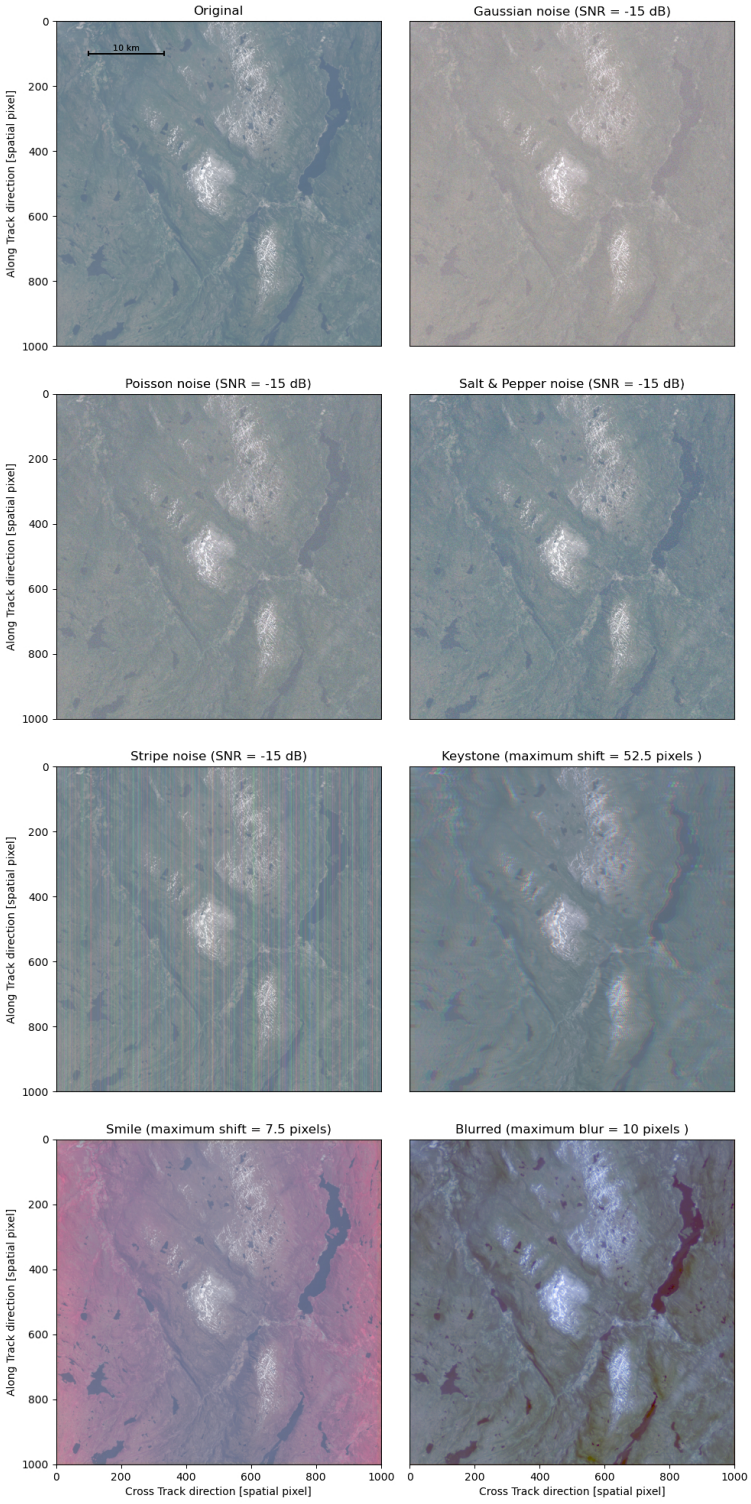
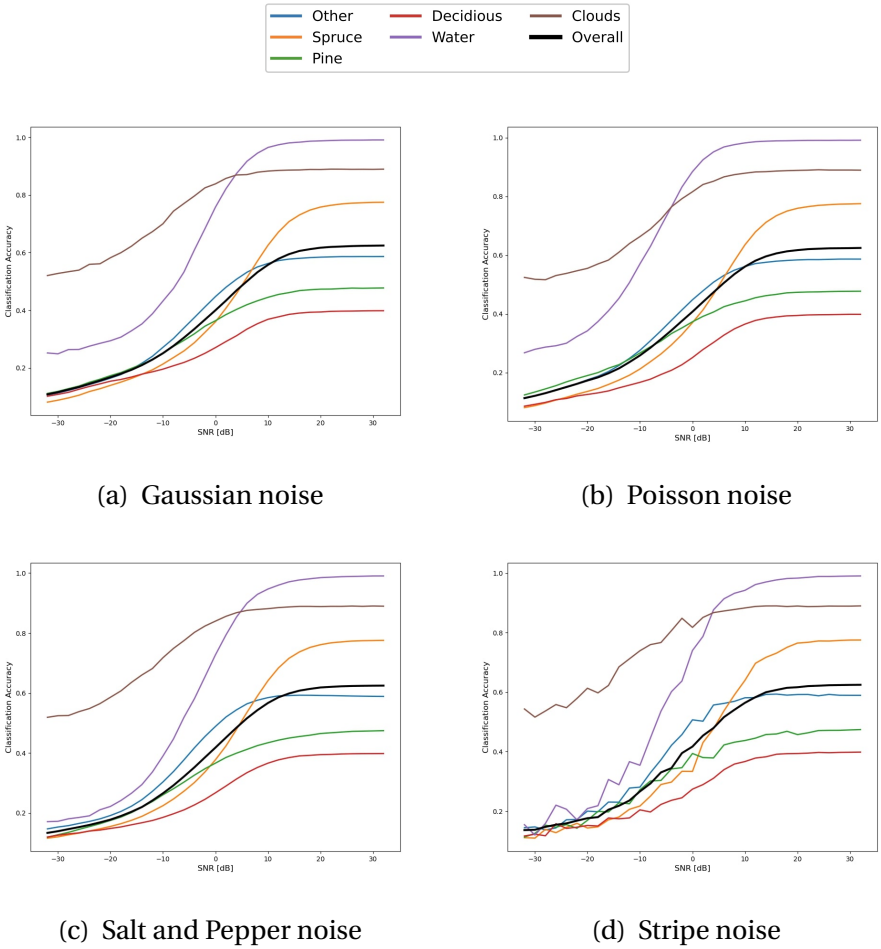


Figure 4.6: True-color image of hyperspectral data with artifacts.



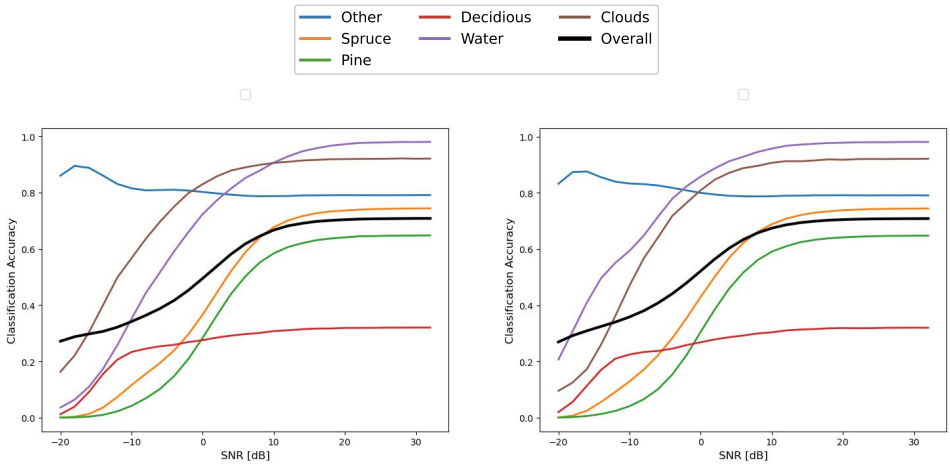
(a) Gaussian noise

(b) Poisson noise

(c) Salt and Pepper noise

(d) Stripe noise

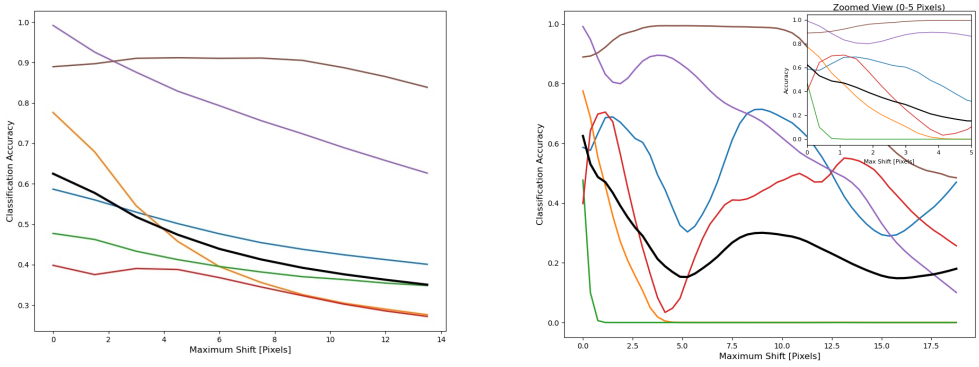
Figure 4.7: SNR (dB) vs **SVM-S** Classification Accuracy, (a) Gaussian noise; (b) Poisson noise; (c) Salt and Pepper noise; (d) Stripe noise.



(a) Gaussian noise (b) Poisson noise

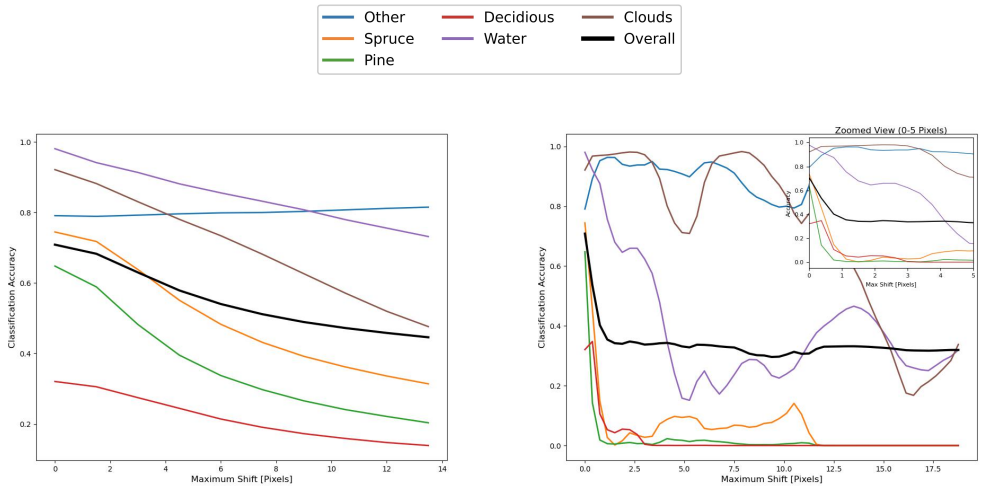
(c) Salt and Pepper noise (d) Stripe noise

Figure 4.8: SNR (dB) vs **CNN-S** Classification Accuracy, (a) Gaussian noise; (b) Poisson noise; (c) Salt and Pepper noise; (d) Stripe noise.



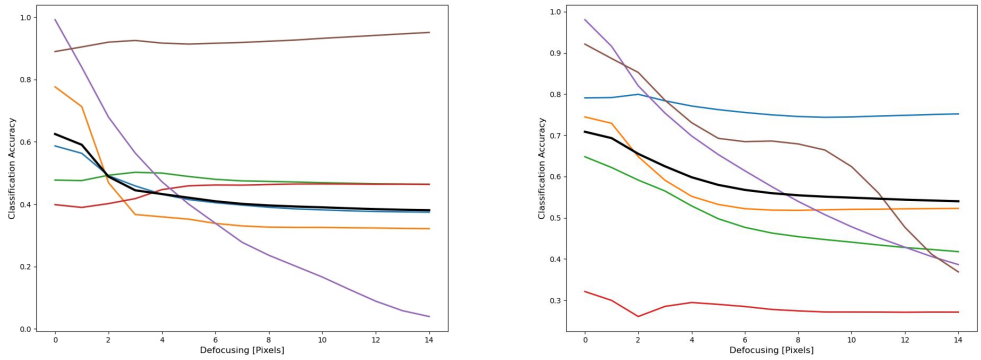
(a) Keystone (b) Smile

Figure 4.9: Maximum pixel-shift (pixels) vs **SVM-S** Classification Accuracy, (a) Keystone distortion; (b) Smile distortion.



(a) Keystone (b) Smile

Figure 4.10: Maximum pixel-shift (pixels) vs **CNN-S** Classification Accuracy, (a) *Keystone* distortion; (b) *Smile* distortion.



(a) WDD SVM-S (b) WDD CNN-S

Figure 4.11: Wavelength Dependent Defocusing (WDD) (pixels) vs (a) **SVM-S**; and (b) **CNN-S** Classification Accuracy.

Chapter 5

Assessment of Network Accuracy

This chapter touch upon multiple sources of uncertainty outlined in Table 1.1. The *explanatory variables* are investigated by analyzing how the SVM-F and CNN-F, trained on the full Faubai dataset, behave when artifacts are applied to the full test dataset. Subsequently, *response* uncertainty from additional contextual parameters such as environmental conditions, geographical coordinates, and the age of the trees are scrutinized. Lastly, a *parameter estimates* assessment is performed by unveiling the accuracies of the training batches.

5.1 Species-type Assessment

Assessing models trained on the complete training dataset is essential to evaluate the models and establish comprehensive conclusions regarding the models' response to artifacts. This entails uniformly degrading the entire test dataset for each artifact and testing the models on the degraded datasets. The visual depiction of the prediction made by the SVM-F after the introduction of artifacts, applied to a degraded example image from the test set, is depicted in Figure 5.5. This figure provides a graphical representation of the SVM-F's output, showcasing the predicted outcome after incorporating artifacts into the image.

Assessing the accuracy of the individual tree species is essential for several reasons. Different species might have varying tolerance to disturbance or relationships with other species. By identifying the species in which the model has the most difficulty, it is possible to identify specific areas that need improvement, such as data quality, labeling accuracy, or model architecture.

To analyze the results, the degraded data sets are compared towards the *Original* confusion matrices, both normalized and non-normalized. The normalized confusion matrix for the whole test dataset is found by summing up all the non-normalized confusion matrices and normalizing it at the end.

5.1.1 SVM-F Results

The influence of the artifacts on the SVM-F species predictions can be seen in Figure 5.1 and 5.3, providing the normalized and non-normalized confusion matrices.

The results show that the tree species' accuracy is highly degraded when noise is applied. However, for *poisson* noise, the prediction accuracy of *Water* exhibits a comparatively superior value when compared against the accuracy associated with the other noise types. The SVM-F performs marginally superiorly over the CNN-F when predicting tree species in noisy conditions, verifying the observed tendencies from the earlier analysis.

Upon examining the off-diagonal elements of the confusion matrix, it becomes evident that certain distortions, namely *smile*, *keystone*, and *WDD*, contribute significantly to the misclassification of tree species. Similarly to the SVM-S, small *smile* distortions generally result in the SVM-F consistently misidentifying tree species as *Clouds*. This suggests that the nature of the distortion, rather than its degree, plays a crucial role in determining the impact on classification accuracy.

A thorough analysis was conducted to investigate the intricacies and outcomes of the *smile* distortion, encompassing a range of distortions and examining the resulting prediction tendencies. Figure 5.6 depicts a specific prediction example. The comprehensive analysis revealed that an increased degree of *smile* distortion leads the classifier to misclassify pixels located in the spatial edges. This effect becomes even more pronounced when the center contains instances of *Clouds*. Additionally, when the pixel distortion exceeds a threshold of 25, the classifier incorrectly classifies the distorted edges as *Clouds*.

Conversely, both *keystone* and *WDD* distortions lead to a pervasive misclassification of all tree species into the *Other* category. Interestingly, despite having a greater pixel shift in the *keystone* distortion, its propensity to misclassify is considerably less pronounced than that of the *smile* distortion. Thus, the earlier observed tendencies regarding *smile* are also verified.

5.1.2 CNN-F Results

The confusion matrices for the CNN-F model predictions, both non-normalized and normalized, are depicted in Figure 5.2 and Figure 5.4, respectively. The confusion matrices generally show that the CNN-F is more sensitive toward noise and distortions than the SVM-F. Similarly to the SVM-S and previous observations, they reveal that most misclassification errors in the noisy data involve the tree species being misidentified as the *Other* class. As observed in the CNN-S, the *Deciduous* class is more robust towards the noise. In contrast, both *Pine* and *Spruce* are significantly degraded at SNR equal to zero.

The impact of *smile* distortion on the classification performance of the CNN-F model aligns with the observed behavior of the SVM-F classifier. Specifically, the distortion induces the misclassification of pixels in the edges of the image. However, in contrast to the SVM-F classifier, where the distorted edges are misclassified as *Clouds*, the comprehensive analysis of the *smile* response reveals that the misclassification of edges occurs as *Other* when the distortion magnitude exceeds 25 pixels. Notably, this consistent misclassification pattern observed in the CNN-F model's performance agrees with the outcomes obtained from the CNN-S model.

The *keystone* distortion misidentifies classes as the *Other*. However, both distortions appear to have a pronounced effect on the CNN-F tree species prediction, with a poorer performance than for the SVM-F model.

Regarding *WDD* images, the results reveal a degradation in the accuracy of species identification across all categories. This artifact has a particularly pronounced impact on the prediction accuracy for the *Water* class. Again, the misclassification rate of different species as the *Other* class is high. As earlier mentioned, this suggests that both models rely heavily on sharp, distinct features for accurate classification.

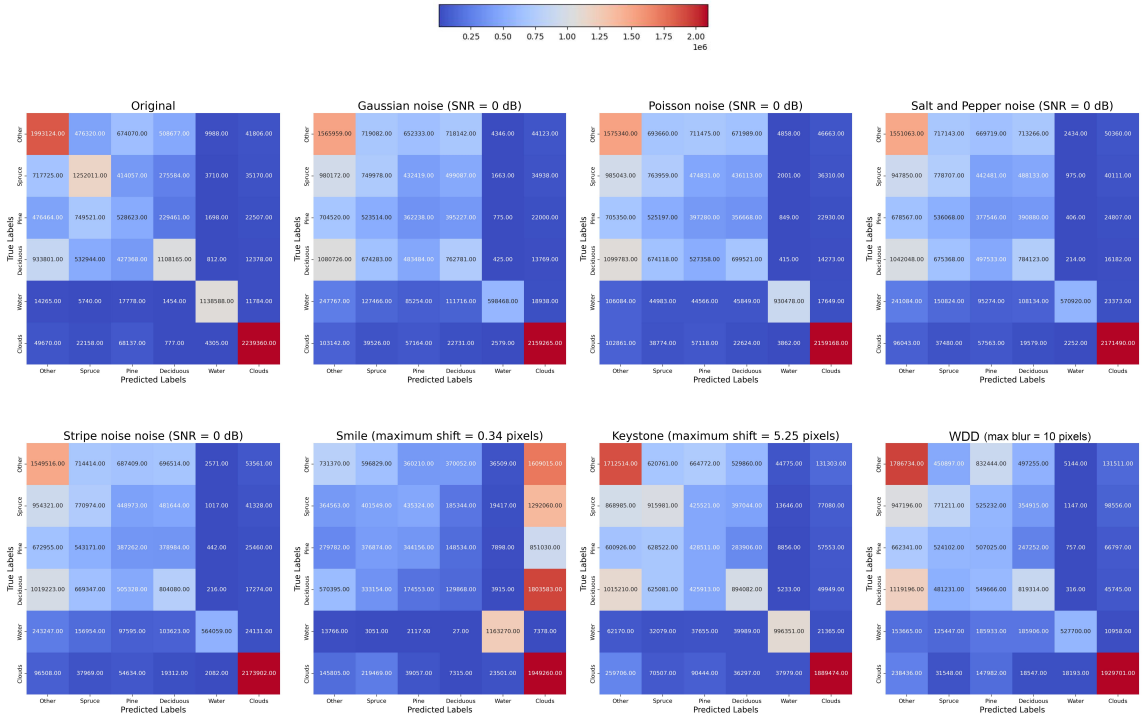


Figure 5.1: Confusion matrix of **SVM-F** for original and degraded test data sets.

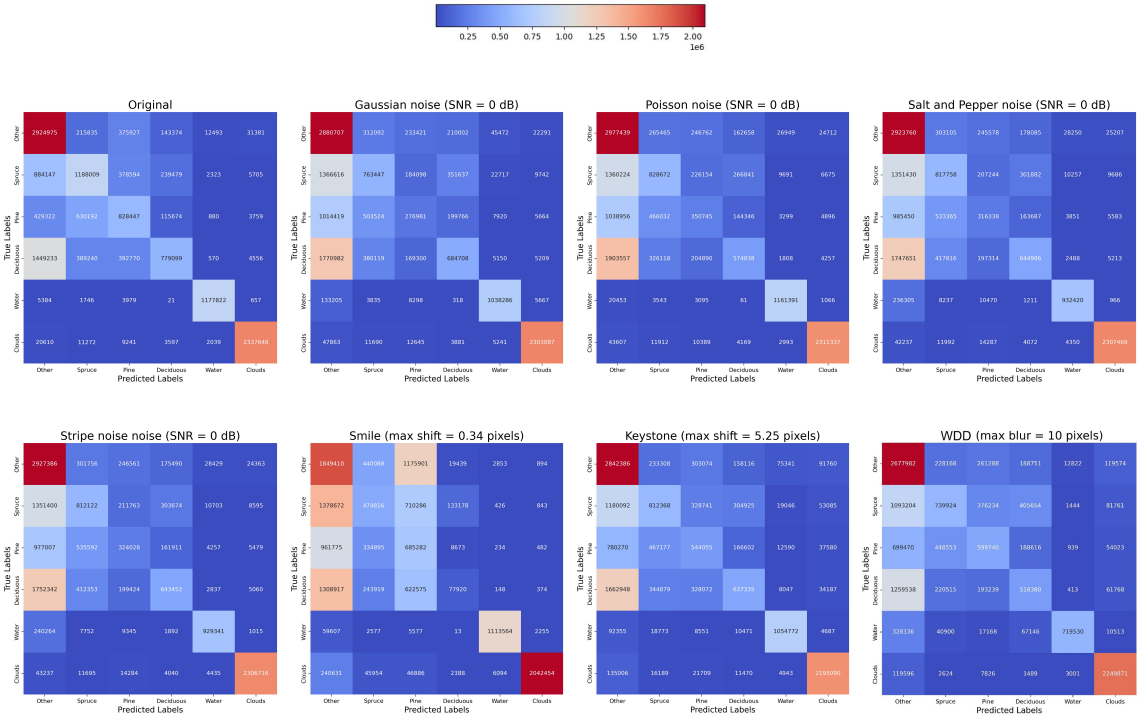


Figure 5.2: Confusion matrix of **CNN-F** for original and degraded test data sets.

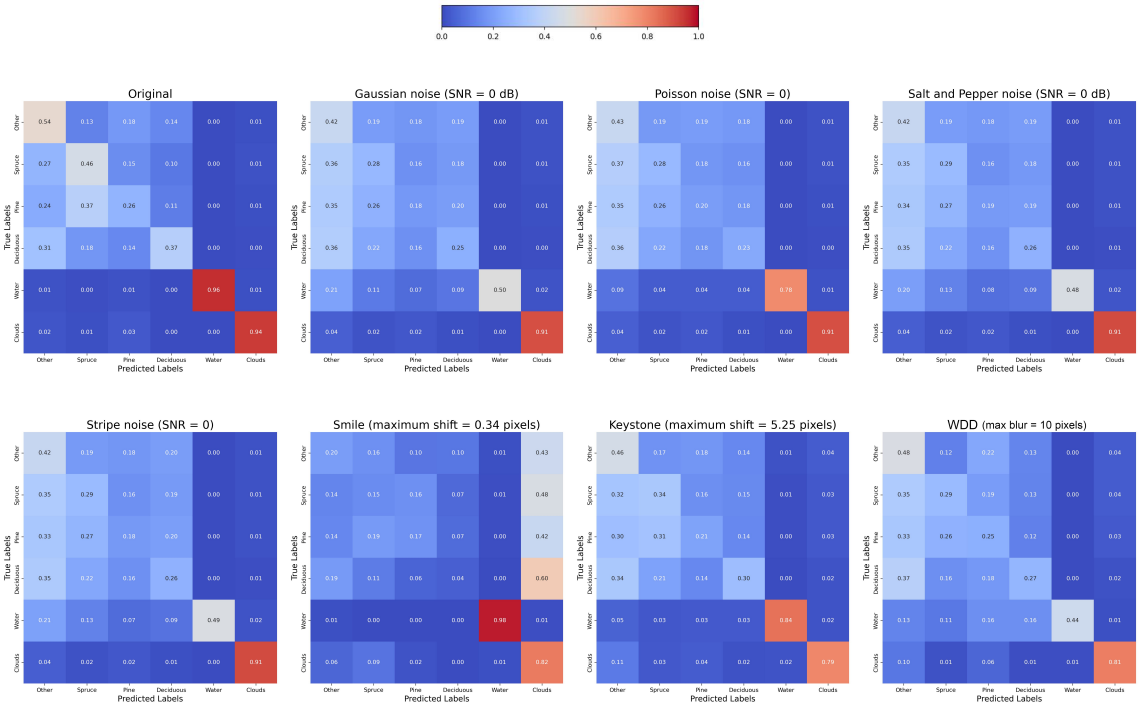


Figure 5.3: Normalized confusion matrix of **SVM-F** for original and distorted test data sets.

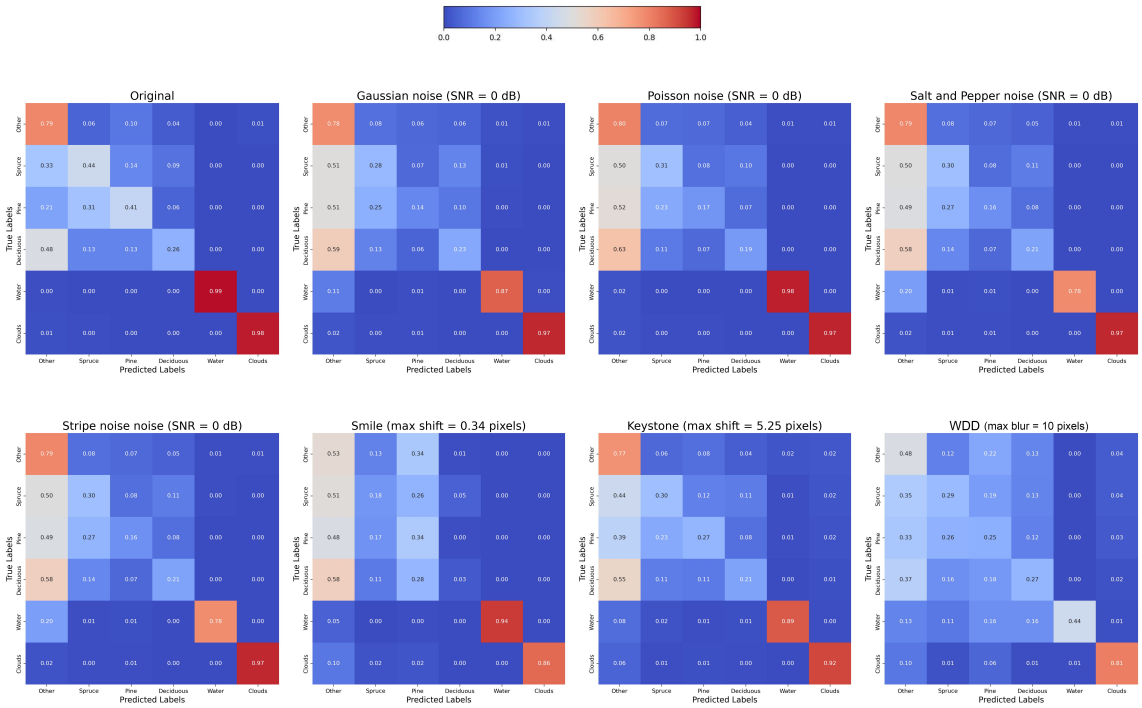


Figure 5.4: Normalized confusion matrix of **CNN-F** for original and distorted test data sets.

5.2 Geographic Assessment

Ensuring that the data for model training is representative of all geographical areas can ensure accurate classification in all geographical areas. At the same time, it can reveal potential uncertainties if the data has not adequately captured tree characteristics and environmental conditions across various regions or countries, as outlined for the *response* variable in Table 1.1.

Thus, an assessment of the geographic locations' influence on the accuracy of the models is essential to conduct a thorough spatial analysis of the results. This is performed by analyzing if the models exhibit varying classification performances of species within geographical areas, such as regions and countries.

Countries

The species distribution within each country is shown in Figure 3.3 and 3.2. A universally consistent identification pattern for *Water* is evident across all geographical areas. Interestingly, Norway demonstrates superior accuracy in classifying *Clouds*, even though Sweden and Finland have similar amounts of *Clouds* in the training dataset. On the other hand, Canada exhibits a complete lack of *Cloud* identification, which can be attributed to the absence of such elements in the Canadian image captures.

A closer inspection of the SVM-F results unveil that *Spruce* is the most accurately classified

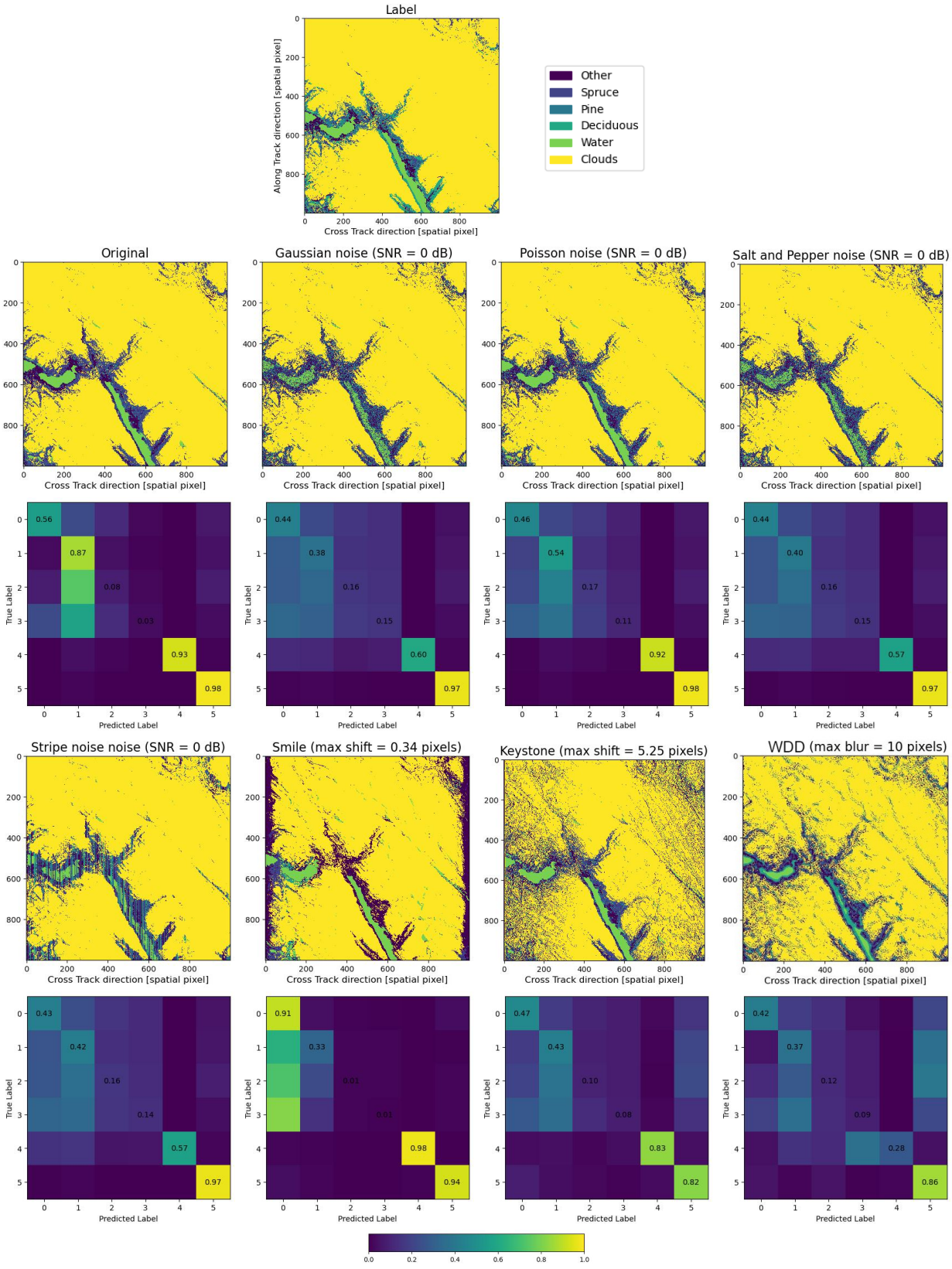


Figure 5.5: Predictions by **SVM-F** classifier.

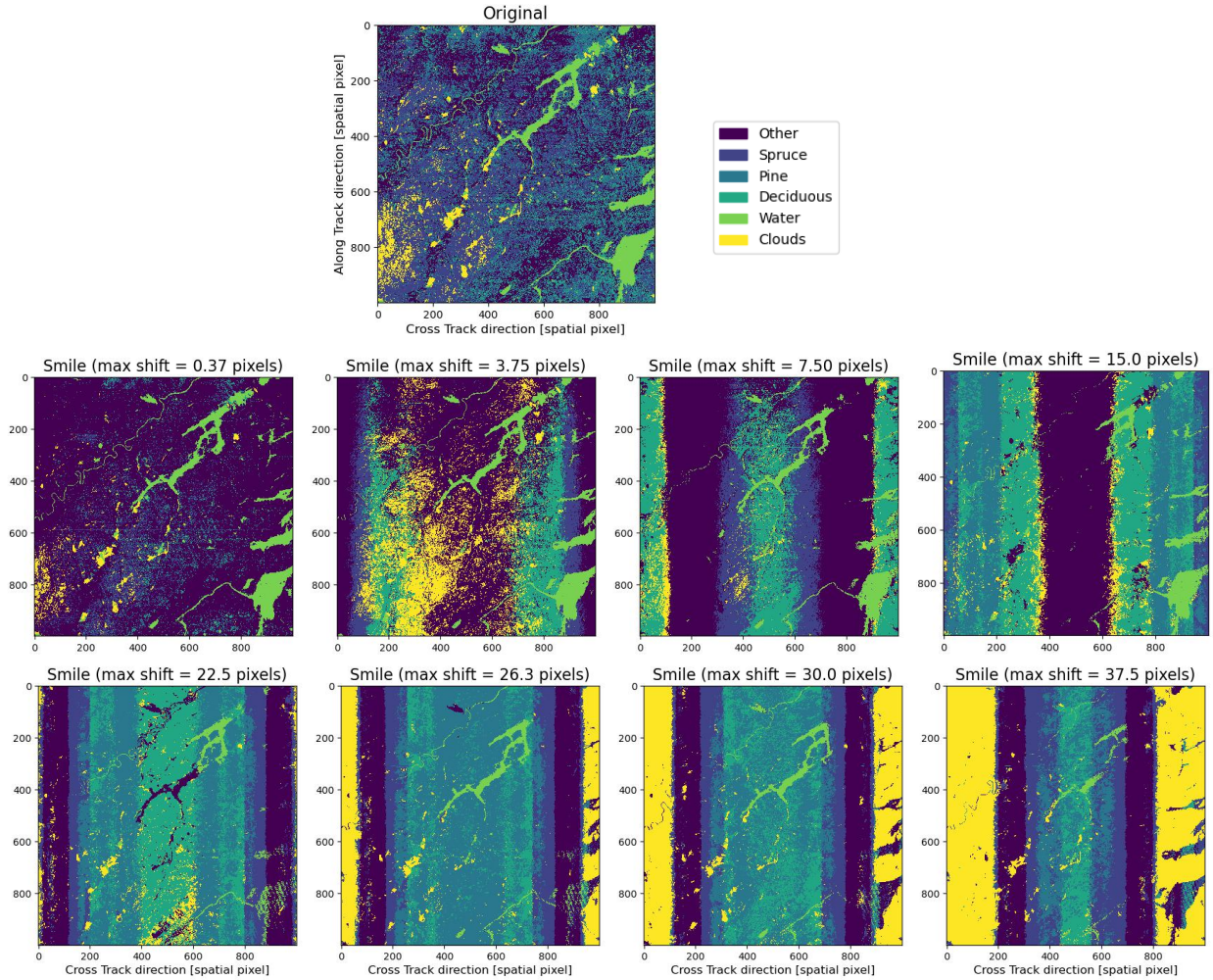
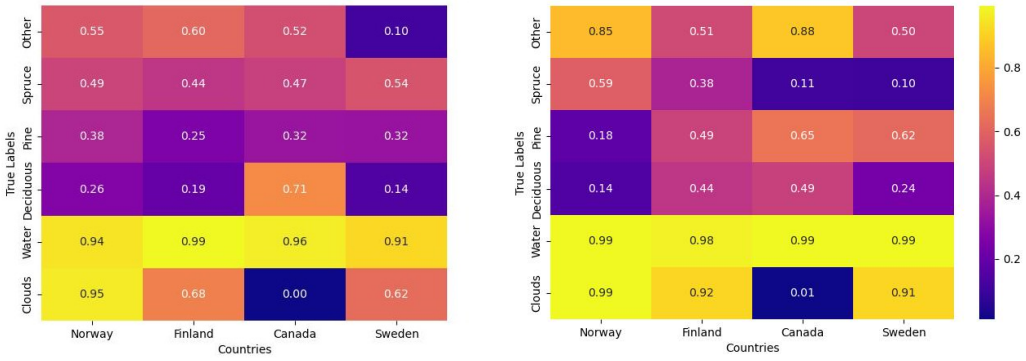


Figure 5.6: Smile predictions by **SVM-F** classifier, for a single image.

species in all countries, with the notable exception of Canada, where *Deciduous* trees are correctly classified with an accuracy of 71%.

Switching the focus toward the CNN-F model, a distinct divergence in classification patterns is noticeable. Norwegian *Spruce* is classified with the highest degree of accuracy, while in the other countries, *Pine* exhibits the most significant classification accuracy. Interestingly, even though the SVM-F classified the Canadian *Deciduous* better, the CNN-F performs better for the *Pine* class. Despite the high classification accuracy of Canadian *Pine*, the identification accuracy of Canadian *Deciduous* trees is not as robust as that observed in the SVM-F model.



(a) SVM-F (b) CNN-F

Figure 5.7: Accuracies of Countries for the test dataset.

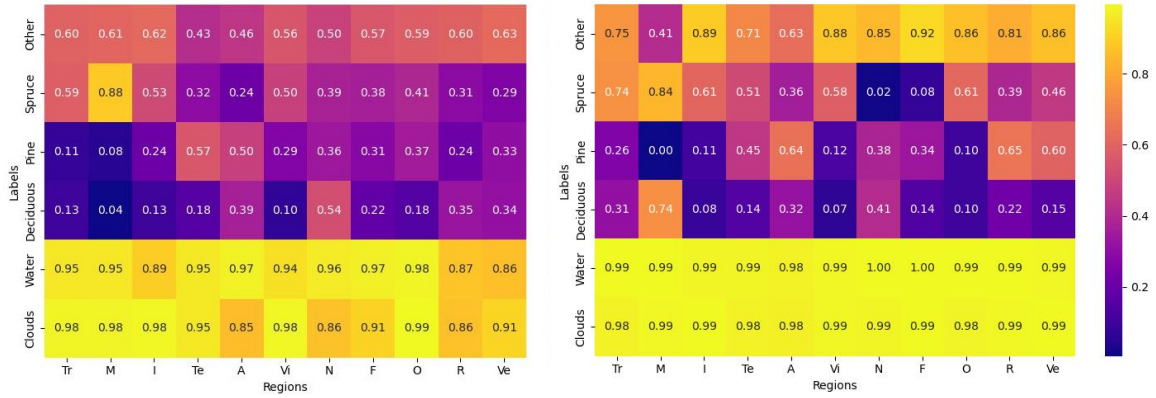
Regions

The accuracy mapping within the Norwegian regions is presented in Figure 5.8. It can be seen that the CNN manages to classify both *Clouds* and *Water* accurately in all regions. Furthermore, *Spruce* has the highest accuracies in Trøndelag and Møre and Romsdal and the lowest in Nordland and Finnmark.

The SVM-F seems to get a higher median accuracy for the *Spruce* and *Pine* species across regions, in comparison to the varying accuracy of the CNN-F. *Spruce* are shown to be particularly well classified in Trondheim, Møre and Romsdal, Innlandet, and Viken, and poorly classified in Nordland and Finnmark (for the CNN-F). *Pine* has a low accuracy for Trøndelag, Møre and Romsdal, Innlandet, and Viken, while achieving better results in Telemark, Agder, Rogaland, and Vestland. *Deciduous* are generally poorly classified within Norway, but perform well in Nordland for both classifiers and in Møre and Romsdal for the CNN-F.

In a broader perspective, it is notable that regions with high accuracy in classifying the *Spruce* species often display relatively lower accuracy in classifying the *Pine* species, and vice versa. Furthermore, an interesting observation is that while the CNN-F performs excellently in

certain regions, the prediction accuracies tend to vary more than the SVM-F. Consequently, the SVM-F demonstrates a more balanced prediction of tree species across different regions. This finding implies that the SVM-F model achieves a more consistent and evenly distributed classification performance for the given Norwegian dataset.



(a) SVM-F (b) CNN-F

Figure 5.8: Accuracy of Regions for the Norwegian test dataset.

5.3 Tree Age Assessment

Similarly to the geographical assessment, variables within the representation of tree age can affect the *response* variable in Table 1.1. As presented in subsection 3.2.1, the SK16 Beta version by NIBIO comprises ".tif" files that include tree species estimated ages. This valuable attribute allows for investigating whether certain three age groups are more readily classifiable than others. By incorporating the environmental variable of tree age into the predictive model, a more comprehensive and nuanced understanding of tree species classification across different life stages can be achieved. This analysis specifically focuses on the Norwegian captures within the dataset.

Young trees, whether *Spruce*, *Deciduous*, or *Pine*, tend to fall within the 20-50-year-old range. Mature trees show slight variation between species: mature *Spruce* and *Pine* trees are generally between 60-100 years old, while for *Deciduous* trees, this age range is 50-100 years [Loehle (1988), Spies et al. (1991), for Nature (2016)]. Trees aged over 150 years are classified as old for all three species. With these considerations, we can broadly categorize the tree ages into three groups: 0-50 years representing *young* trees, 50-100 years standing for *mature* trees, and those over 100 years as *old* trees. It's worth noting that the latter category includes exceptionally long-lived individuals surpassing 150 years.

As displayed in Figure 5.9, the SVM-F predictions for a single test set image are elucidated, accompanied by the corresponding confusion matrix (5.9d), and the accuracy of each tree species by age (5.9e). Also included are the distributions of tree species (5.9a) and tree ages (5.9b) for the examined image. The distribution of tree species is approximately uniform. However, the distribution of tree ages exhibits an over-representation of *young* trees.

The corresponding confusion matrix elucidates that the image primarily classifies *Spruce* with superior accuracy compared to other tree species, as seen in the accuracy distribution per tree species. Interestingly, while the image contains a larger number of *young* trees, the prediction accuracy for each species improves with age. This trend indicates that older trees might possess more distinctive features that the models can more easily detect and differentiate, thus reducing the data uncertainties in the *response* variable for the SVM-F.

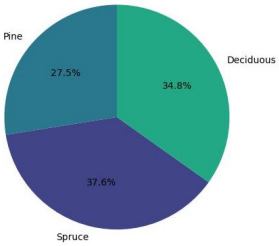
Conversely, the outcomes are ambiguous when observing results from the entirety of Norwegian captures within the test dataset. The distributions shown in Figure 5.10b indicate a significant representation of *young* trees (66%). Additionally, a smaller pixel group represents *Deciduous* trees (9.9%), compared to *Pine* (13.4%) and *Spruce* (15.7%).

Both models classify *Spruce* trees across all age groups quite competently. The SVM-F manages to classify 41% *Deciduous* trees, whereas the CNN-F finds it challenging to classify other types than *Spruce* for the test dataset. The corresponding confusion matrices reveal that the SVM-F surpasses the CNN-F in classifying *Deciduous* trees. Consequently, these results are anticipated.

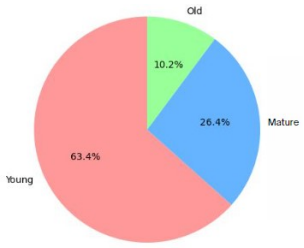
The inconsistency between the single image and full testing set could be attributed to the substantial variation in classification accuracy among different tree species and a varying number of species within images. Specifically, the influence of tree age on classification accuracy seems less definitive when considered across a wider variety of species and geographic locations. This suggests the presence of additional complexities and potentially unaccounted sources of uncertainty within our models, reaffirming the need for further investigation and refinements.

5.4 Parameter Estimates Assessment

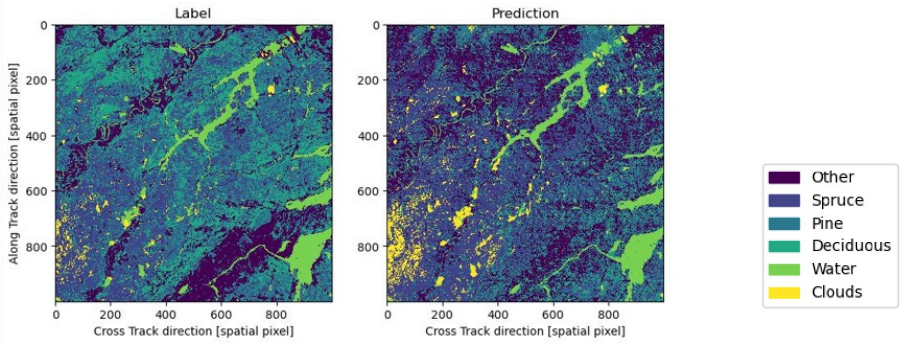
As defined in Table 1.1, the uncertainty in *parameter estimates* arises from the model training process. It underlines the intrinsic variability and unpredictability associated with the training procedure of the SVM-F and CNN-F, introduced in section 2.7 and 2.8. This variability and uncertainty become evident by “assessing the performance and convergence of partially trained models,” as outlined in Table 1.2. This is performed by finding the model convergence across batches for the SVM-F and CNN-F.



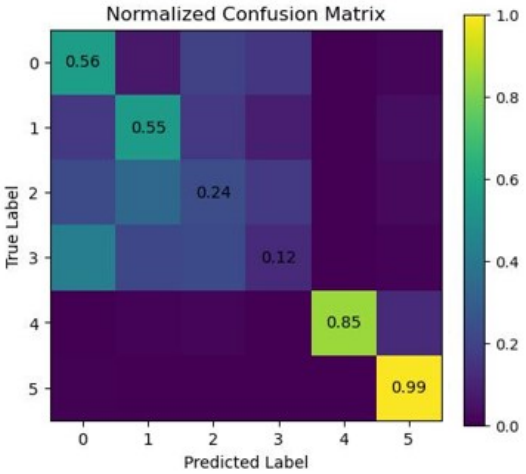
(a) Tree Species



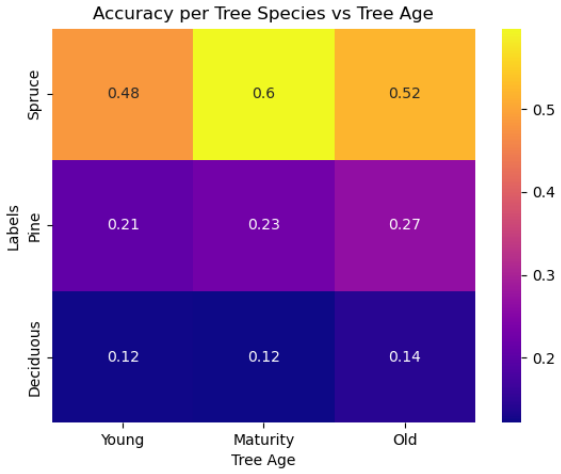
(b) Tree ages



(c)

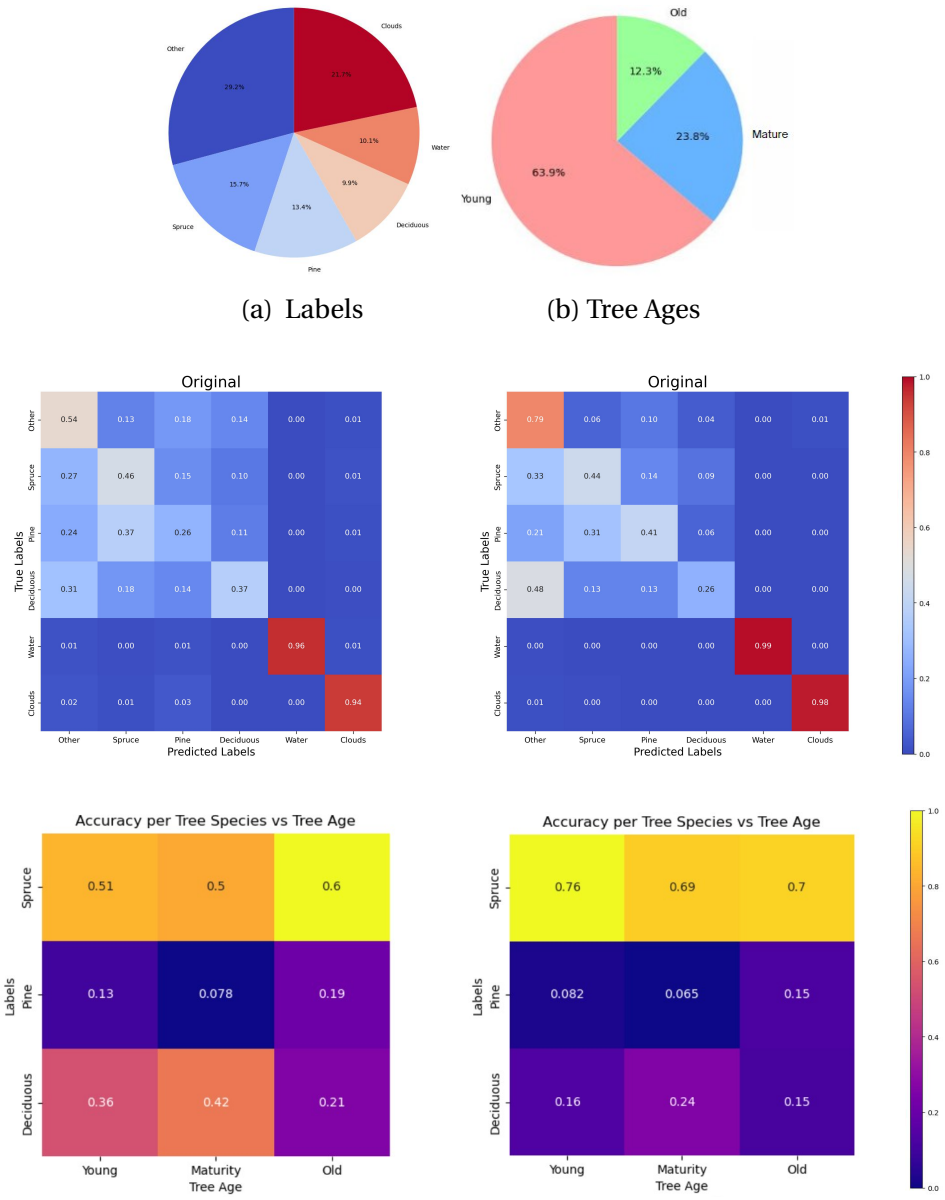


(d)



(e)

Figure 5.9: Example of Tree Age distribution and Accuracy per Tree Species in SVM-F classifier for a specific image.



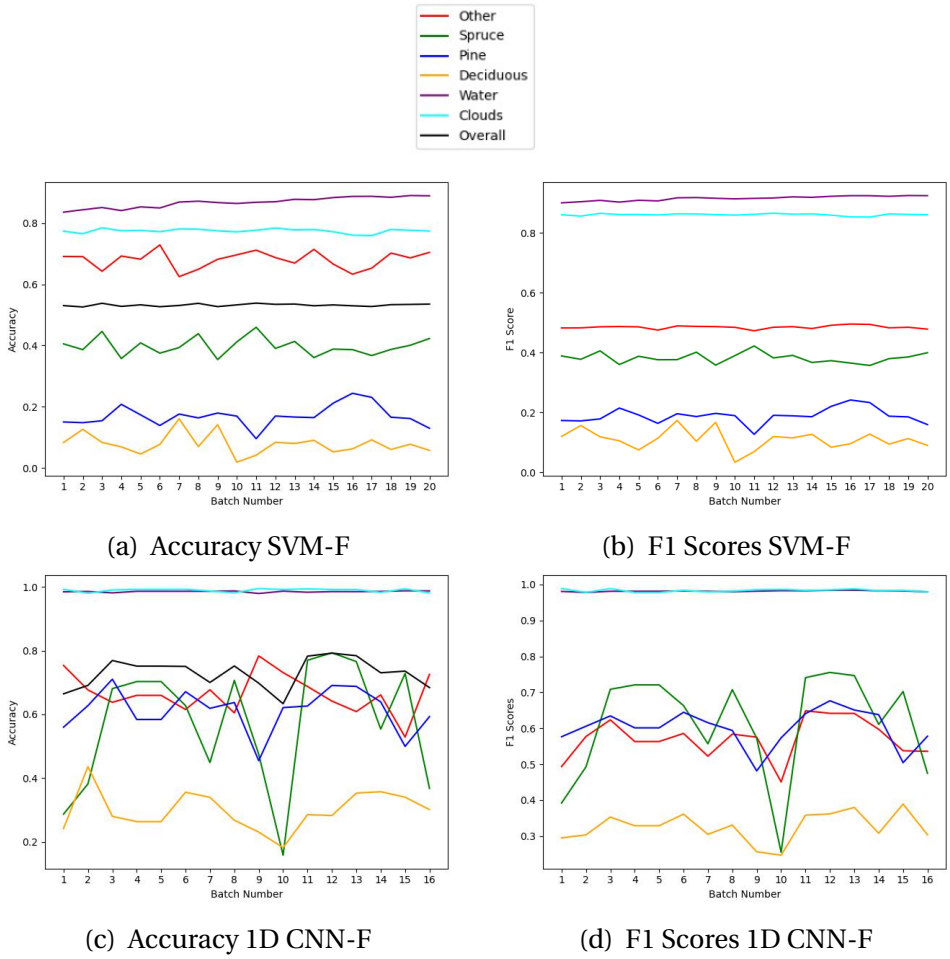
(a) Labels (b) Tree Ages

(c) SVM-F (d) CNN-F

Figure 5.10: Label Accuracy in test dataset and Accuracy per Tree Species for Tree Age.

The class accuracies and F1 scores for different SVM-F batches can be seen in Figure 5.11a and 5.11b. These show that even though the prediction accuracy for the specific image was 32%, the average accuracy based on the whole test set equals 54%. Moreover, it shows that the SVM-F struggle to exhibit significant performance improvement throughout twenty batches, underscoring the challenges of ensuring consistent model improvement during training.

Similarly, the visualization of CNN-F training progress across different batches (Figure 5.11c and 5.11d) shows comparable challenges. Despite the increase in batch size, the CNN-F model, like the SVM-F, has difficulties improving class accuracies and F1 score. Notably, the performance outcomes for the *Spruce* class in the CNN-F model exhibit considerable variability across different batches, as demonstrated by the substantial accuracy drops in batch 10. This suggests the presence of additional uncertainties and potential overfitting in the CNN-F model's training process.



(a) Accuracy SVM-F

(b) F1 Scores SVM-F

(c) Accuracy 1D CNN-F

(d) F1 Scores 1D CNN-F

Figure 5.11: Accuracy and F1 Scores for each batch in the models trained of the full Faubai dataset.

Chapter 6

Discussion

This chapter adheres to the hierarchical assessment framework outlined in Table 1.2, discussing the findings from the various sources of uncertainty given in Table 1.1. These uncertainties encompass *response*, *explanatory variables*, *parameter estimates*, and *model structure*, in which the findings are listed in Table 6.1. Furthermore, the chapter evaluates the framework used.

6.1 Response Uncertainty

By evaluating uncertainties inherent in the *response* variable, the tree species have been systematically addressed through the steps outlined in Table 1.2. A comprehensive examination of the ground truth data was undertaken, encompassing a detailed review of its distribution, potential sources of error, and the methodology employed in labeling.

6.1.1 Irregularities in Data Representation

Details about the distribution of the dataset were presented in subsection 3.2.1. The figures 3.3 and 3.2 accurately illustrate the distribution across labels, geographic coordinates, and distinctive regions. Nevertheless, upon further scrutiny of label distribution across various countries, certain anomalies come to light, such as the relative scarcity of cloud captures in Canada when juxtaposed with other countries (referring to Figure 3.3g and 3.2g).

Moreover, when examining the geographic distribution of class accuracies, it became apparent that the predictions of the Canadian results deviated from those of the other countries. Specifically, within the results of the SVM outlined in section 5.2, the classification of *Spruce* is predominantly accurate in most countries, except Canada. This finding is notable considering that the Canadian datasets prominently feature the *Spruce* species. This suggests that what is nominally considered the same type of tree may correspond to different species within Canada. Specifically, trees with the same genus designation may, upon closer examination, be identified

as different species. Conversely, the classification of *Clouds* in the Canadian dataset was consistently poor, which can be attributed to the absence of *Clouds* instances. This discrepancy in class accuracies between countries suggests the presence of country-specific characteristics and variations in the data that influence the performance of the classification models.

Compared to the factual data obtained from Sweden and Finland, unifying the Canadian labels necessitates a more comprehensive consolidation to align with the format of Norwegian tree species. This might elucidate the diminished predictive accuracy of the *Spruce* species. Given the results, one can conjecture that unifying 37 species into the three categorizations applied in Norway yields less ideal outcomes. Moreover, other ways of merging labels should be investigated.

Variables such as tree age, climatic variables, soil composition, and altitude, which may exhibit significant regional differences, could potentially influence the model's precision in tree species classification. Such discrepancies might influence the outcome of tree species classification.

The distributions demonstrated in Figure 5.10b indicate a preponderance of *young* trees. Even though the varying results make it hard to conclude regarding tree age, general models may encounter challenges in accurately classifying older trees with the current dataset, given the overrepresentation of juvenile trees. Thus, the tree age is attributable to inaccuracies in age determination or disproportionate representation across diverse life stages.

The complexity of forest ecosystems makes their detailed characterization daunting. Forests have diverse attributes and characteristics that pose significant challenges to comprehensive representation. Consequently, applying an AI model in such a context might result in the loss or obscurity of specific data. This occurrence is plausible, especially in scenarios where the model, faced with a slight probabilistic edge, opts for one label over another. Thus, the model's classification might not entirely reflect the species distribution or variety inherent in the data.

While drawing a definitive conclusion from the tree age assessment remains challenging, it is unequivocally observed that geographical disparities have negatively impacted the accuracy of cloud classification in Canada. Nonetheless, considering the focus is primarily on the Norwegian captures, the ground truth dataset maintains a significant degree of representativeness of the intended population.

6.1.2 Quality and Reliability of Ground Truth Data

The performance assessment of different ground truth datasets revealed specific observations. The Norwegian dataset achieved an overall accuracy of 91% at the stand level, with misclassifications primarily occurring in younger spruce stands, mixed forests, and areas with incomplete

coverage [Rahlf et al. (2021)]. This observation warrants further attention since the test dataset predominantly comprises juvenile trees (as shown in Figure 5.10b). Concurrently, systematic inaccuracies were detected in the predictive outputs of SR16, specifically in forest stands characterized by elevated tree densities and in geographical regions featuring a blend of deciduous and coniferous forests.

In the case of the Swedish captures, stand age information was excluded from the 2015 map due to insufficient quality. Errors were larger in southern areas, attributed to denser and larger forests in the south. As is evident from the map of data captures referenced in Figure 3.5, merely two scenes have been collected from southern Sweden. For the Finnish dataset, adjustments were made to minimize errors, resulting in comparable parameters across four distinct study areas representing different geographical regions [Katila (2004)]. Lastly, the Canadian dataset's accuracy was assessed using independent validation data from the NFI, indicating an overall accuracy of 93.1% [Hermosilla et al. (2022)]. However, as mentioned earlier, the complex labeling procedure to fit the Norwegian tree species introduces uncertainties affecting prediction accuracies.

In the context of multi-source forest inventories, the presence of errors tends to escalate with an increased number of data sources. In this study, errors in each ground truth dataset introduce substantial uncertainties when making predictions using the models. Since intermediate ground truth is already based on uncertain information, it is reasonable to assume that the true accuracy in our results is even lower. Consequently, while the data can serve as a valuable basis for forest management planning, caution is advised, as highlighted by Breidenbach, Rahlf, Räty, Hauglin & Bergseng (2022), suggesting the incorporation of field measurements for verification purposes.

6.2 Explanatory Variables Uncertainty

The evaluation process for the *explanatory variables* encompassed an examination of the HSI data derived from the PRISMA satellite, in conjunction with an exploration of the influence of artifacts on model performance. This was conducted by meticulously analyzing the HSI data quality and evaluating the various models' susceptibility to image artifacts.

6.2.1 Quality of Hyperspectral L1 PRISMA Data

Multiple studies have examined HSI data quality obtained from the PRISMA satellite, as discussed in section 2.5. Notably, it was discovered that the sensors employed by PRISMA are calibrated using internationally recognized test sites. Furthermore, an incidence of coherent noise

was identified in PRISMA imagery.

However, earlier studies imply that relatively small errors and differentials have been observed in the spectrum of the hyperspectral data used in this thesis. Additionally, the L1 product enhances data quality and integrity by addressing and rectifying challenges such as bad pixels and geometric distortions. These factors collectively contribute to less uncertainty within *explanatory variables*.

6.2.2 Impact of Artifacts on Model Performance

The impact of artifacts on model performance was first assessed for the SVM-S and CNN-S models in section 4.3, offering a simplified investigation of noise characteristics. Further species assessment on the larger dataset was conducted in section 5.1 to ensure the conclusions drawn are valid in a broader context. This analyzed how the SVM-F and CNN-F behaved when artifacts were applied to the full test dataset.

Model Response to Noise

In analyzing the single image trained models' response to noise (subsection 4.3.1), the SVM-S and CNN-S classifiers exhibited similar trends, with accuracy increasing as SNR increased. A general tendency was misclassifying classes as *Other* or *Clouds*. It was also found that the CNN-S was more vulnerable to noise-related variations than the SVM-S.

From section 5.1, it was found that most misclassification errors involve the species being misidentified as the *Other* class. Interestingly, the CNN-F results show that the *Deciduous* class seems more robust towards the noise, probably due to the diversity of tree species it encapsulates. Consequently, the higher heterogeneity within the *Deciduous* class potentially demands less sophisticated computational strategies to achieve accurate predictions in CNNs, even with a noisy dataset.

Model Response to Distortions

The responses of varying *smile* and *keystone* distortion to SVM-S and CNN-S were presented in subsection 4.3.2. The results revealed that the SVM-S and CNN-S classifiers exhibit different robustness to *keystone* and *smile* distortions.

The SVM-S maintains relative stability in the face of *keystone* distortions, barring *Clouds*, while the CNN-S falters significantly for tree species with increased pixel shift. For the *smile* distortion, both classifiers exhibit rapid declines in accuracy for tree types within the first pixel of distortion. The CNN-S accuracy of *Spruce*, *Pine*, and *Deciduous* accuracy is degraded 84%, 98%, and 76% within a pixel. This is also observed, albeit to a lesser extent, in the case of SVM-S, in which the accuracy of *Spruce* and *Pine* is degraded 26%, and 100%, respectively, within the first

pixel of distortion.

From the results found in section 5.1, it is clear that *smile* distortions lead the classifiers to misclassify pixels in the spatial edges. The confusion matrices and in-detail *smile* analysis found that the SVM-F misclassifies edges as *Clouds*, while the CNN misclassifies them as *Other*. This discrepancy in the classification outcomes of the two models under the same distortion condition points towards the different ways SVMs and CNNs handle and interpret image data.

The presence of the *keystone* distortion, in contrast to the *smile* distortion, results in a more generalized misclassification of all classes, assigning them to the *Other* category. Interestingly, even with minor degrees of *smile* distortion, the models demonstrate significant sensitivity to this type of distortion. On the other hand, when considering similar magnitudes of *keystone* distortion, it is observed that the models are not significantly affected.

Model Response to Defocusing

The model response for varying amounts of pixel defocusing was presented in subsection 4.3.3 and section 5.1. The results showed decreased classification accuracy, with the CNNs being more sensitive to defocusing than the SVMs. This might be because the CNNs search for clear patterns and edges in the images, which fades with an increased blur effect.

Specifically, it was found that *Water* is particularly degraded with an increased defocus. Additionally, *Pine* and *Deciduous* exhibit a slight accuracy increase, which implies that the soothing might reduce the overlap between the spectral features of these classes and make them more distinguishable in the feature space. Both classifiers were significantly affected by the first defocused pixel.

6.3 Parameter Estimates Uncertainty

The performance and convergence analysis of partially trained SVM-F and CNN-F models involved monitoring the training progress, as depicted in Figure 5.11.

Examining the accuracy and F1 scores revealed that both models encountered challenges in achieving notable improvements, even with an increased number of training batches. The CNN-F model exhibited inconsistent outcomes when classifying the *Spruce* category, as seen for the 10th model. This variability suggests the *parameter estimates* are suboptimal or not adequately tuned for the specific characteristics of *Spruce* instances within the dataset, implying that the specie might be harder to predict than the others. Furthermore, even though the class accuracies are higher in the CNN-F, the results seem less reliable.

The complex structure of CNNs makes investigating inconsistent outcomes more challenging. It requires careful analysis, considering various factors influencing the model's behavior and performance. Specifically, examining the data for the specific batch of analyzing the hyperparameters choices could be performed. However, due to time restrictions, this was not prioritized in this thesis. Further analysis is warranted to investigate the factors contributing to the observed performance disparities and potential strategies to enhance the classification capabilities of both models.

6.4 Model Structure Uncertainty

The model structure uncertainty lies in the investigation process, which refers to the structure of the equations that link the *response*, *explanatory variables*, and *parameter estimates* in the SVM and CNN models. This involves investigating the assumptions made by each model to assess whether they are reasonable for the problem domain.

6.4.1 Model Assumptions

The results in section 4.3 and 5.1 have shown that both the single (SVM-S, CNN-S) and full Faubai-trained models (SVM-F, CNN-F) struggle to classify tree species while performing well for *Clouds* and *Water*. Part of the reason may be that the assumptions underlying SVM and 1D CNN models affect tree species classification.

SVMs assume linear relationships exist between the *explanatory variables* and the *response*. This presupposition may prove inadequate in complex classification scenarios such as the differentiation of tree species. Moreover, SVMs strive to optimize the margin, a characteristic that may not always be realizable, particularly in cases where data points lie close to the decision boundary. Consequently, this could result in misclassifications, exemplified by the incorrect identification within the *Other* class.

Compared to SVM models, the underlying assumptions concerning CNN are inherently more intricate. Specifically, 1D CNNs leverage local connectivity based on the principle that nearby values in the sequence data are more likely to be related than distant ones.

The interconnection between the *explanatory variables* and the *response* is inherently non-linear and multifaceted due to the utilization of convolutional processes and non-linear activation functions, which are learned empirically from the data. Another consideration is that the pooling operations, employed in 1D CNNs for downsampling, can inadvertently lead to the loss of detail-rich information.

The accuracy scores for the different training batches in Figure 5.11c revealed that some CNN models suffered from a sudden drop in class accuracy (as seen for the 10th model). The complex structure of CNNs, including non-linear and hierarchical feature extraction, many parameters, lack of interpretability, deep and distributed representations, and potential overfitting, makes investigating accuracy drops more challenging.

Additionally, the CNN results have shown that artifacts affect 1D CNN models more than SVMs. Specifically for the CNN, the WDD was shown to influence the *Water* prediction accuracy. From the model assumptions mentioned, noise can disrupt the local context, affecting the accurate extraction of meaningful features. Moreover, the downsampling procedure can make the model more susceptible to noise. Thus, the SVM benefits from its global structure-based approach and focus on support vectors.

6.4.2 Variable and Model Relations

Figure 2.21 gave an example of all models' predictions for a specific image. Specifically, a decline in classification performance of 48.8% and 42.2% when trained on a larger dataset was observed for the SVM and CNN, respectively. This is a notable generalization error that can be attributed to several factors.

When transitioning from training a model on a single image to utilizing a larger dataset, several factors can contribute to a decline in performance. Overfitting becomes a concern as the model may have learned specific features or patterns unique to the single training image, resulting in poor generalization to new, unseen data. The increased complexity and diversity within the expanded dataset pose additional challenges, making it harder for the classifier to find an optimal decision boundary that encompasses all the images. Additionally, hyperparameters that were effective for the single-image scenario require adjustments to accommodate the larger dataset. By fine-tuning hyperparameters and considering the diverse characteristics of the full dataset, it is possible to achieve better overall performance on the test dataset than the example prediction suggests.

The CNN model might have exhibited relatively better generalization performance than the SVM model due to its inherent ability to leverage local connectivity from the input data (as earlier mentioned). It is important to note that the improved generalization of the 1D CNN model in this scenario is not guaranteed and can vary depending on several factors, including the specific architecture, hyperparameter settings, and the nature of the dataset.

Both SVM-F and CNN-F perform differently for tree species in different countries and regions

(Figure 5.7 and 5.8), suggesting that the mathematical relationships and assumptions between the *response* and *explanatory variables* might differ for these two models. Similarly, an over-representation of *young* trees was found. Moreover, the inconsistencies imply that models must be adjusted to account for the geographical differences or the datasets must be more equally distributed.

Source	Uncertainty Findings
Response Ground truth	1. An irregular representation was detected among the Canadian species, excluding the *Clouds* label, which comprised only 0.1% of the observed labels. This suggests a lack of uniformity in the frequencies of labels. 2. In the Norwegian tree population, there was a noticeable dominance of *young* trees, accounting for 66% of the total. This indicates an age imbalance in the Norwegian dataset, where younger trees are more prevalent. 3. Labeling choices contributed to misclassifications for Canadian tree species.
Explanatory variables Optical spectra	1. *Smile* distortion significantly impacted the model's performance, leading to a degradation in tree species accuracy of at least 76% within a pixel when utilizing the 1D CNN-S. 2. Artifacts affected the accuracy of the *Deciduous* class less than *Spruce* and *Pine* class. The robustness of the *Deciduous* class might stem from its diversity. 3. Artifacts affected the accuracy of the 1D CNNs more than the SVMs.
Parameter estimates Model parameters	1. The 1D CNN was more sensitive to the details of the training procedure than the SVM.
Model structure Model design	1. The 1D CNNs exhibit higher overall accuracies compared to SVMs. 2. A generalization error was significant on the Faubai dataset. SVM-F and CNN-F showed 48.8% and 42.2% decreases in accuracy relative to the SVM-S and CNN-S when tested on the single image test set.

Table 6.1: Findings for sources of uncertainty presented in Table 1.1, assessed according to Table 1.2.

6.5 Framework Evaluation

The framework employed in this thesis provides insights into important aspects of model performance for the introduced models, namely the SVM-S, SVM-F, CNN-S, and CNN-F. The uncertainty assessment can be guided by identifying and defining the sources of uncertainty to verify their results [Simmonds et al. (2022)]. Unfortunately, due to the uncovered uncertainties, verification of the model results has not been met for the models utilized in this study.

The assessment of uncertainty within the data (*response* and *explanatory* variables) is often overlooked, despite its significance. This can be attributed to the time-consuming nature of such investigations. In the context of remote sensing data, earlier studies have predominantly focused on validating the datasets, neglecting the comprehensive examination of uncertainties inherent within them.

A thorough analysis of the ground truth data sources uncovered a predominant presence of *young* trees within the Norwegian dataset. Furthermore, previous studies have consistently indicated the misclassification tendency associated with *young Spruce* specimens. This investigation of individual components within the dataset revealed uncertainties that may not have been evident without comprehensive scrutiny.

Both models demonstrated sensitivity to artifacts, with the *smile* distortion having a notable degrading impact. To further assess data uncertainty, a sensitivity analysis could be conducted to evaluate the influence of potential misclassifications, measurement errors, or biases on model performance.

The *parameter estimates* assessment highlights the importance of incorporating uncertainty considerations into the training process of models, emphasizing the need for continuous evaluation. Specifically, planning uncertainty assessment before training is recommended to ensure robust parameter estimates. Further investigation could examine the influence of various hyperparameter choices and optimization techniques on model performance, providing additional insights into the uncertainties associated with *parameter estimates*.

Model structure uncertainty assessment is often observed in studies, offering direct accuracy measurements to compare models easily. Such an approach allows for a holistic understanding of the assessed uncertainties and facilitates drawing meaningful conclusions regarding the overall uncertainty of the models. However, this alone is not sufficient to fully assess model uncertainty.

In summary, assessing uncertainty as explained by Simmonds et al. (2022) is feasible. Notably, this approach considers uncertainty across all levels, including often overlooked uncertainty

aspects associated with ground truth investigation and *parameter estimates*. It is essential to acknowledge that in cases where favorable outcomes are attained, studies need to document the steps taken to reach those results so that users can reproduce them.

The framework provides a structured way of assessing uncertainty, with the absence of a straightforward assessment procedure. This allows the users to adapt and tailor the framework to their needs and circumstances. On the other hand, this may limit the user-friendliness of the framework for studies that do not solely focus on uncertainty. For this reason, the assessment table presented in Table 1.2 was needed.

Chapter 7

Conclusion

The present thesis has emphasized the significance of assessing uncertainty in ML models by systematically evaluating various sources. This comprehensive analysis encompassed the assessment of uncertainty in ground truth data, hyperspectral data, sensitivity to artifacts, training progress, and model structures for both SVM and CNN classifiers.

The examination of the dataset representation revealed variations in the geographical distribution of labels and tree ages, underscoring the heterogeneity within the dataset. Furthermore, the CNN classifiers exhibited greater sensitivity to artifacts than the SVM classifiers. At the same time, the *Deciduous* class was shown to be more robust towards noise than *Spruce* and *Pine*. Even minor *smile* distortions significantly degrade prediction accuracy, leading the classifiers to misclassify pixels in the spatial edges. Notably, when employing the 1D CNN, the presence of *smile* results in a significant drop of 76% in the accuracy of identifying tree species. This decline occurs within the first pixel of distortion. The analysis of converged models across different batches revealed that the SVM model might offer greater reliability, despite the CNN exhibiting higher median class accuracy.

The findings of this study underscore the presence of considerable remote sensing ML model uncertainty in this use case. Efforts are required to enhance the accuracy and robustness of the classifiers, particularly in mitigating the impact of artifacts. It is crucial for future research to prioritize the assessment of uncertainties, using the acquisition of reliable ground truth and hyperspectral data as a fundamental requirement for the comprehensive validation of models.

Appendix A

Code Created for Modelling Artifacts

A.1 Smile Distortion

```
def add_smile_distortion(img, a, b):
    # Get height, width of HSI imagery
    h, w = img.shape[:2]
    # Define center of image
    center = (h // 2, w // 2)

    # Create maps for x, y coordinates
    map_x, map_y = np.meshgrid(np.arange(img.shape[1]).reshape(-1, 1), np.
   arange(img.shape[0]).reshape(-1, 1))

    # Define the distortion function using a lambda function
    y_max = map_y.max() / 2
    x_max = map_x.max() / 2

    # Adds frown effect by combining smile effect in the vertical dir. w.
   frown effect in the horizontal dir.
    g = lambda x,y: -b*y**2 - b*y_max**2/2 - a*x**2 + a*x_max**2/2
    delta_x = g(map_x - center[1], map_y - center[0])
    max_delta_x = np.abs(delta_x).max()

    # Calculate the new_x values
    new_x = map_x + delta_x

    # Clip the values in the new_x array to fall within the bounds of the
   original image
    new_x = np.clip(new_x, 0, img.shape[1] - 1)

    # Create distorted image
```

```
    distorted_img = cv2.remap(img, new_x.astype(np.float32), map_y.astype(
    np.float32), cv2.INTER_CUBIC)

    # Return distorted HSI imagery and the maximum pixel shift
    return distorted_img, max_delta_x
```

A.2 Keystone Distortion

```
def add_keystone_distortion(img, a):
    # Get height, width of HSI imagery
    h, w = img.shape[:2]
    # Define center of image
    center_x, center_y = w // 2, h // 2

    # Create maps for x, y coordinates
    map_x, map_y = np.meshgrid(np.arange(w), np.arange(h))

    f = lambda x: a * x

    # Calculate delta_y
    delta_y = (map_y - center_y) * f(map_x - center_x)
    max_delta_y = np.abs(delta_y).max()
    # Add delta_y to map_y to create new_y
    new_y = map_y + delta_y

    # Clip new_y values to ensure they are within image bounds
    new_y = np.clip(new_y, 0, h-1)

    map_x, new_y = map_x.astype(np.float32), new_y.astype(np.float32)

    # Create distorted image
    distorted_img = cv2.remap(img, map_x, new_y.astype(np.float32), cv2.
    INTER_CUBIC)

    # Return distorted HSI imagery and the maximum pixel shift
    return distorted_img, max_delta_y
```

Bibliography

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R. & Sousa, J. J. (2017), 'Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry', *Remote Sensing* **9**(11).
URL: *https://www.mdpi.com/2072-4292/9/11/1110*

Albawi, S., Mohammed, T. A. & Al-Zawi, S. (2017), Understanding of a convolutional neural network, *in* '2017 International Conference on Engineering and Technology (ICET)', pp. 1–6.

ASI (2020), 'Prisma products specification document issue 2.3'.
URL: *http://prisma.asi.it/missionselect/docs/PRISMA%20Product%20Specifications_Is2_3.pdf*

Bakken, S., Orlandic, M. & Johansen, T. A. (2019), The effect of dimensionality reduction on signature-based target detection for hyperspectral remote sensing, *in* T. S. Pagano, C. D. Norton & S. R. Babu, eds, 'CubeSats and SmallSats for Remote Sensing III', Vol. 11131, International Society for Optics and Photonics, SPIE, p. 111310L.
URL: *https://doi.org/10.1117/12.2529141*

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. & Chanussot, J. (2013), 'Hyperspectral remote sensing data analysis and future challenges', *IEEE Geoscience and Remote Sensing Magazine* **1**(2), 6–36.

Breidenbach, J., Ellison, D., Petersson, H., Korhonen, K. T., Henttonen, H. M., Wallerman, J., Fridman, J., Gobakken, T., Astrup, R. & Næsset, E. (2022), 'Harvested area did not increase abruptly—how advancements in satellitebased mapping led to erroneous conclusions', *Annals of Forest Science* .
URL: *https://doi.org/10.1186/s13595-022-01120-4*

Breidenbach, J., Rahlf, J., Räty, J., Hauglin, M. & Bergseng, E. (2022), 'Validering av sr16 med kontrollmålinger i asker, alver og elverum/våler'.
URL: *https://www.nibio.no/tema/skog/kart-over-skogressurser/skogressurskart-sr16/_attachment/inline/b1351797-d448-4a67-b099-961efaa6bf80:*

*ee*86*cfca*94*ca*89*db*5*c*17610*fd*52*b*04*b*1*fc*6290*f*1/*Notat* − *Validering* − *av* − *SR*16 − *med* − *kontrollm*%*C*3%*A*5*linger* − *i* − *Asker* − *Alver* − *og* − *ElverumV*%*C*3%*A*5*ler.pdf*

Brownlee, J. (2019), 'How to choose a feature selection method for machine learning', *Machine Learning Mastery* **10**.

Brynjolfsson, E. & Mcafee, A. (2017), 'Artificial intelligence, for real', *Harvard business review* **1**, 1–31.

Campbell, J. & Wynne, R. (2011), *Introduction to Remote Sensing*, Vol. 5, The Guildford Press.

Centre, E. S. D. (2023), 'Lucas'.
URL: *https*://*esdac.jrc.ec.europa.eu*/*projects*/*lucas*

Chien, S., Silverman, D., Davies, A. G. & Mandl, D. (2009), 'Onboard science processing concepts for the hyspiri mission', *EEE Intelligent Systems* pp. 12–19.

Cogliati, S., Sarti, F., Chiarantini, L., Cosi, M., Lorusso, R., Lopinto, E., Miglietta, F., Genesio, L., Guanter, L., Damm, A. et al. (2021), 'The prisma imaging spectroscopy mission: Overview and first performance analysis', *Remote sensing of environment* **262**, 112499.

Coops, N. C., Wulder, M. A., White, J. C. et al. (2007), 'Identifying and describing forest disturbance and spatial pattern: data selection issues and methodological implications', *Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches* p. 264.

Czapla-Myers, J., Ong, L., Thome, K. & McCorkel, J. (2016), 'Validation of eo-1 hyperion and advanced land imager using the radiometric calibration test site at railroad valley, nevada', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(2), 816–826.

Dawson, C. W. & Wilby, R. (1998), 'An artificial neural network approach to rainfall-runoff modeling', *Hydrological Sciences Journal* **43**(1), 47–66.
URL: *https*://*doi.org*/10.1080/02626669809492102

Edeling, W., Arabnejad, H., Sinclair, R., Suleimenova, D., Gopalakrishnan, K., Bosak, B., Groen, D., Mahmood, I., Crommelin, D. & Coveney, P. V. (2021), 'The impact of uncertainty on predictions of the covidsim epidemiological code', *Nature Computational Science* **1**(2), 128–135.

ESA (2023), 'Sentinel-2 mission guide'.
URL: *https*://*sentinel.esa.int*/*web*/*sentinel*/*missions*/*sentinel* − 2

Ewald, F., Kölling, T., Baumgartner, A., Zinner, T. & Mayer, B. (2016), 'Design and characterization of specmacs, a multipurpose hyperspectral cloud and sky imager', *Atmospheric Measurement Techniques* **9**, 2015–2042.

Fernandes, M. R., Aguiar, F. C., Martins, M. J., Rico, N., Ferreira, M. T. & Correia, A. C. (2020), 'Carbon stock estimations in a mediterranean riparian forest: A case study combining field data and uav imagery', *Forests* **11**(4), 376.

for Nature, R. (2016), 'Trees – a lifespan approach'.
URL: *http://www.ancienttreeforum.co.uk/wp-content/uploads/2017/04/Trees-a-lifespan-approach-Nev-Fay-et-al.pdf*

Galeazzi, C., Sacchetti, A., Cisbani, A. & Babini, G. (2008*a*), 'The prisma program', *Geoscience and Remote Sensing Symposium* **4**, 105–108.

Galeazzi, C., Sacchetti, A., Cisbani, A. & Babini, G. (2008*b*), The prisma program, *in* 'IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium', Vol. 4, pp. IV – 105–IV – 108.

Genesio, L., Braga, F., Bresciani, M., Boschetti, M., Carotenuto, F., Cogliati, S., Colella, S., Colombo, R., Giardino, C., Gioli, B., Lopinto, E., Meloni, D., Pepe, M., Pascucci, S., Pignatti, S., Pompilio, L., Sacco, P., Satalino, G. & Miglietta, F. (2022), Updates on prisma: Scientific calibration/validation activities and supporting studies, *in* 'IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium', pp. 4585–4586.

Geography, G. (n.d.), 'Multispectral vs hyperspectral imagery explained'.
URL: *https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/*

Giardino, C. & Brando, V. (2023), 'Prisma data are now available for access'.
URL: *https://sbg.jpl.nasa.gov/news-events/prisma-data-are-now-available-for-access*

Glenn, N. F., Mundt, J. T., Weber, K. T., Prather, T. S., Lass, L. W. & Pettingill, J. (2005), 'Hyperspectral data processing for repeat detection of small infestations of leafy spurge', *Remote Sensing of Environment* **95**(3), 399–412.

Graham, B. (2014), 'Fractional max-pooling'.
URL: *https://arxiv.org/abs/1412.6071*

Grøtte, M., Birkeland, R., Honore-Livermore, E., Bakken, S., Garrett, J., Prentice, E., Sigernes, F., Orliandic, M. & Gravdahl, J. (2021), 'Ocean color hyperspectral remote sensing with high resolution and low latency—the hypso-1 cubesat mission', *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–19.

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S. & Colombo, R. (2021), 'Mapping methane point emissions with the prisma spaceborne imaging spectrometer', *Remote Sensing of Environment* **265**, 112671.

Guarini, R., Loizzo, R., Facchinetti, C., Longo, F., Ponticelli, B., Faraci, M., Dami, M., Cosi, M., Amoruso, L., De Pasquale, V. et al. (2018), Prisma hyperspectral mission products, *in* 'IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium', IEEE, pp. 179–182.

Guarini, R., Loizzo, R., Longo, F., Mari, S., Scopa, T. & Varacalli, G. (2017), Overview of the prisma space and ground segment and its hyperspectral products, *in* '2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)', pp. 431–434.

Hasan, H., Shafri, H. Z. & Habshi, M. (2019), A comparison between support vector machine (svm) and convolutional neural network (cnn) models for hyperspectral image classification, *in* 'IOP Conference Series: Earth and Environmental Science', Vol. 357, IOP Publishing, p. 012035.

Hasinoff, S. W. (2014), 'Photon, poisson noise.', *Computer Vision, A Reference Guide* **4**, 16.

Heller Pearlshtien, D., Pignatti, S., Greisman-Ran, U. & Ben-Dor, E. (2021), 'Prisma sensor evaluation: a case study of mineral mapping performance over makhtesh ramon, israel', *International Journal of Remote Sensing* **42**(15), 5882–5914.

Henriksen, M. B. (2019), Hyperspectral imager calibration and image correction, Master's thesis, NTNU.

Henriksen, M. B., Garrett, J. L., Prentice, E. F., Stahl, A., Johansen, T. A. & Sigernes, F. (2019), Real-time corrections for a low-cost hyperspectral instrument, *in* '2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)', pp. 1–5.

Hermosilla, T., Bastyr, A., Coops, N. C., White, J. C. & Wulder, M. A. (2022), 'Mapping the presence and distribution of tree species in canada's forested ecosystems', *Remote Sensing of Environment* **282**, 113276.

Hillenbrand, M., Grewe, A., Bichra, M., Mitschunas, B., Kirner, R., Weiß, R. & Sinzinger, S. (2012), 'Chromatic information coding in optical systems for hyperspectral imaging and chromatic confocal sensing', *Optical Systems Design* .
URL: *https://doi.org/10.1117/12.981147*

Hong-Xia, D., Lu, X.-S., Wang, C., Shen, H.-Z., Zhuo, Y.-W. & Deng, L.-J. (2022), 'Patchmask: A data augmentation strategy with gaussian noise in hyperspectral images.', *Remote Sensing* **14**.
URL: *https://doi.org/10.3390/rs14246308*.

Hsieh, T.-H. & Kiang, J.-F. (2020), 'Comparison of cnn algorithms on hyperspectral image classification in agricultural lands', *Sensors* **20**(6), 1734.

Jablonski, J., Durell, C., Slonecker, T., Wang, K., Simon, B., Eichelberger, A. & Osterberg, J. (2016), 'Best practices in passive remote sensing vnir hyperspectral system hardware calibrations', pp. 1–31.

Johnson, D. H. (2006), 'Signal-to-noise ratio', *Scholarpedia* **1**(12), 2088. revision #126771.

Kairu, E. (1982), 'An introduction to remote sensing', *GeoJournal* **6**, 251–260.

Katila, M. (2004), 'Controlling the estimation errors in the finnish multisource national forest inventory'.

Kerekes, J. & Baum, J. (2003), 'Hyperspectral imaging system modeling', *Lincoln Laboratory Journal* **14**.

Kreps, S. E. & Kriner, D. L. (2020), 'Model uncertainty, political contestation, and public trust in science: Evidence from the covid-19 pandemic', *Science advances* **6**(43), eabd4563.

Landgrebe, D. A. & Malaret, E. (1986), 'Noise in remote-sensing systems: The effect on classification error', *IEEE Transactions on Geoscience and Remote Sensing* **GE-24**(2), 294–300.

Loehle, C. (1988), 'Tree life history strategies: the role of defenses', *Canadian Journal of Forest Research* **18**(2), 209–222.

Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C. & Varacalli, G. (2018), Prisma: The italian hyperspectral mission, *in* 'IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium', pp. 175–178.

Long, T., Jiao, W., He, G. & Zhang, Z. (2016), 'A fast and reliable matching method for automated georeferencing of remotely-sensed imagery', *Remote sensing* **8**(1), 56.

Luke (2021), 'Luke mvmi lateuspalvelu'.
URL: *https://kartta.luke.fi/*

Mavroforakis, M. & Theodoridis, S. (2006), 'A geometric approach to support vector machine (svm) classification', *IEEE Transactions on Neural Networks* **17**(3), 671–682.

Mazet, V. (2023), 'Basics of image processing'.
URL: *https://vincmazet.github.io/bip/restoration/denoising.html*

Memon, M. (2022), 'Ann vs cnn vs rnn: Neural networks guide'.
URL: *https://levity.ai/blog/neural-networks-cnn-ann-rnn*

Mouroulis, P., Green, R. O. & Chrien, T. G. (2000), 'Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information', *Appl. Opt.* **39**(13), 2210–2220.
URL: $https://opg.optica.org/ao/abstract.cfm?URI = ao-39-13-2210$

Netteland, S. (2022), Exploration and implementation of large cnn models for image segmentation in hyperspectral cubesat missions, Master's thesis, NTNU.

nfis (2020), 'National terrestrial ecosystem monitoring system for canada'.
URL: $https://opendata.nfis.org/mapserver/nfis-change_eng.html$

NIBIO (2022), 'Skogressurskart (sr16)'.
URL: $https://www.nibio.no/tema/skog/kart-over-skogressurser/skogressurskart-sr16$

Noble, W. S. (2006), 'What is a support vector machine?', *Nature biotechnology* **24**(12), 1565–1567.

Oche, P. A., Ewa, G. A. & Ibekwe, N. (2021), 'Applications and challenges of artificial intelligence in space missions', *IEEE Access* pp. 1–1.

Ohgi, N. & A., I. (2010), 'Japanese hyper-multi spectral mission', *IEEE International Geoscience and Remote Sensing Symposium* .

O'Shea, K. & Nash, R. (2015), 'An introduction to convolutional neural networks', *CoRR* **abs/1511.08458**.
URL: $http://arxiv.org/abs/1511.08458$

Pratihast, A. K., DeVries, B., Avitabile, V., De Bruin, S., Kooistra, L., Tekle, M. & Herold, M. (2014), 'Combining satellite data and community-based observations for forest monitoring', *Forests* **5**(10), 2464–2489.

QGIS (2023), 'Welcome to qgis'.
URL: $https://www.qgis.org/en/site/$

Rahlf, J., Hauglin, M. & Breidenbach, J. (2021), 'Sammenligning av sr16 treslagskart med en skogbruksplan i tjeldsund, sør-troms'.
URL: $https://www.nibio.no/tema/skog/kart-over-skogressurser/skogressurskart-sr16/_attachment/inline/b1351797-d448-4a67-b099-961efaa6bf80:3b8b5737b9f49aa3196cffa3e55cbc560c968788/Treslag_SR16_Troms_4.pdf$

Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G. & Chanussot, J. (2018), 'Noise reduction in hyperspectral imagery: Overview and application', *Remote Sensing* **10**, 482.

Riese, F. M. & Keller, S. (2019), 'Soil texture classification with 1d convolutional neural networks based on hyperspectral data', *arXiv preprint arXiv:1901.04846* .

Roli, F. & Fumera, G. (2001), Support vector machines for remote sensing image classification, *in* 'Image and signal processing for remote sensing VI', Vol. 4170, SPIE, pp. 160–166.

Romaniello, V., Silvestri, M., Buongiorno, M. F. & Musacchio, M. (2020), 'Comparison of prisma data with model simulations, hyperion reflectance and field spectrometer measurements on 'piano delle concazze'(mt. etna, italy)', *Sensors* **20**(24), 7224.

Ronneberger, O., Fischer, P. & Brox, T. (2015), 'U-net: Convolutional networks for biomedical image segmentation', *CoRR* **abs/1505.04597**.
URL: *http*://*arxiv.org*/*abs*/1505.04597

Rossmann, K. (1969), 'Point spread-function, line spread-function, and modulation transfer function', *Radiology* **93**(2), 257–272. PMID: 5822701.
URL: *https*://*doi.org*/10.1148/93.2.257

Sang, B., Schubert, J., Kaiser, S., Mogulsky, V., Neumann, C., Förster, K., Hofer, S., Stuffler, T., Kaufmann, H., Müller, A., Eversberg, T. & C., C. (2008), 'The enmap hyperspectral imaging spectrometer: instrument concept, calibration and technologies'.

scikit learn (2023), 'scikit-learn machine learning in python'.
URL: *https*://*scikit* − *learn.org*/*stable*/

Scott, J. C. (1998), *Seeing like a state: How certain schemes to improve the human condition have failed*, yale university Press.

Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C. et al. (2019), 'Arthropod decline in grasslands and forests is associated with landscape-level drivers', *Nature* **574**(7780), 671–674.

Sexton, J., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K., Channan, S., Dimiceli, C. & Townshend, J. (2013), 'Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of modis vegetation continuous fields with lidar-based estimates of error', *International Journal of Digital Earth* **6**, 1–22.

Shi, D. & Yang, X. (2015), *Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery*, Springer, pp. 265–279.

Sigernes, F., Syrjäsuo, M., Storvold, R., Fortuna, J., Grøtte, M. E. & Johansen, T. A. (2018), 'Do it yourself hyperspectral imager for handheld to airborne operations', *Opt. Express* **26**.

Simmonds, E. G., Adjei, K. P., Andersen, C. W., Aspheim, J. C. H., Battistin, C., Bulso, N., Christensen, H., Cretois, B., Cubero, R., Davidovich, I. A., Dickel, L., Dunn, B., Dunn-Sigouin, E., Dyrstad, K., Einum, S., Giglio, D., Gjerlow, H., Godefroidt, A., Gonzalez-Gil, R., Cogno, S. G., Grosse, F., Halloran, P., Jensen, M. F., Kennedy, J. J., Langsaether, P. E., Laverick, J. H., Lederberger, D., Li, C., Mandeville, E., Mandeville, C., Moe, E., Schroder, T. N., Nunan, D., Parada, J. S., Simpson, M. R., Skarstein, E. S., Spensberger, C., Stevens, R., Subramanian, A., Svendsen, L., Theisen, O. M., Watret, C. & OHara, R. B. (2022), 'How is model-related uncertainty quantified and reported in different disciplines?'.
URL: *https://arxiv.org/abs/*2206.12179

SLU (2023), 'Slu forest map'.
URL: *https://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/foreststatistics/slu-forest-map/*

Spies, T. A., Franklin, J. F. et al. (1991), 'The structure of natural young, mature, and old-growth douglas-fir forests in oregon and washington', *Wildlife and vegetation of unmanaged Douglas-fir forests* pp. 91–109.

S&T, NTNU, U. (2021), Esa iod proposal, faubai: Fpga acceleration for push broom satellite imagery, Technical report, ESA.

Talbot, H., Phelippeau, H., Akil, M. & Bara, S. (2009), Efficient poisson denoising for photography, *in* '2009 16th IEEE International Conference on Image Processing (ICIP)', IEEE, pp. 3881–3884.

Torkildsen, H. E. & Skauli, T. (2018), 'Full characterization of spatial coregistration errors and spatial resolution in spectral imagers', *Optics Letters* **43**(16), 3814–3817.

van Beers, W. C. M. & Kleijnen, J. P. C. (2003), 'Kriging for interpolation in random simulation', *The Journal of the Operational Research Society* **54**(3), 255–262.
URL: *http://www.jstor.org/stable/*4101619

Volodina, V. & Challenor, P. (2021), 'The importance of uncertainty quantification in model reproducibility', *Philosophical Transactions of the Royal Society A* **379**(2197), 20200071.

Wallerman, J., Axensten, P., Egberth, M., Janzén, J., Sandström, E., Fransson, J. E. & Nilsson, M. (2021), Slu forest map-mapping swedish forests since year 2000, *in* '2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS', IEEE, pp. 6056–6059.

WindPRO (2023), 'Norwegian elevation models'.
URL: *https://help.emd.dk/mediawiki/index.php?title=Norwegian_Elevation_Models*

Wolfewicz, A. (2022), 'Deep learning vs. machine learning – what's the difference?'.
URL: $https://levity.ai/blog/difference-machine-learning-deep-learning$

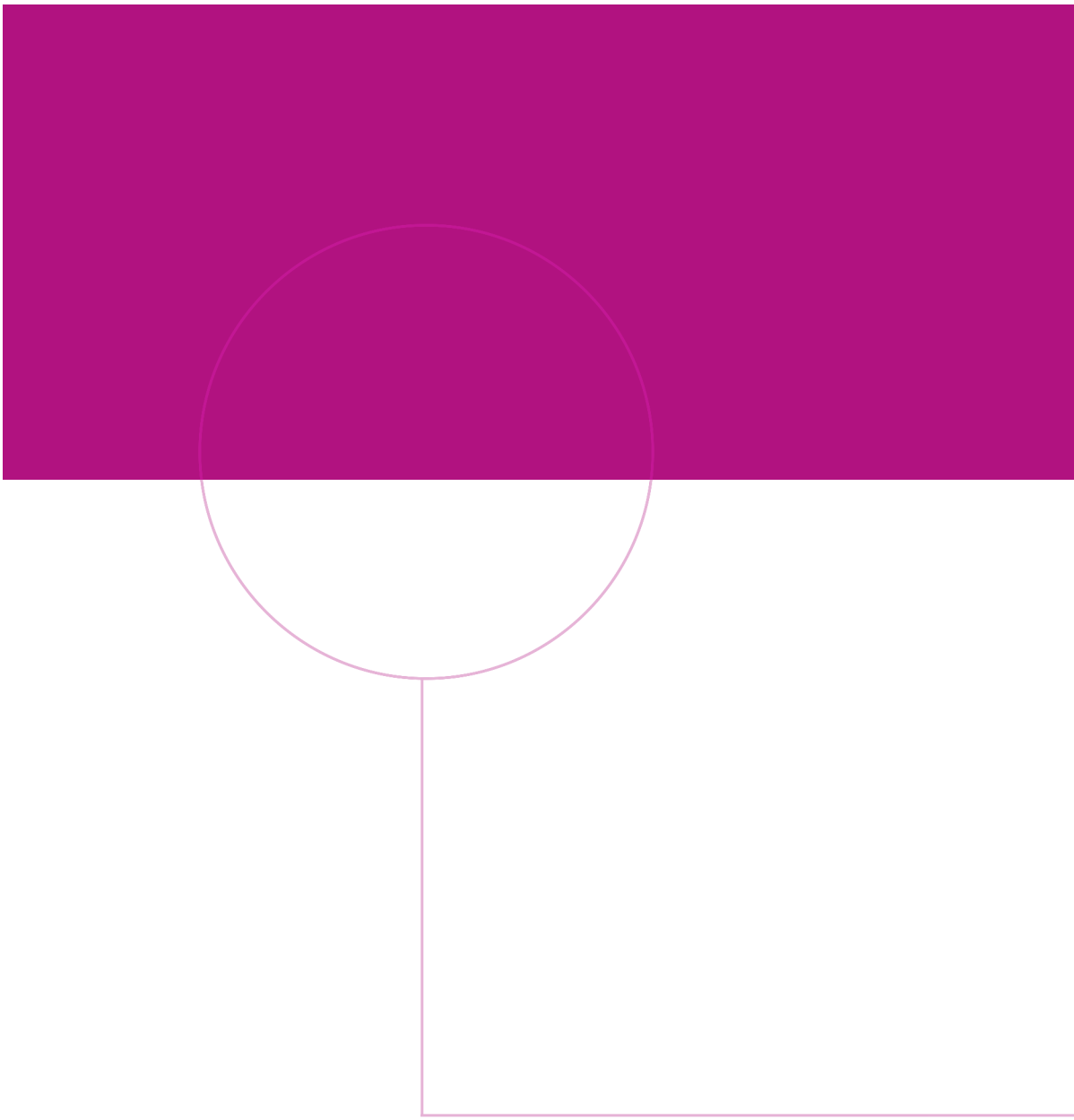
Yang, X., Ye, Y., Li, X., Lau, R., Zhang, X. & Huang, X. (2018), 'Hyperspectral image classification with deep learning models', *IEEE Transactions on Geoscience and Remote Sensing* **PP**, 1–16.

Yi, D. & Choi, B. (2010), Chapter 5 - nonlinear partial differential equations for noise problems, *in* 'Advances in Imaging and Electron Physics', Vol. 164, Elsevier, pp. 329–343.
URL: *https://doi.org/10.1016/B978-0-12-381312-1.00005-5*.

Yiwei, M., Betters, C. H., Evans, B., Artlett, C. P., Leon-Saval, S. G., Garske, S., Cairns, I. H., Cocks, T., Winter, R., & Dell, T. (2022), 'Openhsi: A complete open-source hyperspectral imaging solution for everyone', *Remote Sensing* **14**(9).
URL: *https://doi.org/10.3390/rs14092244*

Yokoya, N., Miyamura, N. & Iwasaki, A. (2010), 'Preprocessing of hyperspectral imagery with consideration of smile and keystone properties', *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications III* pp. 1–10.

Zhao, X., Zhang, J., Huang, Y., Tian, Y. & Yuan, L. (2022), 'Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis', *Computers and Electronics in Agriculture* **193**, 106717.



NTNU

Norwegian University of
Science and Technology