

INVESTIGATION OF TARGET DETECTION IN HYPERSPECTRAL IMAGES FOR  
AIDED TARGET RECOGNITION

by

Mazharul Hossain

A Dissertation Proposal

Submitted for the Committee's Review

Academic Advisor:

Dr. Lan Wang and Dr. Aaron L Robinson

Major: Computer Science

The University of Memphis

May 2025

## Abstract

Aided Target Recognition (AiTR) improves human target detection and tracking by enhancing sensor data for better accuracy and visual understanding. Remote sensing (RS) collects information from a distance, often via satellites or aircraft. We are exploring a network of unmanned aerial vehicles (UAVs) equipped with various sensors to enhance Intelligence, Surveillance, and Reconnaissance (ISR), agriculture, and environmental assessment applications. The combination of AiTR and remote sensing is a key advancement in military surveillance, and research is ongoing in this area. Our research involves the development of algorithms for target detection in collected hyperspectral sensor data. A key challenge we face is the need for lightweight CPU-based anomaly detection to reduce computation overhead and battery consumption, a crucial aspect in designing an optimized UAV surveillance system for long, steady flight missions with broad area coverage. We must also further investigate those scenes containing the anomalies on more capable drones or powerful ground stations.

Hyperspectral (HS) imaging, a powerful tool for capturing detailed spectral information, is susceptible to variations in data quality due to weather conditions. Factors such as time of day, cloud cover, dust, and water vapor can introduce noise and scattering, underscoring the necessity of image preprocessing techniques. These techniques, including noise reduction, brightness correction, and contrast enhancement, are crucial for improving data quality, bolstering feature extraction, and enhancing the performance of standard detection methods and machine learning models.

Statistical methods have been used extensively for anomaly detection. They typically make assumptions about the background in an HS image to detect anomalies. However, they can fail when the background does not meet their assumptions. In contrast, machine learning (ML) techniques can enhance anomaly detection in cluttered backgrounds. Thus, we propose a supervised stacking ensemble called GE-AD to combine the outputs from multiple statistical methods and ML algorithms (i.e., base methods to GE-AD). The selection of these base methods

is crucial, and we employ a greedy search algorithm to identify the most effective ones for the ensemble. However, GE-AD has difficulty handling HS datasets collected using diverse sensors or datasets collected using the same sensor but under different weather patterns, changed collection techniques, and other uncertainties. I propose generalizing our anomaly detection solution using domain adaptation to address this. Domain Adaptation (DA) is an ML technique that assumes no access to target domain data. It aims to learn a generalized model to differentiate between datasets from one or several training domains with different probability distributions, thereby achieving good out-of-distribution generalization.

Semantic segmentation proves to be a valuable asset when applications necessitate a more in-depth analysis of a scene for precise object localization and boundary delineation, such as in military surveillance scenarios. It allows us to comprehend the detailed composition of a scene by distinguishing different objects at a granular level. While high-performing image semantic segmentation models like U-Net, FCN, and DeepLab are readily available, their direct use on HS imager data, which has more channels than the RGB images they were designed and trained with, is not feasible. Simply updating the input layers of existing deep learning (DL) models may compromise their performance. Therefore, I propose a separate channel attention module to extract crucial spectral information from the hundreds of channels in HS images without disrupting the original DL model. By combining the spectral feature in the first layers of the original DL model and fine-tuning it, we can achieve improved results, thereby demonstrating the adaptability of this method for AiTR's use.

**Keywords:** hyperspectral remote sensing, near-infrared NIR, unmanned aerial vehicles UAV, aided target recognition AiTR, anomaly detection methods, machine learning, deep learning, crop classification, semantic segmentation, stacking ensemble learning

## Table of Contents

Chapter	Page
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Goals and Future plan	4
1.2 Background	5
1.2.1 Hyperspectral Imaging	5
1.2.2 Aided Target Recognition (AiTR)	6
1.2.3 Potential Applications of AiTR	8
1.2.4 Hyperspectral Unmixing	8
1.2.5 Hyperspectral Anomaly Detection (HS-AD)	9
1.2.6 Generalization of Detection	9
1.2.7 Hyperspectral Image Semantic Segmentation	10
1.3 Organization	10
2 Hyperspectral Image Preprocessing	12
2.1 Previous Work	12
2.1.1 Initial HS Image Preprocessing	12
2.1.2 Data Normalization for Machine Learning	13
2.1.3 Image Processing Algorithms for Visual Enhancement	14
2.1.4 Performance Evaluation Metrics	15
2.2 Preliminary Work	15
2.2.1 Preprocessing Methods to Improve Hyperspectral AD Algorithms for Arizona Dataset	15
2.2.2 Normalization of Anomaly Detection Results	16
2.2.3 Image Enhancement Methods for Human Understandability	18
2.3 Further Research Plan	18
2.3.1 Process HS images to Improve Machine Detection Accuracy	18
2.3.2 Image Enhancement Methods for Human Understandability	19
3 Hyperspectral Anomaly detection	22
3.1 Previous Work	22
3.1.1 Individual HS-AD Algorithms	22
3.1.2 Ensemble HS-AD Algorithms	25
3.1.3 Hyperspectral Unmixing-Based Voting Ensemble Anomaly Detector (HUE-AD)	29
3.2 Greedy Ensemble Anomaly Detection (GE-AD)	31
3.3 Evaluation	36
3.3.1 Datasets	36
3.3.2 Evaluation Metrics	38

3.3.3	Comparison with Individual AD Methods	39
3.3.4	Comparison with Baseline SOTA Ensemble Methods	44
4	Generalization of Hyperspectral Anomaly detection	45
4.1	Previous Work	47
4.2	Preliminary Work	49
4.2.1	Replacing Supervised Meta Method in GE-AD with Unsupervised Method	49
4.2.2	Adding Spectral Information as Input to GE-AD	49
4.2.3	Evaluation Results	49
4.3	Further Research Plan	53
5	Hyperspectral Image Semantic Segmentation	55
5.1	Previous Work	57
5.2	Preliminary Work	58
5.3	Preliminary Results	59
5.4	Further Research Plan	60
6	Proposed Work Timeline	61
	References	63

## List of Tables

		Page
1	Examples from various types of HS-AD algorithms.	23
2	Feature comparison of ensemble learning algorithms.	26
3	Comparison of F1-macro scores showing our proposed ensemble method (GE-AD) performs better compared to HUE-AD, and individual methods using the ABU dataset.	42
4	Performance results of all methods regarding AUC metric (%) along with Friedman rank using the ABU-airport dataset.	42
5	ABU-IV quantitative performance (as reported in the authors' papers [54, 56, 57]) comparison between our proposed ensemble method (GE-AD) and other ensemble methods. Here, the best scores are shown in bold.	44
6	The average ROC-AUC scores of our newly proposed Unsupervised ensemble UGE-AD over various datasets. The first column shows the dataset used in the greedy search to find suitable base methods, and the second column shows those base methods. Where the dataset names in rows and columns match, those cells show the training score after the greedy search. The boldface marks the best performance in each test dataset column.	50
7	The average and standard deviation of our ensemble methods' ROC-AUC scores computed over the test datasets. We generated this table using a model trained on randomly selected, at most, fifty percent of the datasets from the ABU, Arizona, HYDICE urban, Salinas, and San Diego datasets as training and evaluated on other datasets. The boldface marks the best performance in each column.	51
8	The average and standard deviation of our ensemble methods' F1-macro scores were computed over the test datasets. We generated this table using a model trained on randomly selected, at most, fifty percent of the datasets from the ABU, Arizona, HYDICE urban, Salinas, and San Diego datasets as training and evaluated on other datasets. The boldface marks the best performance in each column.	52
9	Timeline to complete my proposed dissertation works.	62

## List of Figures

Figure		Page
1	Generic algorithm processing pipeline for an AiTR system [1].	2
2	The Electromagnetic spectrum [6].	5
3	A generic scheme of HS imager mapping a scene identifying soil, vegetation and water through reflectance signature [4].	6
4	AiTR is helping identifying enemy vehicles in military applications.	7
5	Example of plant diseases and pests detection problem [10].	8
6	Positive impact of removing noisy bands from Image V in the Arizona dataset.	16
7	Irregular results from kernel isolation forest anomaly detection are transformed into a normal distribution. (a) Distribution of Original data. (b) Irregularities. (c) Transformed normal distribution.	17
8	Normalization can give a boost in performance for tricky data like ABU-IV. (a) RGB, (b) ground truth, (c) GE-AD with no normalization ( $F_1 = 0.727$ ), (d) GE-AD with z score normalization ( $F_1 = 0.744$ ), (e) GE-AD with quantile normalization ( $F_1 = 0.757$ ).	17
9	Example of digital image processing. From left to right, it shows the original image, the image applied auto brightness adjust (SSIM 0.685), and the image processed using HSV CLAHE (SSIM 0.677).	19
10	Visual comparison showing quantitative improvement of removing noisy bands from Florida Image 1. The green line shows SNR in [dB] after deleting noisy bands, whereas the blue line shows the dropped channels.	20
11	Image Enhancement Method Selection using ML Model	21
12	The complete method to get target abundance using target endmember [14]. A target endmember is a specific reference spectrum extracted from a reference HS image. The process of extracting target abundance is similar to object detection.	29
13	Design for Hyperspectral Unmixing-Based Voting Ensemble Anomaly Detector (HUE-AD).	30
14	The architecture of a stacked model, where the new training data consist of the outputs of the first-level models.	32
15	Overall flowchart of GE-AD to find the best AD methods for ensemble fusion.	33
16	Flowchart of the greedy search to find the best AD methods for ensemble fusion.	34
17	Visual comparison showing better performance of our proposed ensemble method (GE-AD) compared to other methods using ABU-IV data. (a) RGB, (b) ground truth, (c) GE-AD ( $F_1 = 0.856$ ), (d) HUE-AD ( $F_1 = 0.649$ ), (e) Abundance ( $F_1 = 0.735$ ), (f) AED ( $F_1 = 0.610$ ), (g) KIFD ( $F_1 = 0.587$ ), (h) KRX ( $F_1 = 0.604$ ), (i) LSUNRSORAD ( $F_1 = 0.569$ ), (j) FCBAD ( $F_1 = 0.631$ ), (k) ERRX MF ( $F_1 = 0.666$ ), (l) ERCRD ( $F_1 = 0.545$ ), and (m) SED ( $F_1 = 0.592$ ). Please note that, as the implementations of algorithms for Figures (k–m) were not publicly available, we binarized the original outputs from the authors' papers [54, 56, 57], respectively.	40
18	Detection accuracy evaluation through ROC curve and AUC scores (shown in the legend) of (a) ABU-I, (b) ABU-II, (c) ABU-III, and (d) ABU-IV.	43

19	The visualization shows the generalization of our proposed ensemble method (GE-AD) trained on the ABU dataset compared to other methods using San Diego-02 data. <b>(a)</b> RGB, <b>(b)</b> ground truth, <b>(c)</b> GE-AD ( $F_1 = 0.826$ ), <b>(d)</b> Abundance ( $F_1 = 0.655$ ), <b>(e)</b> AED ( $F_1 = 0.636$ ), <b>(f)</b> FCBAD ( $F_1 = 0.536$ ), and <b>(g)</b> KRX ( $F_1 = 0.627$ ).	45
20	The visualization shows the performance of our proposed ensemble method (GE-AD) trained on the ABU dataset compared to other methods using Arizona-V data. <b>(a)</b> RGB, <b>(b)</b> Ground Truth, <b>(c)</b> GE-AD ( $F_1 = 0.490$ ), <b>(d)</b> Abundance ( $F_1 = 0.701$ ), <b>(e)</b> AED ( $F_1 = 0.584$ ), <b>(f)</b> FCBAD ( $F_1 = 0.810$ ), <b>(g)</b> KRX ( $F_1 = 0.503$ ).	46
21	Visualization of difference of feature importance between models.	47
22	Average ROC-AUC scores of our UGE-AD over various datasets showing consistent performance against various training datasets. The X-axis shows the dataset used in the greedy search to find suitable base methods. The Y-axis shows the testing ROC AUC scores. Except where the dataset names in rows and columns match, those bars show the training score after the greedy search.	51
23	The visualization shows the qualitative difference between our proposed ensemble method (UGE-AD) and our other methods using ABU-airport IV data. <b>(a)</b> RGB, <b>(b)</b> ground truth, <b>(c)</b> GE-AD, <b>(d)</b> HUE-AD, <b>(e)</b> UGE-AD.	53
24	The initial design for generalized GE-AD.	54
25	Example of Crop mapping [114].	56
26	Overview of the TransUNet framework [118]: (a) schematic of the Transformer layer; (b) architecture of the TransUNet.	57
27	A block diagram of the Attention U-Net [121] segmentation model. Input image is progressively filtered and downsampled by factor of 2 at each scale in the encoding part of the network (e.g. $H_4 = H_1 / 8$ ). $N_c$ denotes the number of classes. Attention gates (AGs) filter the features propagated through the skip connections.	58
28	Channel Attention Module [123].	58
29	The visualization shows an example of semantic segmentation using an image from large UAV hyperspectral agricultural imagery datasets. <b>(a)</b> RGB, <b>(b)</b> Ground Truth, <b>(c)</b> Trans-U-Net [118] fine-tuned on this agricultural RGB dataset (dice coefficient 0.13), <b>(d)</b> U-Net [105] trained on this agricultural HS image dataset (dice coefficient 0.06).	60

## Chapter 1

### Introduction

#### **Hyperspectral (HS) imaging helps analyze materials for Aided Target Recognition**

(AiTR) in military, agricultural, and environmental fields. AiTR [1] systems play a crucial role in enhancing how accurately we can understand sensor data for target detection, classification, and identification and assist in tracking. It involves preprocessing sensor data to improve detection accuracy, detecting and classifying potential targets [2], and enhancing the visual effects of data for human understandability. Whereas **Remote sensing (RS)** is gathering information about an area or object from a distance, typically using a satellite or aircraft. However, using these solutions on unmanned aerial vehicles (UAVs) is limited by UAVs' computing power and battery life. We are investigating the development of a network of multiple UAVs with different sensors affixed to each [3]. This network of UAVs aims to enhance imaging and reporting capabilities for intelligence, surveillance, and reconnaissance (ISR), potentially transforming military operations by providing better situational awareness and a clearer picture of what is happening. These UAVs or drones can also help in agriculture and environmental assessment applications.

**HS imaging (HSI)** is a promising tool with the potential to obtain detailed spectral information from a scene. The advancement of HSI technologies [4] combined with machine learning (ML) recognition techniques [5] holds the key to improved scene characterization in numerous applications. This scene characterization includes the detection of anomalies in a cluttered background, finding objects, or identifying materials. HS remote sensing combines spectroscopy and imaging to collect spatial and spectral information and process information across the electromagnetic spectrum. The advent of UAVs indicates that combining HS remote sensing with AiTR promises hope. Figure 1 shows various steps in the generic algorithm pipeline of an AiTR system. The top line shows different stages of processing on sensor data. The bottom pipeline shows the final AiTR stage: processing and presenting them to humans in the loop.

Our research involves the development of algorithms for target detection in collected HS

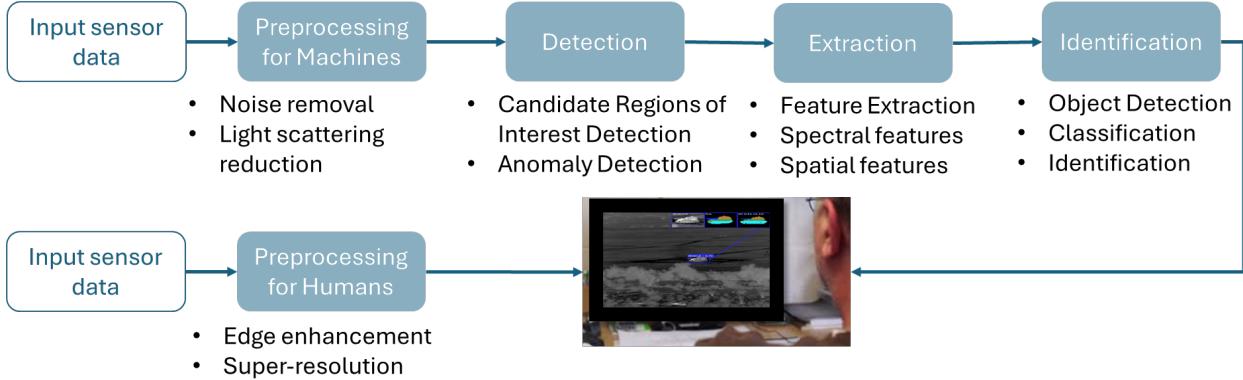


Figure 1: Generic algorithm processing pipeline for an AiTR system [1].

sensor data. A key challenge is that the AiTR's detection performance can degrade due to sensor variations, acquisition protocols, environmental conditions, and weather issues. HS imaging is excellent for spotting material-level issues. However, it's sensitive to environmental noise like clouds, dust, and changes in lighting. Thus, it underscores the necessity of image preprocessing techniques. These techniques, including noise reduction, brightness correction, and contrast enhancement, are crucial for improving data quality. It bolsters feature extraction and enhances the performance of standard detection methods and machine learning models. Another challenge is that standard anomaly detection (AD) methods, especially statistical ones, often struggle when design assumptions are unmet. On the other hand, deep learning models made for regular RGB images do not easily adapt to high-dimensional hyperspectral data.

We propose a lightweight, CPU-efficient anomaly detection framework that combines traditional statistical techniques with machine learning algorithms using a supervised stacking ensemble named GE-AD to address these limitations. A lightweight CPU-based anomaly detection is vital to reduce computation overhead and battery consumption, a crucial aspect in designing an optimized UAV surveillance system for long, steady flight missions with wide area coverage. A greedy search strategy is employed to select optimal base detectors, enhancing ensemble performance under constrained computational resources. However, due to spectral variability and acquisition inconsistencies, GE-AD's performance declines across domains.

To overcome this, we integrate unsupervised domain adaptation methods that align feature distributions between training and deployment domains, enabling robust, out-of-distribution generalization without requiring labeled target data.

Furthermore, we extend our approach with semantic segmentation for tasks requiring fine-grained scene understanding for AiTR identification. Recognizing that existing segmentation models (e.g., U-Net, DeepLab) are not optimized for HS inputs, we introduce a channel attention module to extract relevant spectral features without disrupting the original architecture. By fusing this module with early layers of deep segmentation models and task adaptation, we preserve their performance while adapting them to the rich spectral space of HS data.

Finally, our task-oriented digital image processing enhances human visibility to detect targets and better understand the scene. We can apply these image-processing techniques to 1- or 3-channel digital images (1-channel gray-scale and NIR images or 3-channel RGB and NIR-RG images). Compared to generic image processing, which can introduce artifacts and hinder scene understanding, our proposed framework will significantly improve scene understanding, anomaly detection, and semantic segmentation performance under real-world constraints, enabling effective, long-duration UAV operations across diverse sensing conditions.

This proposal aims to solve AiTR problems for HS data collected using low-flying UAVs, Unmanned Aircraft System (UAS) platforms, and aircraft by providing anomaly detection and segmentation solutions based on HS remote sensing. Here, UAV (Unmanned Aerial Vehicle) refers only to the drone or unmanned aircraft itself, whereas UAS (Unmanned Aircraft System) encompasses the whole system needed to operate a drone, including the aircraft, supporting ground control station and systems, communication links, and software. This work bridges the gap between high-performance HS analysis and domain-robust AiTR systems. The individual components are also valuable and advance the research frontier in various computer vision and remote sensing fields.

## 1.1 Goals and Future plan

The overarching goal of my proposal is to explore, investigate, and propose novel ways to improve HS anomaly detection and semantic segmentation for AiTR and efficiently quantify and measure their performance. The plan includes investigating ensemble anomaly detection and finding innovative ways to generalize the solution on various HS datasets. This plan also includes investigating task adaptation on an existing deep semantic segmentation model pre-trained on a large RGB dataset and spectral channel attention module to optimize it for HS inputs. It will also include unique ways to preprocess the HS images and measure their impact on detections. It will help with both remote sensing and aided target recognition. I will also plan to investigate a novel way to improve visualization for human understanding.

- **Process HS images** to improve machine detection accuracy
- **Identify anomalies** in HS images across various scenarios with greater accuracy to assist humans
- Scenes HS image **semantic segmentation**
- **Process images** to enhance human visibility

Achieving these goals presents multiple challenges. We have found that there is limited public HS data, and annotated data is even more scarce. To address this, we have comprehensively explored and combined unsupervised learning with supervised learning for anomaly detection. The HS image's numerous channels pose memory constraints on fitting deeper networks, leading us to employ no or limited deep-learning algorithms. Leveraging the HS imager's low spatial resolution at a high altitude and high spectral resolution, we have used spectral information and unmixing to improve accuracy. To overcome the biases of simple statistical solutions, We have developed an ensemble using meta-model to generalize them, thereby enhancing their accuracy. We have also devised a greedy solution to select base methods from various statistical HS-AD methods to use in an ensemble. As for future, our proposal

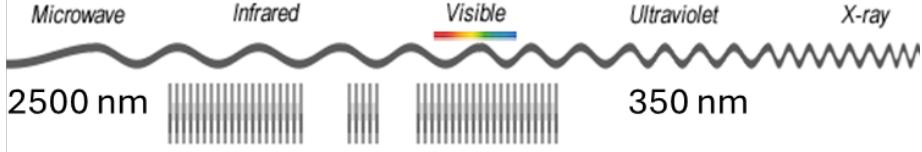


Figure 2: The Electromagnetic spectrum [6].

includes the application of domain adaptation to ensure the widespread application of our solutions. Furthermore, This plan also includes investigate task adaptation to optimize an existing deep semantic segmentation pre-trained on large RGB dataset for HS inputs. Finally, we will investigate the impact of image processing on detection algorithms and human understanding, potentially leading to significant improvements.

## 1.2 Background

In this section, we briefly introduce all topics related to this proposal. We discussed these topics in detail in the subsequent chapters based on their relevance.

### 1.2.1 Hyperspectral Imaging

Hyperspectral imagers capture data from contiguous electromagnetic bands from the electromagnetic spectrum, as shown in Fig. 2. An HS image has hundreds of channels and additional reflectance information per pixel [2]. Figure 3 shows an HS image stack that an HS imager can capture using the remote sensing (RS) technique. Figure 3 also shows various unique reflectance signatures of different materials, as they reflect light differently. Thus, it helps classify materials, differentiate fakes and decoys, and identify objects that do not belong (anomalies) in the captured scene. For example, targets that deviate spectrally from their surroundings in a scene.

HS imager captures hundreds of contiguous bands from the electromagnetic spectrum. However, weather can affect the quality of data collected. Time of day and cloud can impact the

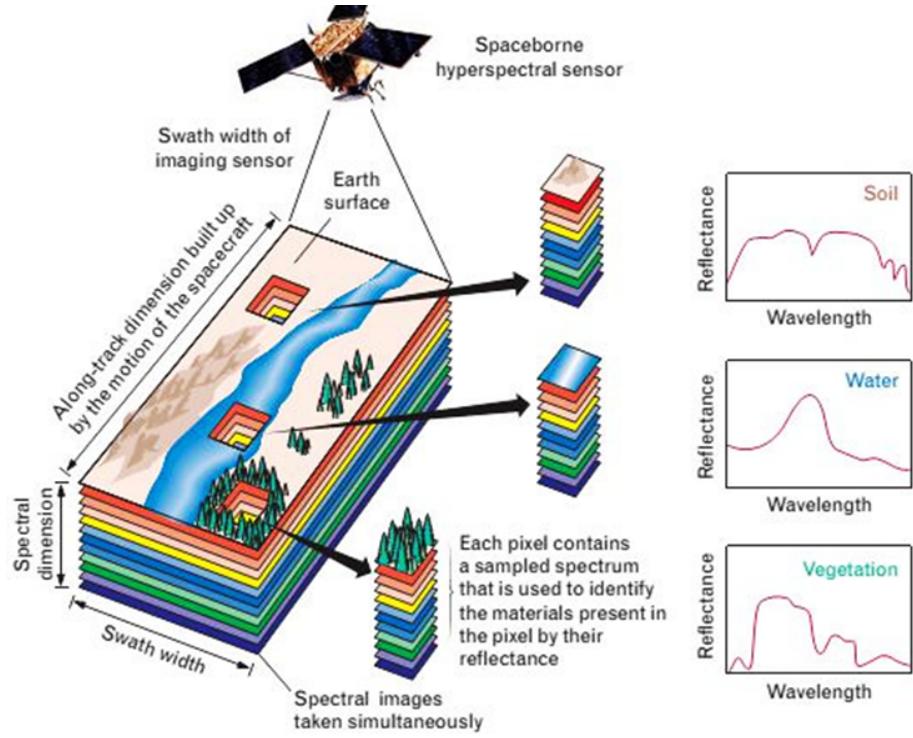


Figure 3: A generic scheme of HS imager mapping a scene identifying soil, vegetation and water through reflectance signature [4].

amount of sunlight available for reflectance, dust can create the scattering effect, and water vapor absorbs some electromagnetic bands completely and leaves only noise. Image enhancement techniques, such as noise reduction, brightness correction, and contrast enhancement, are crucial in curating better data. Thus, image preprocessing can improve the performance of standard detection methods. ML and deep learning (DL) models usually improve their performance with image augmentation using these techniques, too.

### 1.2.2 Aided Target Recognition (AiTR)

Aided Target Recognition (AiTR) refers to the use of advanced technologies and systems to assist in identifying and classifying objects or targets, typically in military, security, and surveillance contexts. AiTR systems enhance the ability to recognize and track targets by utilizing various tools, such as sensors, computer vision, ML algorithms, and automated

processing and analysis. These systems play a crucial role in reducing human error and workload, thereby enhancing detection accuracy and speeding up decision-making processes when identifying potential threats or objects is critical.

Figure 4 shows a military application where AiTR can help identify enemy vehicles using infrared sensors. The system may integrate infrared, radar, or optical sensor data from multiple sources (e.g., satellites, UAVs, or ground-based sensors) to provide more accurate identification in complex environments. Importantly, AiTR systems are designed to learn from previous encounters, continually improving their ability to recognize targets and instilling confidence in their adaptability.

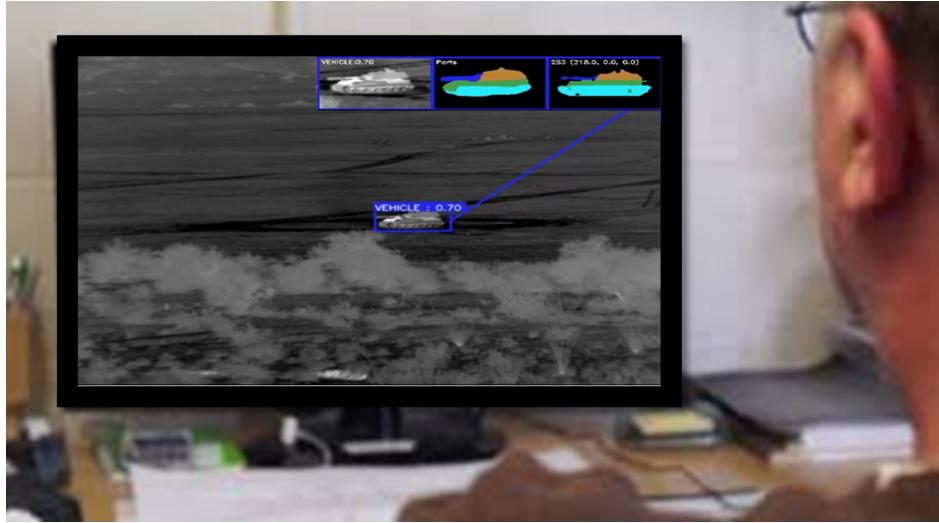


Figure 4: AiTR is helping identifying enemy vehicles in military applications.

AiTR enables operators to distinguish threats that humans cannot discern easily and rapidly in a highly cluttered environment where automated target recognition algorithms are unreliable [2]. AiTR needs a low False Alarm Rate (FAR) [7, 8]. A lower FAR value indicates fewer false targets identified as targets so as not to overwhelm the operator. AiTR requires a high Detection Rate [9] (i.e., recall) or a high ratio between the number of detected targets and the total number of targets. Although the detection and false alarm rates are inversely related, AiTR can prioritize a higher (better) detection rate, which may lead to a higher (worse) false alarm rate. Operators can take action to interrogate further and mitigate the threat.

In my exploration, I will use data from the HS imager and concentrate mainly on anomaly detection from the detection step and classification from the identification step. I will also investigate the impact of Preprocessing on machines and humans.

### 1.2.3 Potential Applications of AiTR

AiTR has the potential to significantly impact the agriculture, the environment, the military, and surveillance. In the agriculture, AiTR's potential in predicting plant health and detecting pests as shown in Fig. 5 is a reason for optimism. In the environment, AiTR's monitoring capabilities, including changes in land cover and pollution level detections, offer a reliable tool. In mineral exploration, AiTR's ability to recognize mineralization-related anomaly patterns and contour geological maps to detect target areas is a promising development. In the military, AiTR's role in camouflaged target detection, differentiating targets from decoys, is a significant advancement.

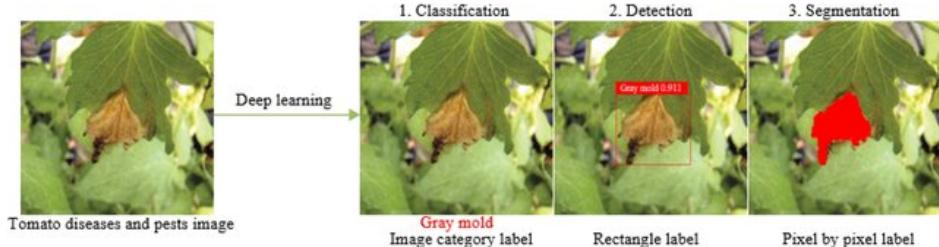


Figure 5: Example of plant diseases and pests detection problem [10].

### 1.2.4 Hyperspectral Unmixing

**Hyperspectral unmixing** identifies and quantifies the individual components within a mixed pixel in a HS image. Because HS data often has a lower spatial resolution, the pixels can represent a combination of multiple materials. Unmixing helps identify these materials and objects by analyzing their physical reflective properties. This process breaks down the measured pixel spectra into a group of constituent spectral signatures (endmembers) and their corresponding fractional contributions (abundances).

### *1.2.5 Hyperspectral Anomaly Detection (HS-AD)*

**Hyperspectral Anomaly Detection (HS-AD)** [11] aims to find targets in HS images that deviate spectrally from their surroundings in a scene. It helps characterize the captured scenes and separates them into anomaly and background classes. Supervised anomaly detection solves this problem using data labeled with ground truth, whereas unsupervised anomaly detection does not. Statistical methods have been used extensively for anomaly detection. They typically make assumptions about the background in an HS image to detect anomalies. Therefore, they will fail to capture the anomalies if the background does not satisfy their assumptions. However, ML recognition techniques can improve the detection of anomalies in a cluttered background. Thus, we propose a supervised stacking ensemble called GE-AD to combine the outputs from multiple statistical methods and ML algorithms (i.e., base methods to GE-AD), as well as a greedy search algorithm for selecting the base methods for the ensemble. However, GE-AD has difficulty handling HS datasets collected using diverse sensors or datasets collected using the same sensor but under different weather patterns, changed collection techniques, and other uncertainties. Both RS and AiTR can benefit from the improvement and generalization of HS anomaly detection algorithms.

### *1.2.6 Generalization of Detection*

Universality or generalization suggests that if a method achieves promising results on known datasets, it will perform similarly well for new or unseen data and provide consistent predictions. Usually, unsupervised HS-AD methods promise generalization as they should work well enough on unknown data. However, these methods typically do not generalize well when tested on other unfamiliar datasets. That's where Domain Generalization (DG) comes into play. DG is an ML technique that assumes no access to target domain data. It aims to learn a generalized model to differentiate between datasets from one or several training domains with different probability distributions. In the context of our research, DG will be used to align input features like the HS dataset and outputs of varying base methods. It weights them based

on their effectiveness for more than one dataset, thereby achieving good out-of-distribution generalization.

#### *1.2.7 Hyperspectral Image Semantic Segmentation*

**Image Semantic Segmentation** is a computer vision technique that classifies each pixel based on its semantic class. It precisely identifies and classifies every pixel within an image. It allows us to understand the detailed composition of a scene by distinguishing different objects at a granular level. **Hyperspectral Image Semantic Segmentation** utilizes spectral pixel information along with spatial data to solve this semantic segmentation problem. It is valuable for applications that require a granular level of image analysis, like medical imaging, military surveillance, and industrial inspection, where precise object localization and boundary delineation are crucial. Image semantic segmentation is a well-investigated problem, and various high-performing DL models designed and trained with RGB images like U-Net, FCN, and DeepLab are readily available. However, AiTR cannot use these models for scenes collected using an HS imager, as these HS images have hundreds of channels. Only updating the input layer of existing DL models to take in HS data can hamper their performance. In contrast, there is insufficient data to utilize these hundreds of channels from HS images to train a model from scratch and extract crucial spectral information. Thus, to address this challenge, the spectral feature should be combined with spatial features in the DL model and fine-tuned for better results.

### **1.3 Organization**

This dissertation proposal and the final dissertation will interpolate material from papers by the author [12, 13] and coauthored [14] with M Younis, A Robinson, L Wang, and C Preza. I incorporated some material from each of these papers into each Chapter.

The current proposal is organized into six major parts. The current introductory Chapter presents the research field of HS Aided Target Recognition (AiTR), existing problems, and

solutions. Chapter 2 discusses various HS and RGB image processing techniques and their impacts on detections and human understandability. It also explores a better way to select and apply these processing techniques. Chapter 3 looks into HS anomaly detection in remote sensing and the details of various detection methods. It goes over our hard voting ensemble HS-AD method [14], which uses predicted class labels for majority rule voting. It details our supervised stacking ensemble HS-AD method [12]. It also summarizes the greedy searching mechanism for base method selection for the ensemble. Chapter 4 outlines our ideas on generalizing HS-AD, which could broaden our HS-AD method's applications. It presents preliminary results, uses material from Reference [13], and describes our proposed solution to address the issue. Chapter 5 discusses our proposed hyperspectral image semantic segmentation method. Chapter 6 outlines the timelines, deliverables, and deadlines to extend and finalize this proposal into a complete dissertation document.

## **Chapter 2**

### **Hyperspectral Image Preprocessing**

This chapter examines various image processing techniques we apply initially to hyperspectral (HS) data before proceeding to anomaly detection and segmentation, which may improve data quality and increase training data for machine learning and deep learning models. These image-processing techniques enhance human visibility and understanding when applied to 1- or 3-channel digital images. These digital images can be 1-channel gray-scale and NIR images or 3-channel RGB and NIR-RG images.

#### **2.1 Previous Work**

Digital image processing, a subcategory of digital signal processing, enables a broad spectrum of algorithms to be applied to the digital input data. Its ability to resolve noise and distortion during processing is a significant feature. It is in considerable demand for its applications in diverse fields such as the agriculture, defense, environment, manufacturing, and medical imaging. The evolution of digital image processing is significantly influenced by the development of mathematics, particularly the creation and enhancement of discrete mathematics theory. In this section, we will dive deeper into a select few existing state-of-the-art image processing and enhancement solutions pertinent to this proposal. We will also review some examples followed by our proposed solution.

##### *2.1.1 Initial HS Image Preprocessing*

Radiometric calibration is the very first step in converting raw digital numbers (DNs) to radiance values. However, the sensor manufacturer supplies this solution to ensure the data represents physical measurements. We also apply geometric Correction (Orthorectification) to correct geometric distortions due to UAV movement (pitch, roll, yaw) and terrain [15].

Atmospheric Correction (AC, also referred to as Atmospheric Compensation,

Characterization, or ACA) Algorithms are used in the application of remotely sensed HS imaging data to correct for the effects of atmospheric scattering, propagation, and absorption on measurements acquired by air and space-borne systems. It is usually the first step after collecting new HS data, and it removes the atmosphere's impact on an image's reflectance values [16].

FLAASH (Fast Line of Sight Atmospheric Analysis of Spectral Hypercubes) [17] is one of the standard AC techniques. The goal of AC is to convert the radiance the sensor receives at the top of the atmosphere to the radiance of the Earth's surface. We apply FLAASH as the first step in our preprocessing pipeline. Wang et al. also found that effective atmospheric correction is essential for minimizing atmospheric interference in UAV-based hyperspectral imagery [18].

Analyzing the signal-to-noise ratio (SNR) for the HS images obtained from a new sensor is essential to ensure that the sensor maintains excellent overall sensitivity and spectral resolution, which supports the capture and preservation of critical information in the scene. Noise reduction techniques can remove unwanted noises identified during SNR analysis through smoothing, blurring, and filtering. Gaussian and median blur methods are some of the standard noise reduction techniques [19]. We selected the Gaussian filter over the median for better performance in our unmixing pipeline.

Deblurring restores a sharp image from a blurred one, which can sometimes reduce noise. A Wiener low-pass filter is one of the first deblurring techniques to reduce additive random noise in images [20]. K. Dabov et al. [21] proposed the BM3D denoising method using modified collaborative Wiener filtering. Y. Mäkinen et al. [22] introduced an improved version called BM4D with exact computation of the noise power spectrum and practical approximations of the spectrum for faster computation. We have tested BM4D; however, we have not seen any improvement in detection accuracy. We plan to test it on RGB images for better visibility.

### *2.1.2 Data Normalization for Machine Learning*

Aggarwal [23], and Zimek et al. [24] investigated outlier detection using ensemble methods. They found using multiple methods with wide ranges of output values challenging

since this often produces unbalanced ensemble results. Thus, we have investigated potential ways for normalizing anomaly detection scores. Normalization scales the feature values in a dataset to a desired similar range, often between 0 and 1, enhancing machine learning models' performance. It prevents features with larger magnitudes from dominating the learning process, allowing the model to learn more effectively from each feature. This leads to better model performance and faster convergence during training. Välikangas et al. [25] comprehensively evaluated various data normalization methods, which significantly helped our understanding and shortlisting choices. Zhao *et al.* [26] demonstrated the effectiveness of quantile normalization in their study. This method involves first ranking the gene of each sample by magnitude, calculating the average value for genes occupying the same rank, and then substituting the values of all genes occupying that particular rank with this average value. The next step is to reorder the genes of each sample in their original order [27]. It transforms the original data to remove any unwanted technical variation and preserve meaning [28]. It standardizes the distribution of each variable in a dataset, making their distribution identical across variables and allowing for easier comparison of patterns across different variables. However, this can obscure the original scale, variability, and magnitude, resulting in a loss of interpretability and original distribution characteristics.

### *2.1.3 Image Processing Algorithms for Visual Enhancement*

Histogram equalization is a contrast adjustment technique that increases the global contrast of an image with a narrow range of intensity values [29–31]. It evenly distributes the densely populated intensity values into the full range of intensities in the histogram. It allows areas with lower local contrast to achieve a higher overall contrast, enhancing the contrast of degraded images. Adaptive histogram equalization (AHE) differs from traditional histogram equalization by calculating multiple histograms for distinct sections of an image, which helps enhance local contrast and edge definitions [32]. However, AHE can overamplify noise in relatively homogeneous regions of an image. To address this, contrast-limited adaptive histogram equalization (CLAHE) is introduced [33]. Its purpose is to limit the amplification and reduce

noise, thereby improving AHE's performance. HSV CLAHE applies CLAHE to the brightness or value part of the HSV image (hue or tint, saturation or amount of gray, value or brightness), and LAB CLAHE applies CLAHE to the lightness part of the LAB image (lightness, green-magenta color axis, yellow-blue color axis). We have evaluated various types of histogram equalization and plan to continue using them for improved visibility.

#### *2.1.4 Performance Evaluation Metrics*

Image quality assessment (IQA) algorithms automatically assess the quality of images compared to perceived image quality by human observers [34, 35]. Comprehensive research efforts have evaluated the image quality and assessed the performance of image enhancement algorithms. Usually, the mean-square error (MSE) and the peak signal-to-noise ratio (PSNR) predict noise reduction performance. The structural similarity index measure (SSIM) assesses the improvement through enhancement. Blind IQA methods like natural image quality evaluator (NIQE) help when there are no reference images to compare with. Usually, it evaluates distortion and blur reduction algorithms.

## **2.2 Preliminary Work**

In Sections 2.2.1 and 2.2.2, we discuss the image processing methods we used before using the HS data for unmixing and anomaly detection. Section 2.2.3 discusses the techniques we used to improve human visibility and understanding.

#### *2.2.1 Preprocessing Methods to Improve Hyperspectral AD Algorithms for Arizona Dataset*

When dealing with HS images from a new sensor, it is crucial to analyze the SNR to ensure that the sensor has an excellent overall sensitivity and spectral resolution. This supports capturing and preserving the most important information in the scene and increases the efficacy of pre-processing techniques such as geo-rectification [15]. We have computed the SNR for all bands using the Avhyas plugin [36] in QGIS. Many HS images contain noisy bands, so further

analysis of the SNR is needed. Figure 6a displays the SNR for Image 5 from the Arizona dataset. Analysis of the figure reveals that the hypercube tends to be noisy for input wavelengths close to 400 nm and those higher than 860 nm. This suggested a procedure to remove those noisy wavelengths/bands, evaluate the anomaly detector on the new hypercube containing less noisy wavelengths, and create a new hypercube consisting of the bands with signal-to-noise ratios above our predefined threshold. The results of the noisy band removal are depicted in Figures 6a and 6b. Once all noisy bands are removed, we selected the most informative bands using the method proposed by Du and Yang [37] to reduce the hypercube's dimensionality and improve the processing times without degrading performance.

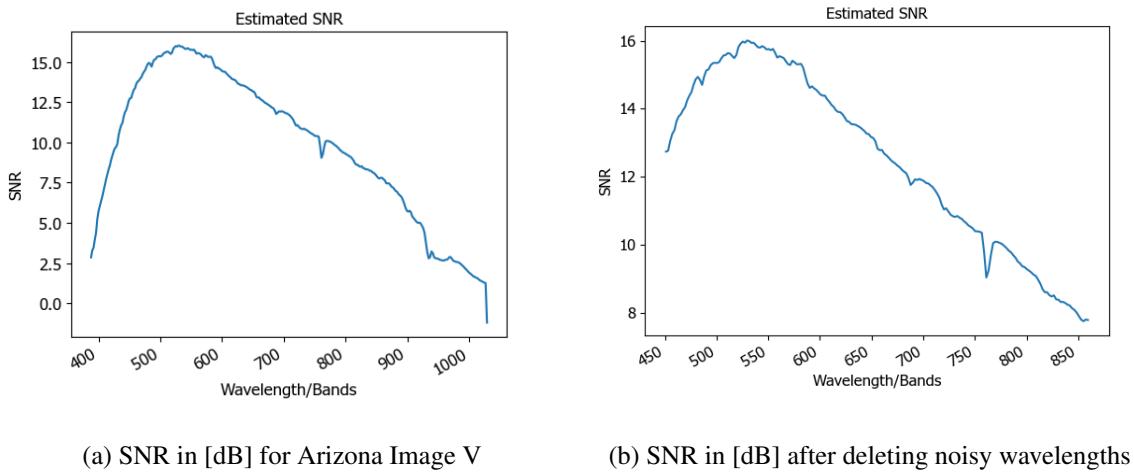


Figure 6: Positive impact of removing noisy bands from Image V in the Arizona dataset.

much better minimum hypercube SNR and does not result in any significant change to the input data characteristics. However, aggressively removing the noisy bands can reduce the spectral content and cause information loss, resulting in worse anomaly detection performance.

### 2.2.2 Normalization of Anomaly Detection Results

We have diverse output values representing anomalies from different AD algorithms.

Quantile normalization (quantile) can convert these results to a new distribution without losing

relative patterns (see Section 2.1.2). Figure 7 shows how quantile normalization transformed the irregular results from KIFD into a normal distribution.

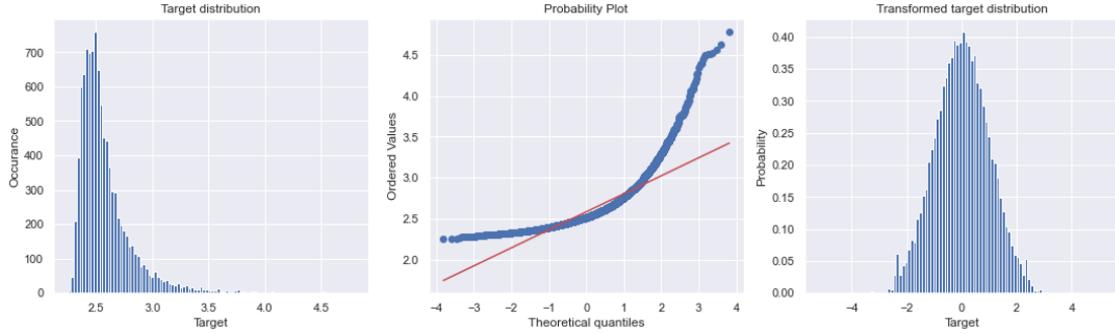


Figure 7: Irregular results from kernel isolation forest anomaly detection are transformed into a normal distribution. (a) Distribution of Original data. (b) Irregularities. (c) Transformed normal distribution.

When aggregating the output values, some meta methods, e.g. logistic regression, require normalization of the values, while others, e.g., Random Forest, inherently use information gain, which is not negatively impacted by scaling. Nevertheless, we applied normalization to ensure fair evaluation between various models. Figure. 8 shows a positive impact of the quantile normalization on the final results. We have also tested Z-score normalization, also known as standardization. It adjusts data values to have a mean of zero and a standard deviation of one. Figure 8 shows that quantile normalization boosts performance for tricky ABU-IV data.

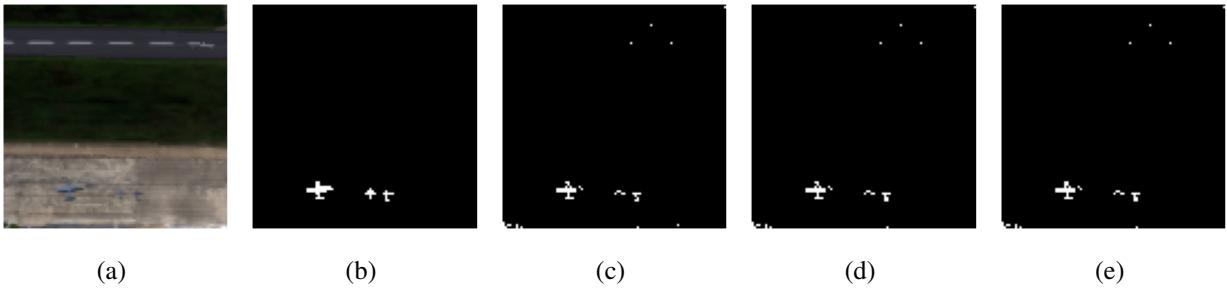


Figure 8: Normalization can give a boost in performance for tricky data like ABU-IV. (a) RGB, (b) ground truth, (c) GE-AD with no normalization ( $F_1 = 0.727$ ), (d) GE-AD with z score normalization ( $F_1 = 0.744$ ), (e) GE-AD with quantile normalization ( $F_1 = 0.757$ ).

### 2.2.3 *Image Enhancement Methods for Human Understandability*

Figure 9 illustrates the challenge of enhancing images to improve human comprehension. The middle image results from auto brightness adjustment, which weights image pixels toward lighter white pixel values and creates a brighter image. In contrast, the right image results from HSV CLAHE (Contrast Limited Adaptive Histogram Equalization), which redistributes intensity values to utilize the entire display range. This results in an overall bright image with more significant details. However, human eyes are naturally drawn to the enhanced brighter image (middle) [38] compared to the more detailed image (right), and the reflective tarp at the bottom is very pronounced in it (middle). There are few details in the original image, and the anomalies are not distinctly visible; still, some may find it more appropriate for anomaly detection. Conversely, increased information (right) can hinder a person's ability to identify anomalies. Thus, the effectiveness of enhancing images to help humans detect anomalies is very subjective. In this scenario, producing a helpful image for stakeholders is more important than developing a new enhancement algorithm.

We will not develop a new algorithm as various standard image enhancement methods are readily available. Instead, we will devise an efficient way to select image enhancement methods for our images, as most methods focus on achieving a high quantitative performance score rather than producing high-quality pictures that can help humans detect anomalies.

## 2.3 **Further Research Plan**

### 2.3.1 *Process HS images to Improve Machine Detection Accuracy*

Previously, we have removed noisy bands below a fixed threshold (5 SNR) or below a certain percentage (97%) [12, 39]. Figure 10 shows various bands with noise. We may recover sensor noise before  $0.40\mu m$  and after  $1.00\mu m$ , weak atmospheric water absorptions at  $0.60$  and  $0.66\mu m$ , and slightly stronger atmospheric water absorptions at  $0.73$ ,  $0.82$ , and  $0.91\mu m$ . However, the strong atmospheric water absorptions at  $0.94$  and  $1.14\mu m$  are not possible to

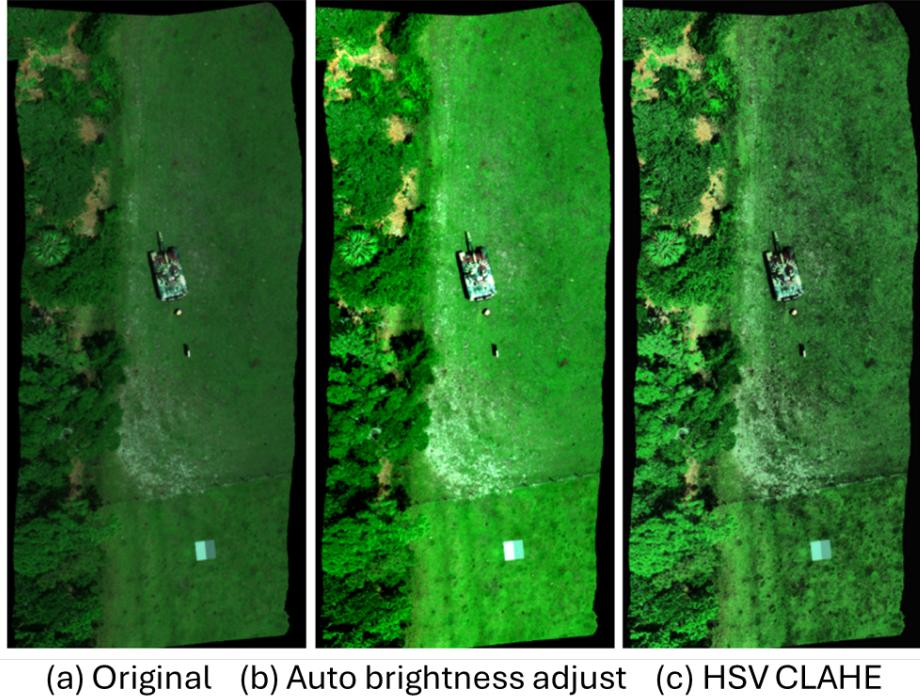


Figure 9: Example of digital image processing. From left to right, it shows the original image, the image applied auto brightness adjust (SSIM 0.685), and the image processed using HSV CLAHE (SSIM 0.677).

recover.

Aggressively dropping noisy channels may cause data loss and reduce anomaly detection performance. I will investigate the impact of removing and recovering some of these bands. Data collection and stitching may introduce artifacts and noises. My objective is to improve the performance of detection methods by preprocessing. I am committed to finding a systematic way to select preprocessing methods. A hidden potential may be untapped due to the manual section of preprocessing methods. We aims to optimize input data for better outcomes in each scenario.

### 2.3.2 *Image Enhancement Methods for Human Understandability*

This part of my research will investigate improving the processing of digital images, with the ultimate goal of enhancing human visibility and helping them detect targets. We will collect various image processing algorithms, create a new dataset, and propose a new algorithm selector

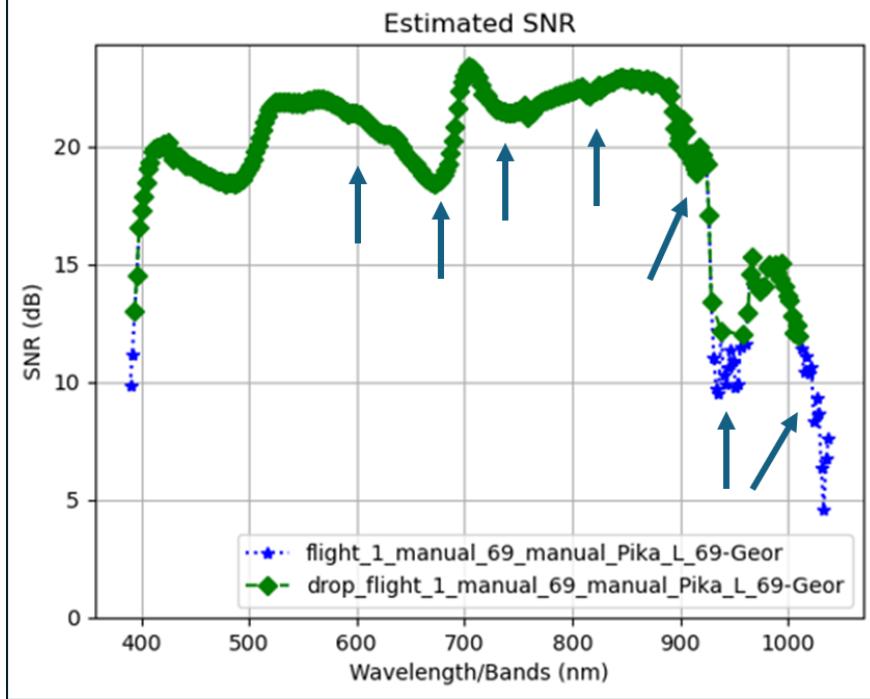


Figure 10: Visual comparison showing quantitative improvement of removing noisy bands from Florida Image 1. The green line shows SNR in [dB] after deleting noisy bands, whereas the blue line shows the dropped channels.

system, all to make a tangible difference in image processing.

At the heart of this research is the understanding that user feedback is invaluable to creating an efficient aided target recognition system. We can simultaneously apply various image processing and enhancement techniques to the same image. To ensure that our model is user-centric, we will pursue a classification model [40] as shown in Fig. 11. This classifier predicts mutually exclusive class labels (i.e., image enhancement methods) for a scene in a sequence based on latent space features created from a scene. The ground truth of the model will come from users who will vote on which image is better than the original image. This user feedback is crucial as the model learns to mimic high-quality pictures of what human eyes like for identifying targets rather than focusing solely on achieving a high quantitative performance score.

We will create a new GUI application (or Web App) for user feedback. This application

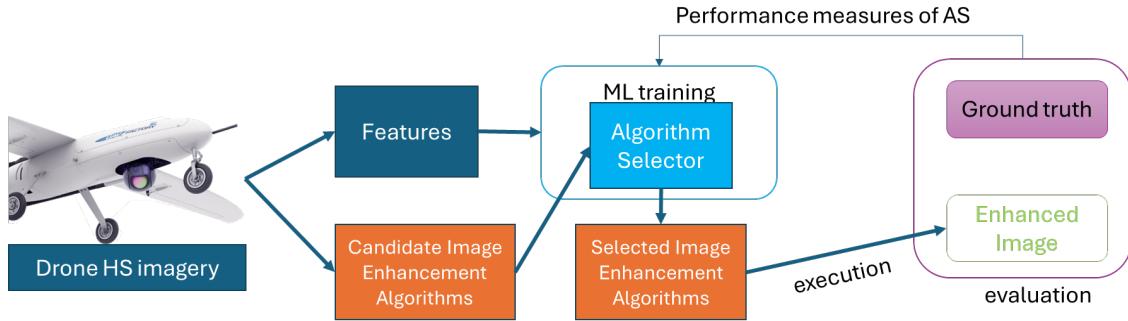


Figure 11: Image Enhancement Method Selection using ML Model

will randomly use any processing or combination of methods to develop enhanced digital images. The crucial element is that the human will select between the original and enhanced images. We plan to showcase one process and image combination to twenty users and identify if one algorithm receives ten votes. If two or more image processing algorithms receive ten votes, they will compete with each other. The application will present enhanced digital images from the two selected processing methods, continuing until it finds an algorithm combination with the highest votes for an image. We can create a new dataset based on the highest vote, consisting of data tuples (original image, algorithm combination).

The challenge we face is significant as we propose a novel data approach. Traditionally, the image enhancement dataset is a tuple of the original and enhanced images. However, our data tuple combines the original image and algorithm. To fully explore this innovative approach, we require sufficient manpower and support. Securing these resources is crucial to the success of my proposal. I will deliver this solution if we can produce this dataset.

## Chapter 3

### Hyperspectral Anomaly detection

This chapter will discuss hyperspectral anomaly detection (HS-AD), which identifies anomalies in HS images. Anomalies are pixels that are spectrally different from their neighboring pixels. HS-AD helps characterize the captured scenes and separates them into anomaly and background classes. Unsupervised AD methods do not require any annotated data. However, their performance varies in different scenarios due to design assumptions.

#### 3.1 Previous Work

##### 3.1.1 Individual HS-AD Algorithms

Table 1 shows various types of individual HS-AD Algorithms. The Reed–Xiaoli (RX) Anomaly Detector algorithm characterizes the HS image’s background using the hypercube’s mean and covariance [41, 42]. RX calculates the Mahalanobis distance between the background and pixel under test. The pixel will be declared an anomaly if the distance exceeds a predefined threshold [43]. The RX algorithm is currently the performance comparison benchmark for most anomaly detection algorithms. It has gained popularity by enabling the detection of anomalies without the need for scarcely available labeled data. However, the RX algorithm does not consider the influence of anomaly targets on the background computation. The Kernel-RX Algorithm (KRX) [44] considers HS image data complex and non-linear. It uses the Gaussian Radial Basis Function (RBF) kernel trick to compute the global normal statistics, project them into a linear model in a higher-dimensional feature space, and apply the RX algorithm in that higher-dimensional feature space. Thus, KRX produces better results than the RX algorithm for a complex scene. One of the main limitations of KRX is that it is computationally intensive. If the dimensions of the HS image are large, then KRX would also require a large amount of memory.

The Gaussian Mixture RX Anomaly Detector (GMRX) [45] fits the Gaussian Mixture Model (GMM), assuming the entire image is the background, and assigns pixels to the highest

Table 1: Examples from various types of HS-AD algorithms.

Type of HS-AD	Examples
Statistics based	RX, MD-RX, WIN-RX
Subspace based	LSUNRSORAD, CSD, SSRX
Cluster-based	CBAD, FCBAD
Spatial-spectral based	AED
ML-based	SVM, iForest
Kernel-based	KRX, GM-RX, KIFD

posterior probability mixture component. Then, it applies the RX algorithm to those pixels.

However, GMRX fails if it assumes an incorrect number of components.

The Complementary Subspace Detector (CSD) [46] considers HS image data to be linear, assumes that the background and target are complementary subspaces of principal component analysis (PCA), and applies the RX algorithm in those subspaces. PCA is a linear dimensionality reduction technique that uses the data's singular value decomposition (SVD) to project it to a lower dimensional space.

Cluster-Based Anomaly Detection (CBAD) [47] considers each pixel to be strictly a member of one cluster in k-means clustering. It computes the pixel's Mahalanobis distance to the component mean. Fuzzy Cluster-Based Anomaly Detection (FCBAD) [48] is a novel extension of the CBAD and GMRX algorithms. It assumes that each pixel can have several possible fuzzy logic membership functions in fuzzy c-means clustering and computes each pixel's Mahalanobis distance to the component mean. The weakness of any clustering-based outlier detection is that the efficacy largely depends on the clustering method used, and these methods are hard to optimize for outlier detection. The model is not more flexible as it needs to be adjusted according to the distribution characteristics of the different datasets. It fails if it assumes an incorrect number of components. Finally, the application of clustering techniques to large datasets is usually expensive.

HS images provide spatial and spectral data about objects of interest. Spatial information

can improve the reliability and stability of anomaly detection if the resolution is high enough. For example, HS-AD with Attribute and Edge-Preserving Filters (AED) [49] consists of two stages. This algorithm assumes that anomalies tend to be smaller and possess unique reflectance signatures and that the pixels belonging to the same class would have a high correlation in the spatial domain. First, it utilizes principle component analysis to reduce the data's dimensionality. Then, a predefined attribute filter is used along with boolean-based fusion to create the initial detection map. The detection map is then refined using edge-preserving filters to reduce the false positive outliers. AED uses a predefined filter, which may not be suitable for various datasets. Moreover, AED assumes that the anomalies must be small objects, making it unsuitable for detecting the large anomalies in a scene.

Du et al. [50] proposed Local Summation Anomaly Detection (LSAD), which utilizes the second-order Mahalanobis distance and a window filter to create the local summation. LSAD represents the correlation between backgrounds via a correlation matrix. The matrix inverse is used during mathematical computation, making the process computationally intensive. Thus, Tan et al. [51] proposed Local Summation Unsupervised Nearest Regularized Subspace with an Outlier Removal Anomaly Detector (LSUNRSORAD) to improve the process by utilizing a linear combination using multiplication and addition. The general idea of the LSUNRSORAD algorithm is the ability to represent background pixels by its neighbors. In addition, LSUNRSORAD uses two thresholds in the outlier removal process to improve the accuracy; one is called the minimum threshold, and the other is called the maximum threshold. If the value of a pixel is higher than the maximum threshold or lower than the minimum threshold, it would be considered an anomaly. The main disadvantage of this algorithm is that it assumes the background obeys a Gaussian distribution. The probability of degraded performance is increased in scenarios where this assumption does not hold.

The isolation forest (iForest) [52] contrsucts multiple binary trees by random sub-sampling to isolate anomalies. The iForest assumes that in an HS image, a background pixel is hard to isolate from the neighboring pixels whereas anomalous pixel separation should

be relatively straightforward. As a result, in the constructed trees, regular pixels reside in a deeper part of the trees, while anomalous pixels reside in the most shallow parts of the trees. The HS-AD with Kernel iForest (KIFD) [53] algorithm assumes that anomalies are easily distinguishable and do not occur frequently. First, it maps the data into the kernel space in order to reduce scene complexity. It then applies the PCA to choose most important features. Next, it uses the iForest algorithm constructs the trees, which outputs an initial detection map that is further improved using locally constructed iForest over sliding windows. The depth of a pixel in the tree determines if the pixel is regular or anomalous.

### 3.1.2 Ensemble HS-AD Algorithms

Classic HS anomaly detection algorithms usually make assumptions, including background distributions and the frequency of anomalies. However, real-life scenarios are much more complicated than this, which is why many anomaly detectors fail to identify anomalies. Various techniques are proposed to solve these anomaly detection methods' instability problems, and ensemble anomaly detection methods are one of them. This section reviews several ensemble learning algorithms related to our work and various approaches to automating the selection and aggregation of the base models for ensemble learning. In Table 2, we summarize all these methods' similarities and dissimilarities.

The detectors derived from the Collaboration Representation Method (CRD) are classic HS-AD methods. CRD represents each test query as a linear combination of all the training samples from all the classes for image classification [60]. It has a high computational cost due to the use of a dual sliding window to explain a complex scene through a linear combination. Lu et al. [54] proposed an Ensemble and Random CRD (**ERCRD**) method that utilizes random sub-sampling on homogeneous CRD (RCRD) base models to reduce the computational complexity of each RCRD method. It then uses summation to aggregate the detection results from the RCRD methods. Although ERCRD reduces the computational complexity of each RCRD method, the ensemble process still brings its own computational burden without the

Table 2: Feature comparison of ensemble learning algorithms.

Literature	Base Models	Type of Base Models	Base Model Optimization	Base Model Selection	Aggregation
ERCRD [54]	homogeneous	unsupervised	random sampling	N/A	bagging, summation
UE-kPCA [55]	homogeneous	unsupervised	random sampling, gradient descent optimization	N/A	bagging, average
ERRX MFs [56]	heterogeneous, mixed with passthrough	unsupervised	random sampling	not mentioned	bagging, average
SED [57]	only one method or heterogeneous, mixed with passthrough	unsupervised	not mentioned	based on ensemble performance	stacking, unsupervised meta method
Fatemifar et al. [58]	heterogeneous	unsupervised	not mentioned	genetic algorithm	stacking, unsupervised meta method
Nalepa et al. [59]	heterogeneous	supervised	not optimized	not investigated	stacking, supervised meta-learner
HUE-AD [14]	heterogeneous	unsupervised	not optimized	based on individual performance	bagging, voting
GE-AD	heterogeneous	unsupervised	not optimized	greedy search	stacking, supervised meta-learner

assurances of complex detection.

Compared to principal component analysis (PCA), kernel PCA (kPCA) can carry out non-linear projections for more complex data. Merrill et al. [55] proposed Unsupervised Ensemble kPCA (**UE-kPCA**) by suggesting an anomaly scoring function using reconstruction error in kPCA feature space. To reduce computational costs, the authors utilized sub-sampling and an ensemble of homogeneous kPCA base models to refine the final anomalies through averaging. Furthermore, they introduced a novel loss function and automated the kernel parameter selection using batch gradient descent. Despite these advancements, UE-kPCA remains a complex method that requires finding parameters for multiple kPCAs, which can be computationally expensive.

Yang et al. [56] proposed an Ensemble and Random RX with Multiple Features (**ERRX MFs**) anomaly detector. Three features, Gabor, Extended Morphological Profile (EMP), and Extended Multiattribute Profile (EMAP), were computed using heterogeneous base models mixed with the original HS image (passthrough). They utilized random sub-sampling to compute multiple results for each feature using the RX algorithm. The results for each feature are fused using a voting average ensemble. Then, the four ensembles (for the three features and the original HS image) produce the final results. This approach avoids the problem of mixing diverse features in an ensemble. However, overrepresented features or noise can dictate the anomaly detection results without standardization.

Wang et al. [57] proposed a subfeature ensemble called **SED**. They removed the noisy bands, normalized the remaining ones, randomly sub-sampled several subfeature sets, and used six different base models in the ensemble. Unlike the ERRX MFs algorithm proposed by Yang et al. [56], SED evaluates the six base models to choose the best-performing methods and uses them to obtain an enhanced feature set. This enhanced feature set is combined with the original HS image (passthrough) to become the input for the PTA algorithm (meta-model). However, some methods used the same distribution model of the background. Due to background interference, feature redundancy, and noise, the feature enhancement process for these processes

may become irrelevant to the final anomaly map. Additionally, SED uses a feature ensemble method rather than an ensemble of different algorithms, as the first-level algorithms aim to obtain the enhanced feature map.

Fatemifar et al. [58] proposed a stacking ensemble for face spoofing anomaly detection that consisted of 63 base classifiers and a Gaussian Mixture Model (GMM) as the meta-classifier for the second stage. The proposed stacking ensemble utilized all 63 classifiers by assigning weights ranging from  $-8$  to  $8$ . While the application for this ensemble is different, it is worth noting that assigning values for each of the 63 weights, even those not performing well, might not result in the most optimized algorithm, especially for the larger dimensional images.

Nalepa et al. [59] proposed a deep learning ensemble for data classification and unmixing that utilized convolutional base models and model augmentation to create new modified models by adding Gaussian noise to the weights of the base models and a second level fuser. The fuser could be as simple as an averaging process, a majority vote, or a supervised second-level fuser. While the proposed method performs well for the tasks of classification and unmixing, it is unclear how effective adding Gaussian noise to the base model weights is when used for detecting anomalous objects.

Ensemble methods yield superior outcomes as they combine the predictions of multiple methods [61]. Aggarwal et al. [62] delved into the effects of using homogeneous weak outlier detectors in the bagging ensemble method in their research. On the other hand, we found that combining heterogeneous weak detectors in an ensemble creates a better understanding of the background and provides better prediction results. Thus, we explored this direction using unsupervised heterogeneous weak detectors in our previous work [14]. However, we found that the process of selecting these algorithms from many options and assigning them weights is not systematic or efficient. In this study, we propose the Greedy Ensemble Anomaly Detection (GE-AD) method with greedy search as the solution to the above problem.

### 3.1.3 Hyperspectral Unmixing-Based Voting Ensemble Anomaly Detector (HUE-AD)

To tackle the shortcomings of the statistical HS-AD methods, we utilized them in an ensemble. HUE-AD [14] is an equal-weight voting method that combines four detectors (Abundance, AED, KIFD, and LSUNRSORAD) to identify anomalies. These methods are selected based on our domain knowledge and promising precision scores. We used the unmixing method N-FINDR to compute the abundance.

**Hyperspectral unmixing** identifies and quantifies the individual components within a mixed pixel in a HS image. Due to lower spatial resolution, pixels in HS data may be a combination of multiple materials, which unmixing can help identify those materials and objects using physical reflective property [63]. It breaks down those measured pixel spectra into a group of constituent spectral signatures (endmembers), along with their corresponding fractional contribution (abundances) [64].

The first step in the spectral unmixing process is estimating the number of unique spectra or endmembers. This can be accomplished by an algorithm such as the noise-whitened Harsanyi Farrand Chang method [65]. Once the number of endmembers in the HS image is known, the N-FINDR algorithm can estimate those individual spectra [66]. Figure 12 displays the complete method to get target abundance using target endmember [14].

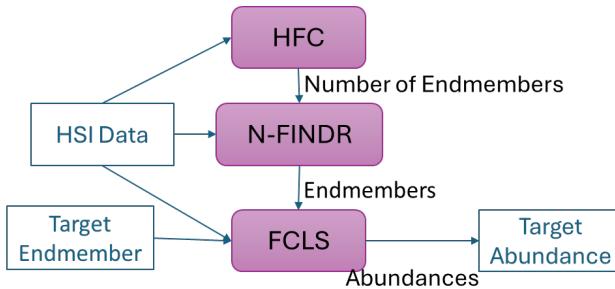


Figure 12: The complete method to get target abundance using target endmember [14]. A target endmember is a specific reference spectrum extracted from a reference HS image. The process of extracting target abundance is similar to object detection.

Once the endmembers have been estimated in an HS image, then the Fully constrained least-squares (FCLS) [63] method could be applied to estimate the abundance map of the unique

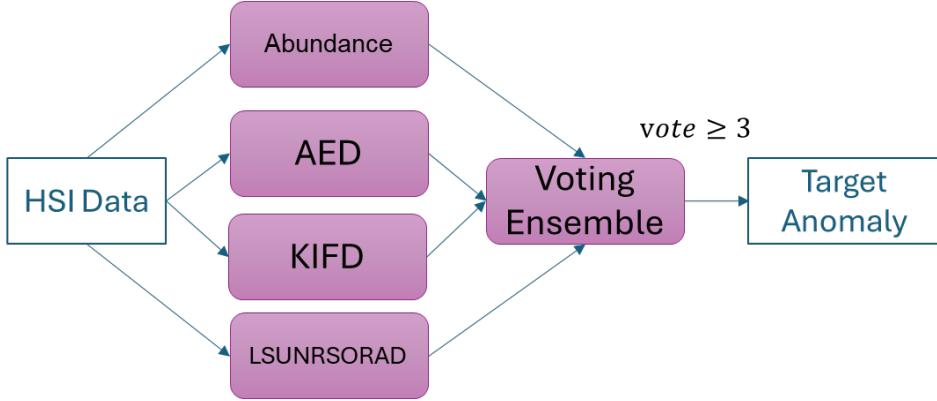


Figure 13: Design for Hyperspectral Unmixing-Based Voting Ensemble Anomaly Detector (HUE-AD).

spectra in the HS image. This method unmixes an HS image into the fractional abundance of each material in each pixel. The output abundance map has the same size as the input HS image and has the same number of abundance maps as the number of estimated end-members in the HS image. Further analysis of the HS image can be done using the spectral similarity information divergence algorithm (SID), which can measure the spectral similarity between pixels and specific reference spectra [67]. The particular reference spectra could be a part of the same image or be extracted from another reference HS image. In this case, the process is similar to object detection as shown in Fig. 12. FCLS assigns a probability distribution to each class as an abundance map. The class with the highest probability is normalized to 1 and classified as that pure material spectrum.

HUE-AD only takes in the binary vote (a vote for a pixel means the pixel is a detected anomaly). We created binarized results by thresholding the four base models' results. Based on the three-sigma rule of thumb, a 99.7th percentile threshold can be used to identify the anomalies in normally distributed data. However, Chebyshev's inequality [68] suggests that at least 88.8% of cases should fall within properly calculated three-sigma intervals, even for variables that are not normally distributed. We conducted tests to improve the F1-macro score and determined that a 97th- percentile threshold would be more appropriate for our purpose.

During our evaluation, we found that systematically assigning weights to inputs improved

the accuracy compared to the traditional equal-weight HUE-AD algorithm. We also found that the threshold selection can be a significant source of information loss, resulting in complete system failure in extreme cases. Thus, we developed our model stacking solution called GE-AD to find those weights using a supervised meta-learner and a normalization method to use the complete information generated by the base AD algorithms without introducing bias.

### 3.2 Greedy Ensemble Anomaly Detection (GE-AD)

Ensemble learning is a machine learning technique that combines multiple individual **base models** or weak learners to create a stronger, more accurate model with better predictions. The individual base models in an ensemble can be different algorithms (heterogeneous) or the same algorithm (homogeneous), and they can be unsupervised or supervised. Note that, even if the base models are the same algorithm, they may be trained on different subsets (random sampling) of the training data. Furthermore, selecting the best parameter values using various approaches, such as cross-validation, i.e., *hyperparameter tuning*, can optimize the individual base model's performance. The three main classes of ensemble learning methods are **bagging**, **stacking**, and **boosting**. Stacking uses another ML model to aggregate the base models' predictions, whereas bagging uses voting, averaging, or summation, and boosting uses a sequence of models to correct the predictions of prior models.

Figure 14 shows the flowchart of a stacked model [61], which uses output predictions from multiple base models trained on the same dataset to train another ML model (**meta-model**) to make the final predictions [69]. Please note that the meta-model will assign different weights to the base models to maximize the performance of the proposed ensemble method.

We previously proposed HUE-AD [14], an equal-weight voting ensemble of four heterogeneous detectors. Our empirical analysis showed that a weighted voting approach outperforms HUE-AD. However, determining the voting system's optimal weights requires a training dataset. Drawing inspiration from Feature Weighted Linear Stacking (FWLS) [70, 71], we developed a two-stage weighted stacking ensemble with unsupervised weak HS-AD methods

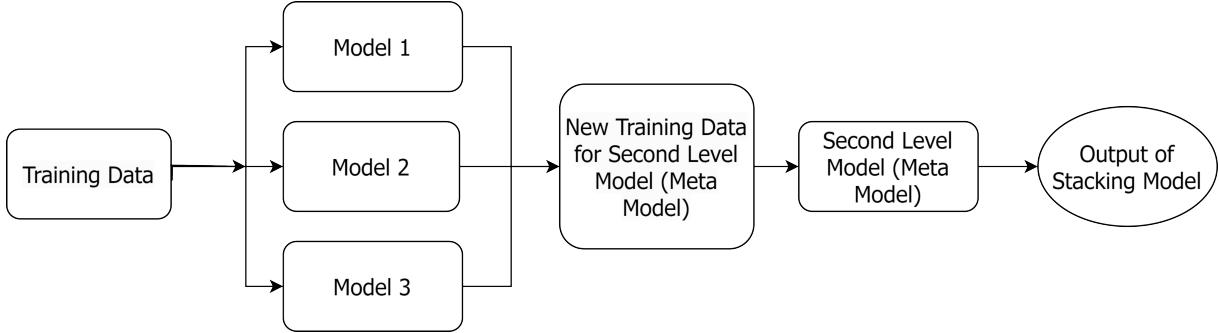


Figure 14: The architecture of a stacked model, where the new training data consist of the outputs of the first-level models.

in the first stage and a supervised machine learning (ML) method in the second stage to compute the weights. For the first stage, as there are limited annotated HS image datasets, a supervised ML model may suffer from under-fitting, so we opted to use unsupervised HS-AD methods. For the second stage, as learning the appropriate weights to assign to multiple inputs requires much less training data than learning features from images, we can utilize a supervised ML method. Initially, we explored Logistic Regression [72] and decision trees [73] to determine the weights in the second stage. Ultimately, we opted for random forest [74] due to its promising results. A random forest is a meta-estimator that fits several decision tree classifiers on various sub-samples of the dataset. These sub-samples are generated by randomly drawing from the dataset with replacement. It uses averaging over the predictions from the decision trees to improve the predictive accuracy and control over-fitting.

Additionally, our empirical analysis showed that the combination of base models that yields the best outcomes varies based on the given scenario. As a result, we put our effort into identifying a more systematic and efficient method for base model selection in a weighted voting ensemble. Figure 15 shows the overall flowchart of the GE-AD algorithm, and our proposed greedy search is depicted in Figure 16.

The GE-AD algorithm [12] uses consistent stratified 2-fold cross-validation.  $K$ -fold cross-validation is a widely used technique to assess the accuracy of a model [75–77]. It splits the

training dataset into  $K$  subsets or folds. The model is trained using  $K - 1$  of the folds as training data, and the resulting model is validated on the remaining part of the data. The model is trained and evaluated  $K$  times, with each fold serving as the validation set in turn and being used for validation exactly once. The performance metrics obtained from each fold are then averaged to estimate the model's generalization performance. In our approach, 2-fold cross-validation divides the pixels in our data into 50–50 split training and testing datasets and evaluates the performance of our GE-AD algorithm twice. Our evaluation found that getting stuck in a local maxima can hinder searching for the best methods. We solved this issue by running the search algorithm five times. We included this  $5 \times 2$ -fold cross validation as a part of the GE-AD algorithm to ensure that we find the best combinations. Out of these ten sets of results, GE-AD identifies which combinations survived the most. If there is a tie, it uses first the average test score and then the average training score to break the tie. For our ensemble fusion, we have considered one HS unmixing algorithm, FCLS, to generate an abundance map and nine AD algorithms, including AED, CSD, FCBAD, GMRX, KIFD, KRX, LSUNRSORAD, Median AD, and RX.

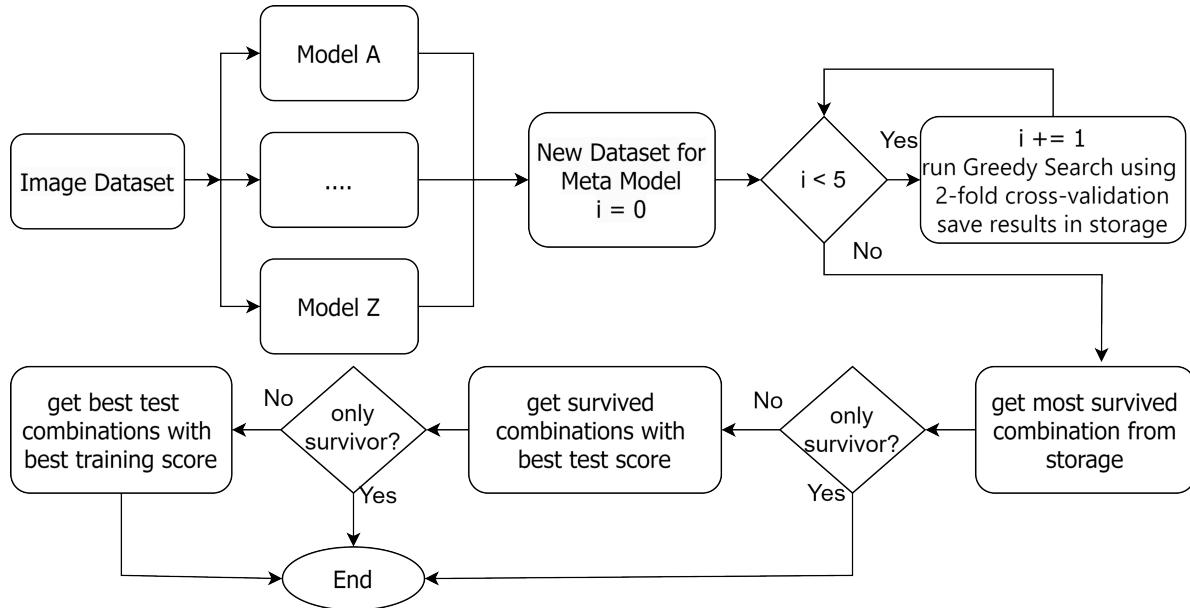


Figure 15: Overall flowchart of GE-AD to find the best AD methods for ensemble fusion.

In Figure 16, the `k_best` variable defines the highest number of base models to be used in combination. We varied the `k_best` variable to determine how many AD methods to

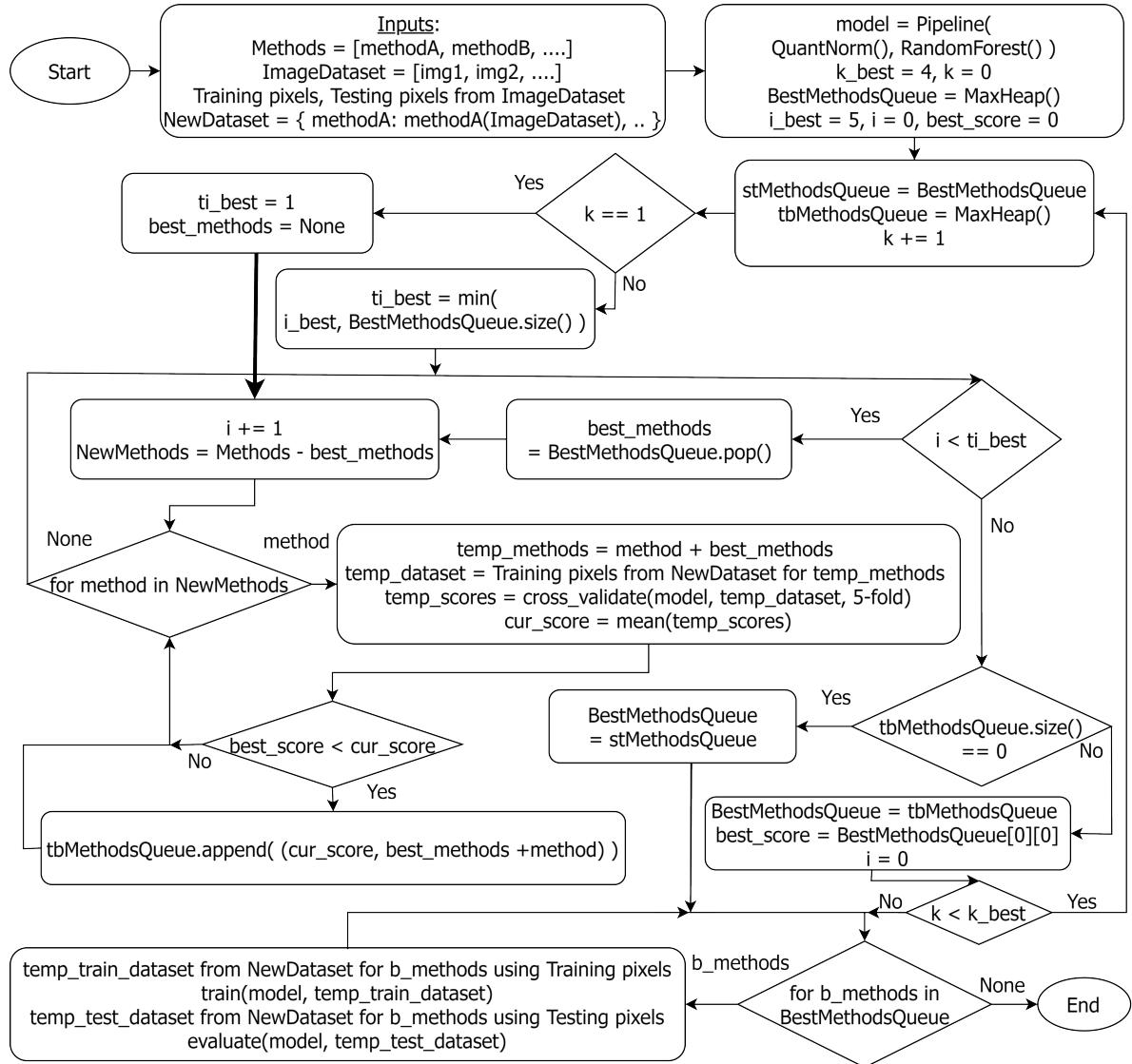


Figure 16: Flowchart of the greedy search to find the best AD methods for ensemble fusion.

include in the ensemble. We found that using four AD methods provides the best results within the constraints of computation complexity. Thus, we set four as the value for `k_best`. We first applied the AD algorithms to the image data to get each method's raw anomaly detection output. Then, we used these detection outputs as input features to train the ML model pipeline, which contains quantile normalization and random forest. The greedy search uses 5-fold cross-validation to compute the average F1-macro score of the trained ML model which fuses the output from the selected methods and stores the scores in a Priority Queue (Max Heap). Initially, the top `i_best` AD methods are selected individually from the priority queue. We varied the `i_best` variable to determine how many combinations to pick up from the priority queue. We found that using the top five unique performers is enough to provide better results. Thus, we set `i_best` to be five. Then, in each subsequent round, one more method is added to the set found in the previous iteration. The methods (or combinations of them) that perform better than the best score of the last round are stored again in the queue, and iteration continues. As the number of methods reaches `k_best` or no better result is found, the greedy algorithm stops its search. We then train the ML pipeline on the training dataset and evaluate the performance using the testing dataset (both datasets are created from the selected methods' outputs). As GE-AD model trained on one dataset may differ from that trained on other datasets, we may identify each of these distinct GE-AD models by a different name.

Now, we analyze the time complexity of the greedy search algorithm. The greedy search starts with finding the performance of the entire set of  $n$  AD methods. The priority queue helps greedily choose only the best-performing combinations. Because ML model training time is orders of magnitude higher than the priority queue's enqueueing and dequeuing time, we do not consider the latter in the complexity analysis.

Thus, the runtime of the first round is  $\mathcal{O}(n)$  as  $n$  AD methods are evaluated. In the second round,  $i_{best}$  best-performing methods were dequeued, each combined with one of the remaining  $(n - 1)$  methods for training and evaluation.

Thus, the runtime of the second round is  $\mathcal{O}(i_{best} \times (n - 1))$  for model training computation

as up to  $i_{best} \times (n - 1)$  AD methods are evaluated and enqueued into the priority queue. Then, in the third round, the  $i_{best}$  best-performing combinations of two methods are dequeued, each combined with one of the remaining  $(n - 2)$  methods for training and evaluation. A table with all combinations is kept for time optimization and to avoid recomputation. Eventually,  $k_{best}$  number of methods are chosen. Then,  $n + i_{best} \times (n - 1) + \dots + i_{best} \times (n - k_{best} + 1)$  training and evaluations are performed resulting in time complexity of

$$\mathcal{O}(n + i_{best} \times (n - 1) + i_{best} \times (n - 2) + \dots + i_{best} \times (n - k_{best} + 1))$$

or,  $\mathcal{O}(n)$ , when  $i_{best}$  and  $k_{best}$  are small constants.

On the other hand, a brute-force search trains and evaluates  $C(n, 1) + C(n, 2) + \dots + C(n, k_{best})$  combinations, with a time complexity of

$$\mathcal{O}(C(n, 1) + C(n, 2) + \dots + C(n, k_{best}))$$

or,  $\mathcal{O}(C(n, k_{best})) = \mathcal{O}\left(\frac{n!}{(n - k_{best})!k_{best}!}\right) \sim \mathcal{O}(n^{k_{best}})$ , when  $k_{best}$  is a small constant.

### 3.3 Evaluation

#### 3.3.1 Datasets

We assessed the AD algorithms' effectiveness using the scenes from the Airport–Beach–Urban (ABU) dataset [49, 78], the San Diego dataset [79, 80], the Salinas dataset [81], the Hydice Urban dataset [82], and the Arizona dataset [15].

We used four airport scenes from the ABU dataset [49, 78] with dimensions of  $100 \times 100$  pixels, manually extracted from larger images obtained from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) website [83]. The AVIRIS is a unique optical sensor that delivers spectral radiance in 224 contiguous spectral channels (bands) with wavelengths from 370 to 2510 nm. The authors removed the noisy bands in the original images using a noise level

estimation method [84]. The four images have different spatial resolutions due to variations in flight heights. ABU-I, II, and III have a spatial resolution of 7.1 m per pixel. In comparison, ABU-IV has a spatial resolution of 3.4 m per pixel. The main background includes roofs, shadows, and aircraft parking aprons; airplanes in the image are considered anomaly targets.

The San Diego airport dataset is also a subset of the AVIRIS dataset. It was preprocessed in a previous study [79, 80] as follows. The low-SNR (signal-to-noise ratio) and water vapor absorption bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224) were eliminated, and the remaining 189 bands were used. The dataset has two images with a size of  $100 \times 100$  pixels and a spatial resolution of 3.5 m per pixel. The main background is similar to the ABU dataset, and airplanes in the image are considered anomaly targets.

The Salinas dataset [81] is also a subset of the AVIRIS dataset. The dataset was collected by an AVIRIS sensor with 224 channels with a high spatial resolution of 3.7-meter pixels and dimensions of  $512 \times 217$  pixels. The 20 removed water absorption bands included 108–112, 154–167, and 224.

The Hydice dataset is a subset of the Urban dataset originating at the US AG center [85]. The dataset was collected by a sensor [86] with a spectral resolution of 10 nm, capturing 162 bands that cover wavelengths ranging from 400 to 2500 nm. The dataset has dimensions of  $80 \times 100$  pixels. The preprocessing done to the dataset included the removal of bands due to atmospheric effects and dense water vapor. The discarded bands included the following bands: 1–4, 76, 87, 101–111, 136–153, and 198–210.

The Arizona dataset, introduced by Watson et al. [15], features a staged scene at the Santa Rita Experimental Range imaged in December 2022. The scene was captured by an unmanned aerial vehicle (UAV) that carries a Pika L sensor manufactured by Resonon [87]. This dataset from Pika L has 281 spectral bands ranging from 400 nm to 1000 nm in wavelength and a spatial resolution of 0.1 m per pixel. The UAV captured the scene from an altitude of 40 m and a speed of 10 m/s. The dataset contains five image scenes of varying sizes from  $358 \times 293$  to  $849 \times 291$ . The main background includes soil, vegetation, and water. Vehicles, plastic inflatables, tarps,

and miscellaneous foreign objects are considered anomaly targets in this dataset.

### 3.3.2 Evaluation Metrics

Traditional classification evaluation metrics could be applied to evaluate the proposed GE-AD algorithm. An example of those metrics is the ROC (receiver operating characteristic) curve. It visually represents a binary classifier model's efficacy across various classification thresholds. The ROC curve plots the true positive rate (TPR) versus the false positive rate (FPR) at each threshold. The decrease in the classification threshold classifies more items as positive, which may increase both false positives and true positives. AUC (area under the ROC curve) [88–90] measures the total two-dimensional area under the ROC curve, comprehensively evaluating the model's performance across all potential classification thresholds.

Another example is the F1 score [91], a harmonic mean of precision and recall. The equations for precision and recall are shown below:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Here,  $TP$  is the number of actual positives predicted as positive,  $FP$  is the number of actual negatives predicted as positives, and  $FN$  is the number of actual positives predicted as negatives. Because of class imbalance in our dataset (many more negatives than positives), the general F1 score, as shown in Equation (3), will not provide us the perfect insight into the algorithmic performance. With the need for different metrics understood, we used the macro-averaged F1 (or F1-macro) score, as shown in Equation (4). The macro-average calculates the metric independently for each class and then computes the average, treating all classes equally. This approach makes it simpler to demonstrate the impact of improved anomaly detection. Here,  $N$  is the number of classes. It is computed by taking the mean of all the classes'

F1 scores. In this case,  $N$  is two as we have two classes.

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$F1\_macro = \frac{\sum_{n=1}^N F1\_score_n}{N} \quad (4)$$

We note that when evaluating binary classifiers on *imbalanced* datasets, the false positive rate is less informative than precision in assessing the quality of a classifier [92]. Since anomaly detection typically involves highly imbalanced data (many more negatives than positives), the F1-macro score is more suitable than the ROC AUC metric.

### 3.3.3 Comparison with Individual AD Methods

We evaluate our GE-AD algorithm using consistent stratified  $5 \times 2$ -fold cross-validation. 2-fold cross-validation randomly selects 50% of the pixels as training data and the other 50% for testing.

In this evaluation with ABU-Airport dataset, the GE-AD algorithm selected the Abundance, AED, FCBAD, and KRX methods as the base models for the ensemble. Subsequently, the individual predictions from these selected algorithms were supplied to the meta-model for the final ML model training and detection. We trained a model with seed 842, which achieved a training F1-macro score of 0.81 and a test F1-macro score of 0.86, a higher F1-macro score than any of the methods used in the ensemble. We are calling this model trained on the ABU dataset GE-AD model-1. GE-AD model-1 produces fewer false positives and creates a clear detection map with all the targets at least partially detected, as shown in Figures 17 from the ABU dataset.

The null hypothesis of the Wilcoxon signed-rank test [93] is that the rank sums of the control algorithm **are not significantly less** than those of the other base models used in the ensemble. If the  $p$ -value is less than the significance level (0.05), the null hypothesis can be

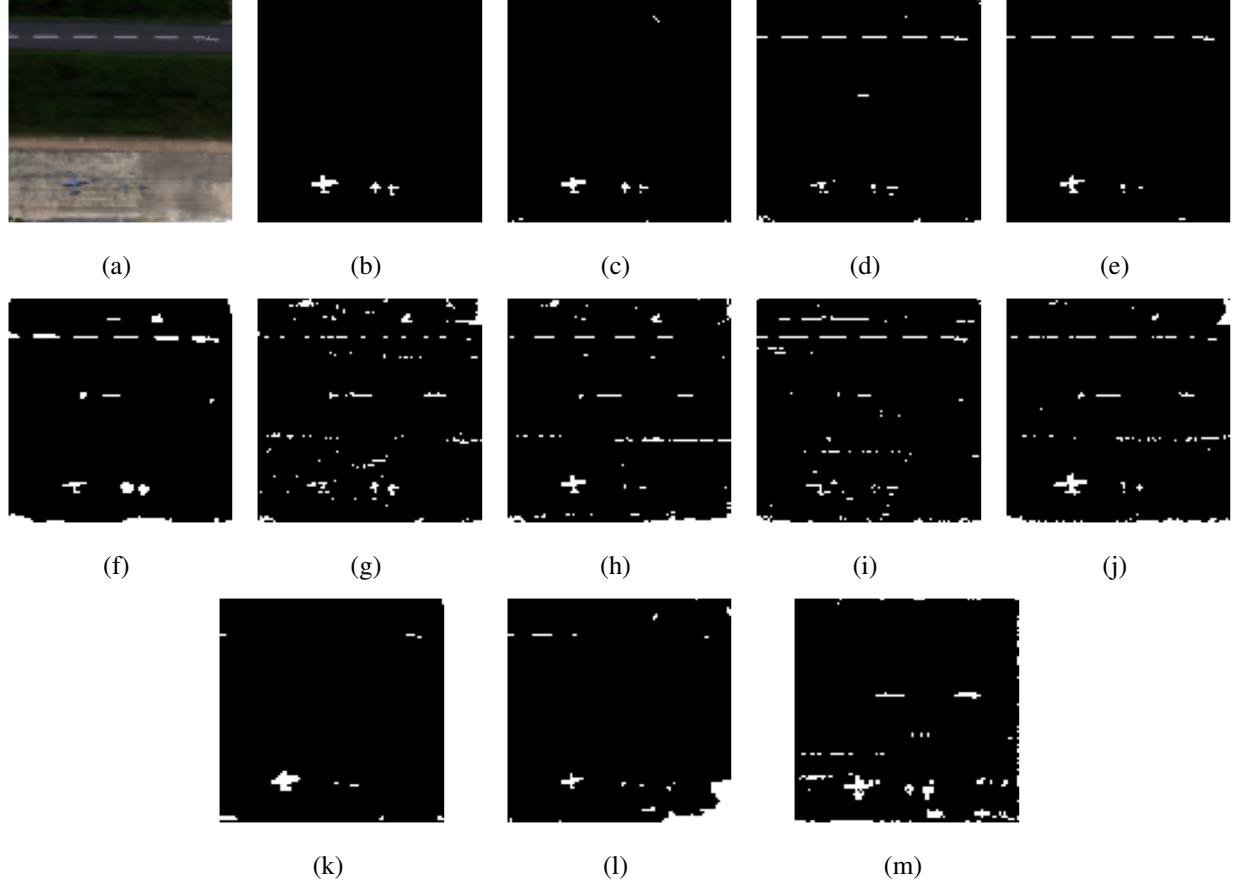


Figure 17: Visual comparison showing better performance of our proposed ensemble method (GE-AD) compared to other methods using ABU-IV data. **(a)** RGB, **(b)** ground truth, **(c)** GE-AD ( $F_1 = 0.856$ ), **(d)** HUE-AD ( $F_1 = 0.649$ ), **(e)** Abundance ( $F_1 = 0.735$ ), **(f)** AED ( $F_1 = 0.610$ ), **(g)** KIFD ( $F_1 = 0.587$ ), **(h)** KRX ( $F_1 = 0.604$ ), **(i)** LSUNRSORAD ( $F_1 = 0.569$ ), **(j)** FCBAD ( $F_1 = 0.631$ ), **(k)** ERRX MF ( $F_1 = 0.666$ ), **(l)** ERCRD ( $F_1 = 0.545$ ), and **(m)** SED ( $F_1 = 0.592$ ). Please note that, as the implementations of algorithms for Figures **(k–m)** were not publicly available, we binarized the original outputs from the authors' papers [54, 56, 57], respectively.

rejected, and the result is statistically significant. We ran our tests ten times and used all the ABU results to run significance tests. In Table 3, GE-AD has the lowest rank, and all the methods have a  $p$ -value less than 0.05; thus, the null hypothesis can be rejected, and our process is better and differs significantly from the other methods.

For comparative analysis, we applied the Friedman rank test [94], which measures the significance of the performance differences among the detection methods considered. The null hypothesis of the Friedman rank test is that the rank sums **do not differ** significantly between the three or more “paired” samples, and they are identical. As in the case of the Wilcoxon test, if the  $p$ -value is less than the significance level (0.05), the null hypothesis can be rejected, and the result is statistically significant. The Friedman test for F1-macro is one-tailed with a  $p$ -value of 0.05, and ROC AUC is two-tailed [95–97] with a  $p$ -value of 0.05/2. Table 3 shows our method is better and differs significantly from the other methods for the F1-macro metric, with the best average rank of 1.07 (lower is better), and the  $p$ -value is  $3.397^{-26}$  (numerical zero), which is less than 0.05.

In addition to the F1-macro metric, we also determined the ROC AUC scores [88] for each algorithm as the authors of other papers evaluated their work using this metric. The results are summarized in Figure 18. Note that even though GE-AD model-1 resulted in outputs that are visually closer to the ground-truth data, it failed to achieve a higher ROC AUC score compared to all the methods used in the ensemble. This demonstrates that a single metric is not sufficient to assess the performance.

As shown in Figures 17 and 18, the proposed method outputs results that are visually closer to the ground-truth data with a lower ROC AUC score; we still conducted the statistical tests on the ROC AUC score and obtained an interesting result. Our evaluation shows that our method is statistically significant for the ROC AUC score over the ten runs. The  $p$ -value of AED in the one-sided Wilcoxon signed-rank test is 0.038011, much higher than the others but still less than the significance level (0.05). Table 4 shows that the rank of the proposed method is the lowest, and the  $p$ -value in the Friedman rank test is  $3.9^{-22}$ , which is less than 0.05/2.

Table 3: Comparison of F1-macro scores showing our proposed ensemble method (GE-AD) performs better compared to HUE-AD, and individual methods using the ABU dataset.

<b>Method</b>	<b>ABU-I</b>	<b>ABU-II</b>	<b>ABU-III</b>	<b>ABU-IV</b>	<b>Avg. F1 -macro Score</b>	<b>Wilcoxon's <i>p</i>-Value [93]</b>	<b>Friedman's <i>p</i>-Value [94]</b>
GE-AD [12]	0.887	0.878	0.875	0.769	0.852 ± 0.087		1.07
HUE-AD [14]	0.793	0.770	0.751	0.649	0.741 ± 0.058	$1.9262 \times 10^{-8}$	2.62
Abundance [63]	0.670	0.741	0.760	0.736	0.727 ± 0.038	$3.033 \times 10^{-8}$	3.10
AED [49]	0.781	0.646	0.697	0.610	0.683 ± 0.065	$1.7847 \times 10^{-8}$	4.40
FCBAD [48]	0.606	0.603	0.591	0.621	0.605 ± 0.014	$1.7847 \times 10^{-8}$	6.75
KIFD [53]	0.694	0.682	0.695	0.587	0.664 ± 0.046	$1.7847 \times 10^{-8}$	5.05
KRX [44]	0.559	0.658	0.659	0.604	0.620 ± 0.043	$1.7847 \times 10^{-8}$	6.42
LSUNRS ORAD [51]	0.695	0.608	0.661	0.568	0.633 ± 0.049	$1.7847 \times 10^{-8}$	6.58

Table 4: Performance results of all methods regarding AUC metric (%) along with Friedman rank using the ABU-airport dataset.

<b>Method</b>	<b>ABU-I</b>	<b>ABU-II</b>	<b>ABU-III</b>	<b>ABU-IV</b>	<b>Avg. AUC Score</b>	<b>Wilcoxon's <i>p</i>-Value [93]</b>	<b>Friedman's <i>p</i>-Value [94]</b>
GE-AD	0.987	0.980	0.975	0.978	0.980 ± 0.018		2.00
Abundance	0.620	0.733	0.834	0.805	0.748 ± 0.085	$1.7847 \times 10^{-8}$	7.15
AED	0.992	0.980	0.975	0.932	0.970 ± 0.025	$3.8011 \times 10^{-2}$	$3.901 \times 10^{-22}$
FCBAD	0.900	0.896	0.889	0.978	0.916 ± 0.038	$1.4139 \times 10^{-7}$	4.28
KRX	0.914	0.971	0.951	0.975	0.953 ± 0.025	$2.7606 \times 10^{-5}$	3.55

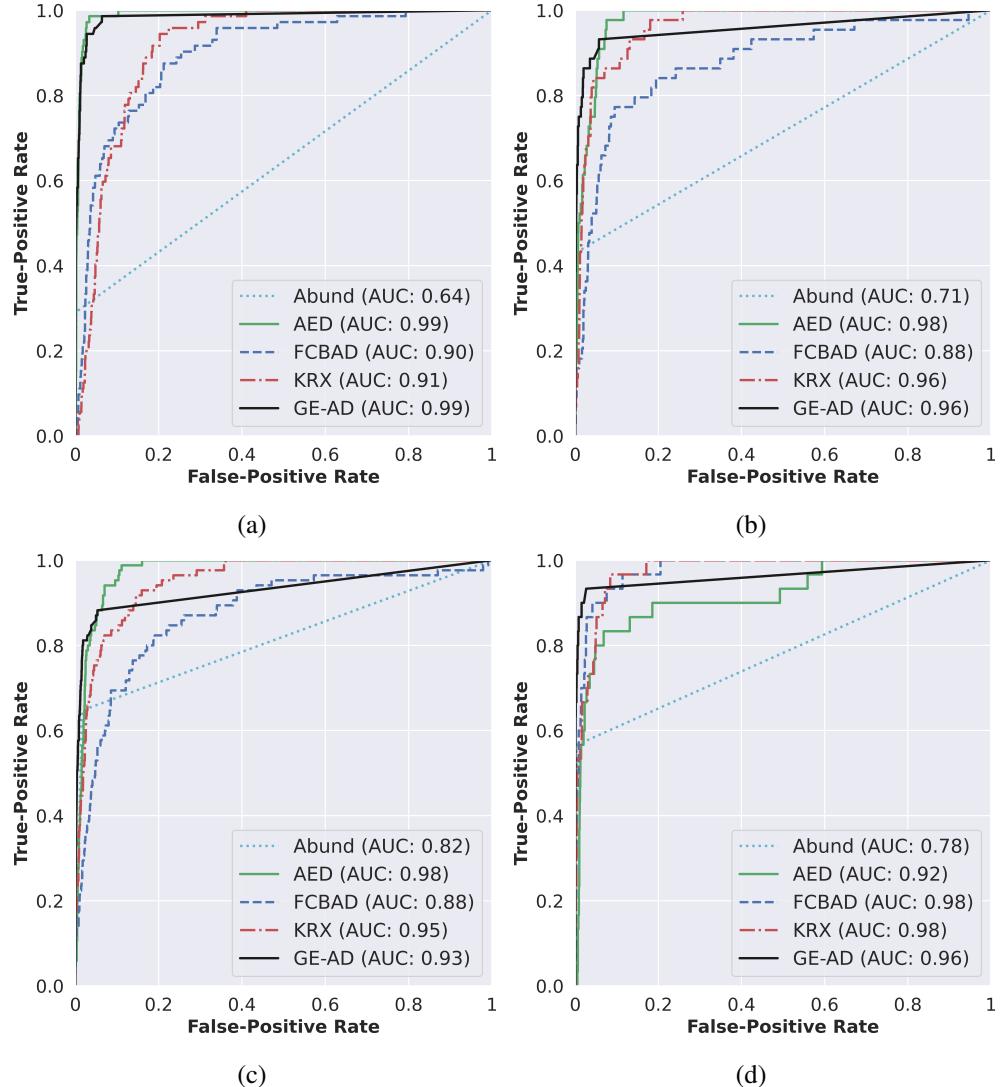


Figure 18: Detection accuracy evaluation through ROC curve and AUC scores (shown in the legend) of (a) ABU-I, (b) ABU-II, (c) ABU-III, and (d) ABU-IV.

### 3.3.4 Comparison with Baseline SOTA Ensemble Methods

We broaden our comparative analysis by elaborating on the performance results of several existing ensemble AD methods. Our GE-AD model-1 struggled to achieve a higher ROC AUC score compared to the other individual methods. Still, it achieved a considerably similar ROC AUC score to the other SOTA ensemble anomaly detection algorithms, as shown in Table 5. However, our previous work [14] discussed how the F1-macro score is a better-suited metric to evaluate anomaly detection tasks as a small number of incorrect or correct predictions can significantly change the AUC score. Additionally, the AUC score does not consider the impact of false positives. Hence, this study has included the ROC AUC score, the F1-macro score, and their corresponding visual comparisons. Our method outperforms the other ensemble methods both visually and in terms of the F1-macro score, as shown in Table 5 and Figure 17, while it provides comparable scores for the ROC AUC.

Table 5: ABU-IV quantitative performance (as reported in the authors' papers [54, 56, 57]) comparison between our proposed ensemble method (GE-AD) and other ensemble methods. Here, the best scores are shown in bold.

	GE-AD	ERRX MF [56]	ERCRD [54]	SED [57]
ROC AUC score	0.9632	0.9970	0.9533	<b>0.9980</b>
F1-macro score	<b>0.856</b>	0.666	0.545	0.592

## Chapter 4

### Generalization of Hyperspectral Anomaly detection

Universality or generalization suggests that if a method achieves promising results on known datasets, it will perform similarly well for new or unseen data and provide consistent predictions. Usually, unsupervised HS-AD methods promise generalization as they fit the given unknown test data before delivering the results. Muhammad et al. [98] evaluated their model's generalization by comparing the performance metrics with other state-of-the-art methods trained on the target dataset and tested only on the test dataset. We also adopted this evaluation methodology. In our previous evaluation [12], the GE-AD algorithm trained on the San Diego Airport dataset selected the Abundance, GMRX, KIFD, and KRX methods as the base input components for the ensemble. It achieved a test F1-macro score of 0.883 and 0.910 on two images, respectively. We have used our GE-AD model trained using the ABU-Airport dataset to test the San Diego dataset and evaluate the model's generalization performance. Figure 19 shows the results of the San Diego Airport data 02. The model achieved F1-macro scores of 0.822 and 0.826. This shows the generalization of our GE-AD model trained using the ABU-Airport for the San Diego Airport dataset. This model's performance is similar to our previous evaluation without seeing any San Diego Airport dataset before testing.

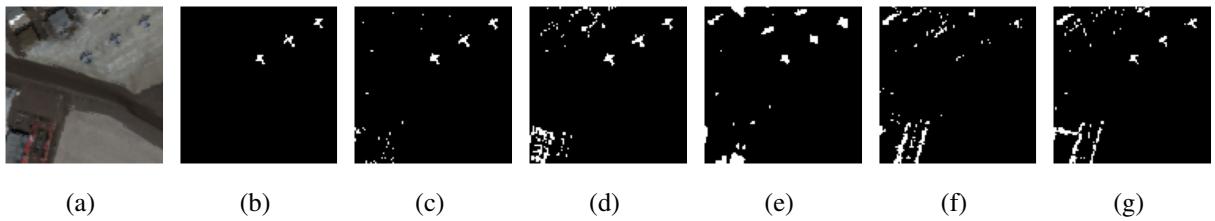


Figure 19: The visualization shows the generalization of our proposed ensemble method (GE-AD) trained on the ABU dataset compared to other methods using San Diego-02 data. (a) RGB, (b) ground truth, (c) GE-AD ( $F1 = 0.826$ ), (d) Abundance ( $F1 = 0.655$ ), (e) AED ( $F1 = 0.636$ ), (f) FCBAD ( $F1 = 0.536$ ), and (g) KRX ( $F1 = 0.627$ ).

We have also tested the same GE-AD model using the Arizona dataset to evaluate the generalization performance. In our previous evaluation [12], the GE-AD algorithm trained on

the Arizona dataset selected the FCBAD, KIFD, KRX, and LSUNRSORAD methods as the base input components for the ensemble. It achieved a test F1-macro score of 0.770, 0.799, 0.882, 0.816, and 0.859 on five images, respectively. Figure 20 shows the result of the GE-AD model trained using the ABU-Airport dataset and tested on Arizona-V data. It achieved F1-macro scores of 0.490. Abundance, AED, FCBAD, and KRX methods are used for the ABU-Airport dataset. This F1-macro score of 0.490 is lower than those of the input methods used in the ensemble trained on the ABU-Airport dataset. It is also lower than our previous evaluation of 0.859 which selected FCBAD, KIFD, KRX, and LSUNRSORAD methods as base methods for the Arizona dataset. Some base methods are different, and the importance assigned by the meta-method for each base method differs significantly between the two datasets as we can see in Fig. 21. When we tested the importance of the Arizona dataset feature in the model trained using the ABU-Airport dataset based on feature permutation, we found that this information had no negative impact on the model (see Fig. 21c). Thus, the GE-AD model trained using the ABU-Airport fails to generalize for the Arizona dataset.

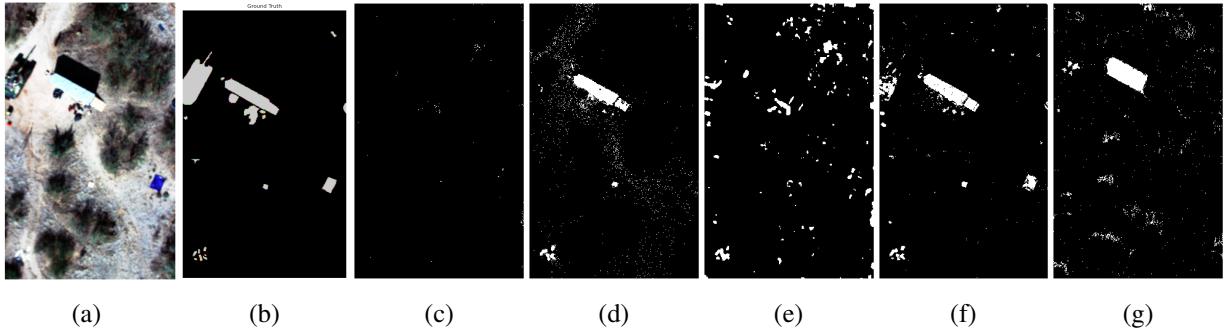


Figure 20: The visualization shows the performance of our proposed ensemble method (GE-AD) trained on the ABU dataset compared to other methods using Arizona-V data. (a) RGB, (b) Ground Truth, (c) GE-AD ( $F1 = 0.490$ ), (d) Abundance ( $F1 = 0.701$ ), (e) AED ( $F1 = 0.584$ ), (f) FCBAD ( $F1 = 0.810$ ), (g) KRX ( $F1 = 0.503$ ).

A supervised classifier usually performs well when tested on data from the same dataset with a similar distribution to the training dataset. However, their performance tends to drop when tested on other unfamiliar datasets. GE-AD generalizes some similar unseen data, as both the ABU Airport and the San Diego Airport dataset contain scenes from airports dissimilar to those

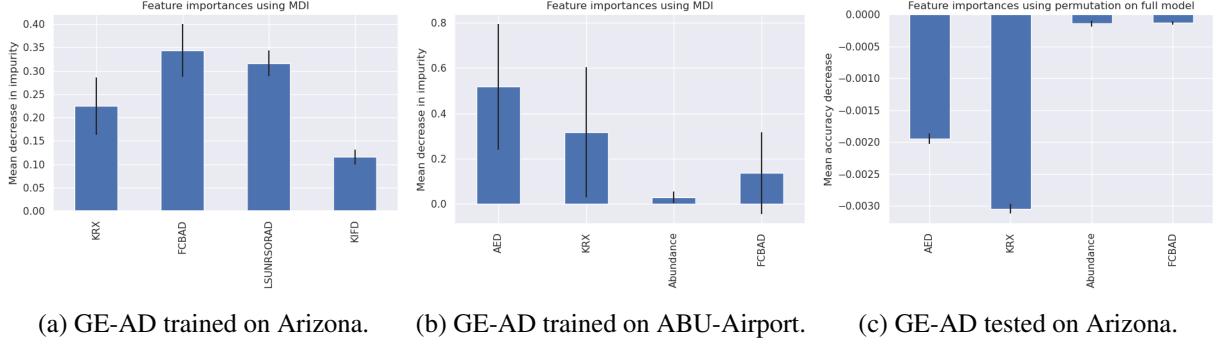


Figure 21: Visualization of difference of feature importance between models.

of the Arizona dataset. Also, the Arizona dataset is collected using different HS sensors. Thus, based on our evaluation, the GE-AD model does not always generalize. It is worth investigating the problem to find a generalized solution that people can readily deploy for AiTR and RS problems.

#### 4.1 Previous Work

Wang et al. [99] generalized a zero-shot learning problem and presented a unified domain adaptation framework for both unsupervised and zero-shot learning conditions. The authors assume they don't have target domain data. Unlike theirs, we consider our target domain data available without labels.

Muhammad et al. [98] proposed a comprehensive solution combining synthetic data generation and deep ensemble learning to enhance face Presentation Attack Detection (PAD) generalization capabilities. They evaluated four PAD datasets and used Half Total Error Rate (HTER) and ROC-AUC score as metrics. A lower HTER signifies that the system makes fewer errors in accepting incorrect identities (false positives) and rejecting correct identities (false negatives). They ran four experiments, using three as training and the other as testing. They compared other state-of-the-art methods tested only on the test dataset. We have integrated their evaluation technique into ours. However, synthetic data generation and deep ensemble are more

complex tasks than the data augmentation and random forest ensemble we have in our current design. We plan to keep our design simple and investigate complex solutions if simpler ones do not work.

Meegahapola et al. [100] proposed a new algorithm called M3BAT using an unsupervised adversarial-based domain adaptation (DA) method. They used an encoder as a feature extractor and two discriminators to identify source and target datasets. In domain adaptation, the dataset changes, not the label. In our case, we want to identify the same set of anomalies in the new dataset. Domain adaptation methods may apply here as our dataset changes, not the target.

Pillai et al. [101] used semi-supervised DA. Semi-supervised DA helps if there are limited labels in target datasets. This DA method, Subspace Alignment (SA), aligns the statistical measures of features of different domains. They also explored multiple other DA methods and evaluated their impacts.

Mancini et al. [102] proposed an ensemble learner with domain predictor to predict the probability of a sample belonging to each domain (weights). Their domain predictor to predict the likelihood of a sample belonging to each domain is intriguing, and we plan to investigate it in our ensemble learners. Ryu et al. [103] investigated the generalizability in random forests. As we have random forests as our meta-model, we can extend their solutions in our scenario.

I plan to investigate the Subspace Alignment (SA) approach adopted by Pillai et al [101] as SA solves a problem similar to our case, and if successful, we do not need to try complex GAN-like adversarial models such as M3BAT [100]. SA methods try to align the statistical measures of features of different domains. Their idea [101] of sub-sampling to find related examples in the source domain is intriguing. It helps a model become data agnostic, reducing the training time without sacrificing performance in a new dataset.

## 4.2 Preliminary Work

### 4.2.1 Replacing Supervised Meta Method in GE-AD with Unsupervised Method

For our first evaluation, we replaced the Random Forest with an unsupervised Mixture Model (GMM) as the meta-model.<sup>1</sup> In our evaluation, we found that it has an AUC ROC score of  $0.943 \pm 0.088$  and F1-macro score of  $0.548 \pm 0.048$  among all datasets (see Section 4.2.3). We plan to investigate other unsupervised models in the future. We also plan to apply all options we used in improving our supervised GE-AD [12] to this new unsupervised GE-AD (UGE-AD) model and investigate their impacts.

### 4.2.2 Adding Spectral Information as Input to GE-AD

As GE-AD uses only the results from base models, the spatial and spectral information from an HS image is not visible. GE-AD can compute different weights for different datasets based on features from the input data. However, we cannot feed all the channels to the our model due to two concerns: (a) this approach may hinder our model from learning essential features from the base models; and (b) the curse of dimensionality leads to increased computational complexity, overfitting, and spurious correlations. As an initial investigation, we modified GE-AD (mGE-AD) to take thirty random spectral channels and four other base methods as input [13]. In our evaluation, we found that it has an AUC ROC score of  $0.593 \pm 0.070$  and F1-macro score of  $0.548 \pm 0.048$  among all datasets (see Section 4.2.3). Instead of randomly selecting spectral channels, we plan to use dimensionality reduction techniques similar to section 5.4.

### 4.2.3 Evaluation Results

For our evaluation, we used only one dataset in the greedy search to find suitable base methods for UGE-AD to maximize the ROC-AUC score. The evaluation results are shown in

---

<sup>1</sup>We also tested the isolation tree, which did not work.

Tab. 6. The results show that the greedy search may fail to find the optimum base methods for a smaller dataset, and the results are generalizing more for larger datasets. As we can see, the best test score for the Salinas dataset comes from a model trained using the ABU airport and urban dataset. Table 6 shows that the performances are correlated to input base methods; for example, GM-RX with UGE-AD achieves a favorable performance score in six out of eleven best score cases. Another issue was that greedy search could not find a better score for the Arizona dataset. GM-RX alone was enough for the Arizona dataset, as seen from Tab. 6. Figure 22 shows the results from the test for better visualization and understandability.

Table 6: The average ROC-AUC scores of our newly proposed Unsupervised ensemble UGE-AD over various datasets. The first column shows the dataset used in the greedy search to find suitable base methods, and the second column shows those base methods. Where the dataset names in rows and columns match, those cells show the training score after the greedy search. The boldface marks the best performance in each test dataset column.

		HYDICE Salinas		San Diego	ABU airport	ABU beach	ABU urban	Arizona
HYDICE	WIN-RX	<b>0.997</b>	0.951	0.962	0.943	0.963	0.964	0.802
Salinas	LSUNRSRAD, RX	0.995	0.999	0.965	0.949	0.971	0.971	0.780
San Diego	GM-RX	<b>0.997</b>	0.997	0.964	0.934	0.961	0.962	<b>0.847</b>
ABU airport	AED, GM-RX, KIFD, LSUNRSRAD	0.995	<b>1.000</b>	0.971	<b>0.962</b>	<b>0.985</b>	0.966	0.712
ABU beach	KIFD, LSUNRSRAD, RX, WIN-RX	0.996	0.997	0.965	0.960	0.973	<b>0.976</b>	0.766
ABU urban	AED, KIFD, LSUNRSRAD, RX	0.996	<b>1.000</b>	0.970	<b>0.962</b>	<b>0.985</b>	0.971	0.706
Arizona	CBAD, FCBAD, GM-RX, KIFD	0.994	0.999	<b>0.972</b>	0.948	0.967	0.966	0.779

For our second evaluation, we used at most fifty percent of datasets in the greedy search to find suitable base methods for mGE-AD and UGE-AD to maximize the ROC-AUC score.

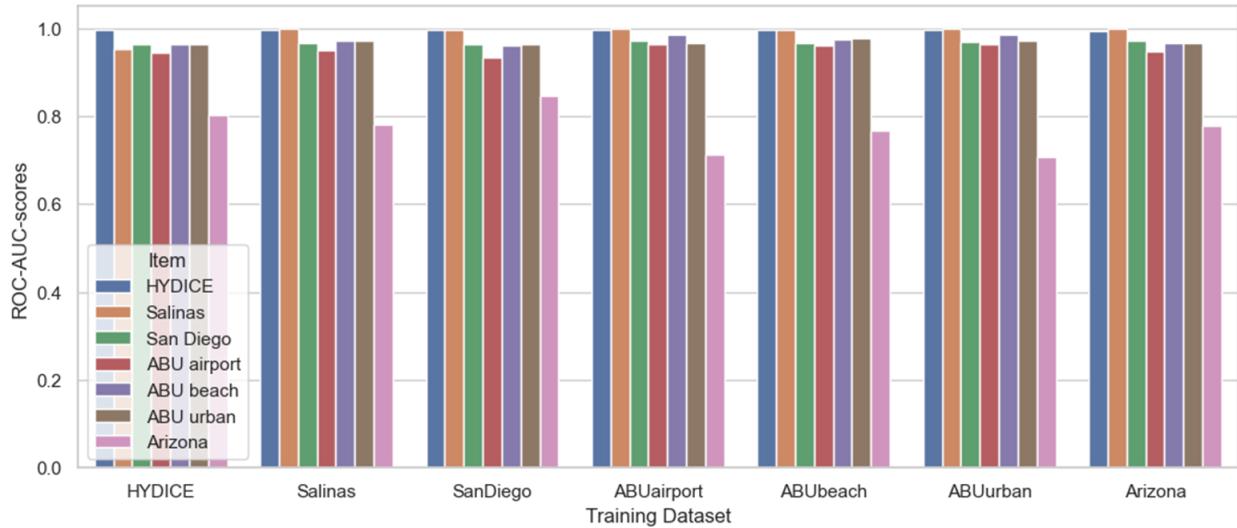


Figure 22: Average ROC-AUC scores of our UGE-AD over various datasets showing consistent performance against various training datasets. The X-axis shows the dataset used in the greedy search to find suitable base methods. The Y-axis shows the testing ROC AUC scores. Except where the dataset names in rows and columns match, those bars show the training score after the greedy search.

Table 7: The average and standard deviation of our ensemble methods’ ROC-AUC scores computed over the test datasets. We generated this table using a model trained on randomly selected, at most, fifty percent of the datasets from the ABU, Arizona, HYDICE urban, Salinas, and San Diego datasets as training and evaluated on other datasets. The boldface marks the best performance in each column.

Methods	HYDICE urban	Salinas	San Diego	ABU airport	ABU beach	ABU urban	Arizona	Average score
baseline								
HUE-AD (AED, KIFD, LSUNRSORAD)	0.962	<b>1.000</b>	<b>0.988</b>	<b>0.980</b>	<b>0.986</b>	<b>0.987</b>	0.411	$0.902 \pm 0.201$
mGE-AD								
(GM-RX, KIFD, LSUNRSORAD, MD-RX), RF	0.927	0.997	0.939	0.907	0.965	0.970	0.472	$0.882 \pm 0.170$
UGE-AD (AED, FCBAD, GM-RX, KIFD), GMM	<b>0.995</b>	0.999	0.970	0.960	0.980	0.968	<b>0.731</b>	$0.943 \pm 0.088$

Table 8: The average and standard deviation of our ensemble methods' F1-macro scores were computed over the test datasets. We generated this table using a model trained on randomly selected, at most, fifty percent of the datasets from the ABU, Arizona, HYDICE urban, Salinas, and San Diego datasets as training and evaluated on other datasets. The boldface marks the best performance in each column.

Methods	HYDICE urban	Salinas	San Diego	ABU airport	ABU beach	ABU urban	Arizona	Average score
baseline HUE-AD	0.519	0.512	<b>0.619</b>	<b>0.629</b>	0.564	<b>0.663</b>	0.541	$0.578 \pm 0.055$
mGE-AD	<b>0.525</b>	<b>0.729</b>	0.560	0.584	<b>0.628</b>	0.624	0.503	$0.593 \pm 0.070$
UGE-AD	0.500	0.480	0.559	0.566	0.514	0.605	<b>0.613</b>	$0.548 \pm 0.048$

Compared to our first evaluation, we want to review whether a model can perform better when it sees multiple datasets during training. Table 8 shows the F1-macro scores, and Table 7 shows the ROC-AUC scores of our ensemble methods evaluated over the leftover datasets. We ran our experiment multiple times and stored the performance scores whenever the dataset was in the test fold. We reported the average of those scores, the overall average score, and the standard deviation. When comparing Tab. 6 and 7, UGE-AD trained on one dataset could achieve higher ROC-AUC scores than learning from multiple datasets. However, these scores in Tab. 6 do not come from the same model. The unsupervised UGE-AD method shows better ROC-AUC scores but lower F1-macro scores than our mGE-AD. Figure 23 shows the qualitative difference of UGE-AD with multiple false-positives that resulted in lower F1-macro scores than mGE-AD.

Finally, from Tab 4, our original GE-AD trained on 50% of the pixels from ABU-airport data and tested with the other 50%. It achieved a better ROC-AUC score of  $0.980 \pm 0.018$  than UGE-AD of 0.960, as seen in Tab. 7. The performance of UGE-AD is inline with the path of generalization.

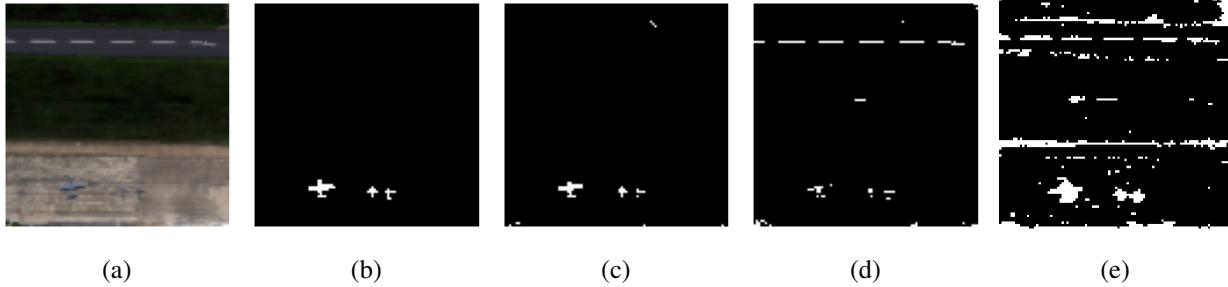


Figure 23: The visualization shows the qualitative difference between our proposed ensemble method (UGE-AD) and our other methods using ABU-airport IV data. (a) RGB, (b) ground truth, (c) GE-AD, (d) HUE-AD, (e) UGE-AD.

### 4.3 Further Research Plan

Figure 22 shows that unsupervised models provide a stable result for each dataset irrespective of the training dataset with slight variation, except for the Arizona dataset. On the other hand, we can modify a supervised model more straightforwardly using ground truth. Thus, we plan to evaluate both types for our meta-learner and investigate other supervised and unsupervised models.

Compared to transfer learning and domain adaptation that assume the availability of target domain data, Domain Generalization (DG) takes a step further and does not require access to target data [104]. We can use DA (domain adaptation) for our semantic segmentation problem in chapter 5.2, as the complete dataset is available during the model’s devising and training. DA method SA (Subspace Alignment) aligns the statistical measures of features of different domains. We will investigate when to apply DA methods. We can use them for spectral information before providing them to the base methods of our GE-AD algorithm. We have investigated providing spectral information directly to the meta-model. We will explore dimensionality reduction techniques to reduce spectral information and apply DA and DG methods here to inspect its impact. We will also use these DA and DG methods to outputs from base methods without exposing spectral information to the meta-model.

Here, the challenge of DG (Domain Generalization) is that the target dataset is

unavailable during training, which is usually the case in real life. DG aims to learn a generalized model from one or several training domains with different probability distributions that can achieve good out-of-distribution generalization. After completing the DA (domain adaptation) investigation, I will expand our investigation on DG if I have sufficient time and support from others.

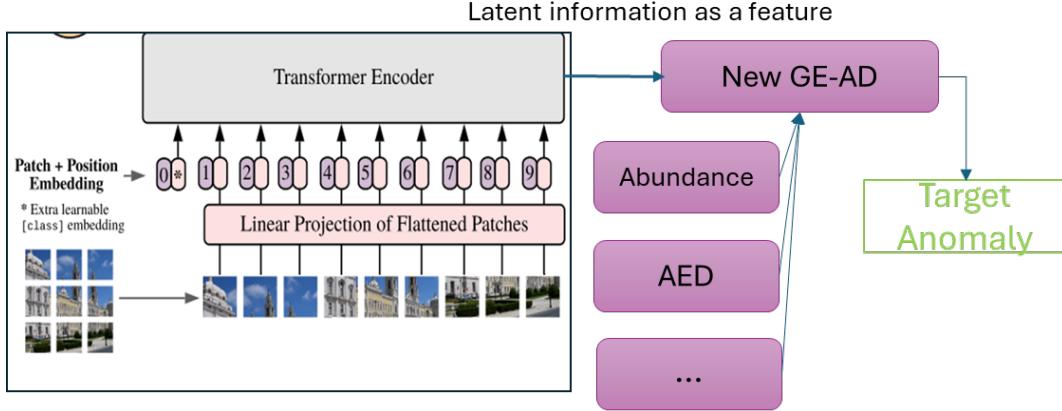


Figure 24: The initial design for generalized GE-AD.

Alternatively, we will investigate ensemble learners with domain predictor [102] to predict the probability of a sample belonging to each domain (weights). I will investigate this if I fail to achieve the expected results using the first proposal. I propose an initial system in Fig. 24 that takes latent information from a transformer as a feature for the meta-model to decide the domain source and weightage of other inputs. As we run iterations of training and validation, we expect the meta-model to learn to differentiate different datasets. If both the previous cases fail, we will also investigate the generalizability in random forests. We have random forests as our meta-model, and we can extend their solutions [103] in our scenario. Although various researchers are contributing to this active generalization field, we can contribute to generalizing the HS AD problem through our work. We will do ablation studies for the model architecture and update the design based on evaluation results.

## Chapter 5

### Hyperspectral Image Semantic Segmentation

Image Semantic Segmentation is a computer vision technique that classifies each pixel based on its semantic class. When applications require precise object localization and boundary delineation, such as military surveillance, semantic segmentation identifies and classifies every pixel within an image. It allows us to understand the detailed composition of a scene by distinguishing different objects at a granular level. It is also valuable for applications in other fields, like autonomous driving, medical imaging, and industrial inspection, where precise object localization and boundary delineation are also crucial.

HS Image Semantic Segmentation (SS) utilizes spectral and spatial information to solve this semantic segmentation problem. When an anomaly detector identifies targets in a scene, AiTR (Aided Target Recognition) wants to interrogate the scene further and distinguish target objects at a granular level. HS Image SS helps AiTR further interrogate a scene. Image SS is a well-investigated problem, and various high-performing deep convolutional neural networks like U-Net [105], FastFCN [106], and DeepLab [107] and Transformer based models like Segmenter [108] and ViT [109] are readily available. However, AiTR cannot use these models, which are designed and trained with RGB images, directly for HS image semantic segmentation as these HS images have a higher number of channels. Crop mapping is a good use case for investigating this problem in agricultural background. We can transfer this solution to AiTR to classify the scene for our desired objects of military interest.

Crop mapping plays a vital role in agricultural monitoring, enabling essential tasks such as crop identification, assessing plant health, and estimating crop production [110,111]. Figure 25 shows an example of crop mapping. Pinpointing the types of crops grown in specific areas and their spatial distribution offers invaluable insights for farmers and stakeholders. Remote sensing technology is currently replacing traditional, labor-intensive, on-the-ground surveys [112]. HS imagery gathered via unmanned aerial vehicles (UAV) strikes a remarkable balance by providing

better spatial and spectral data for crop mapping [113, 114].

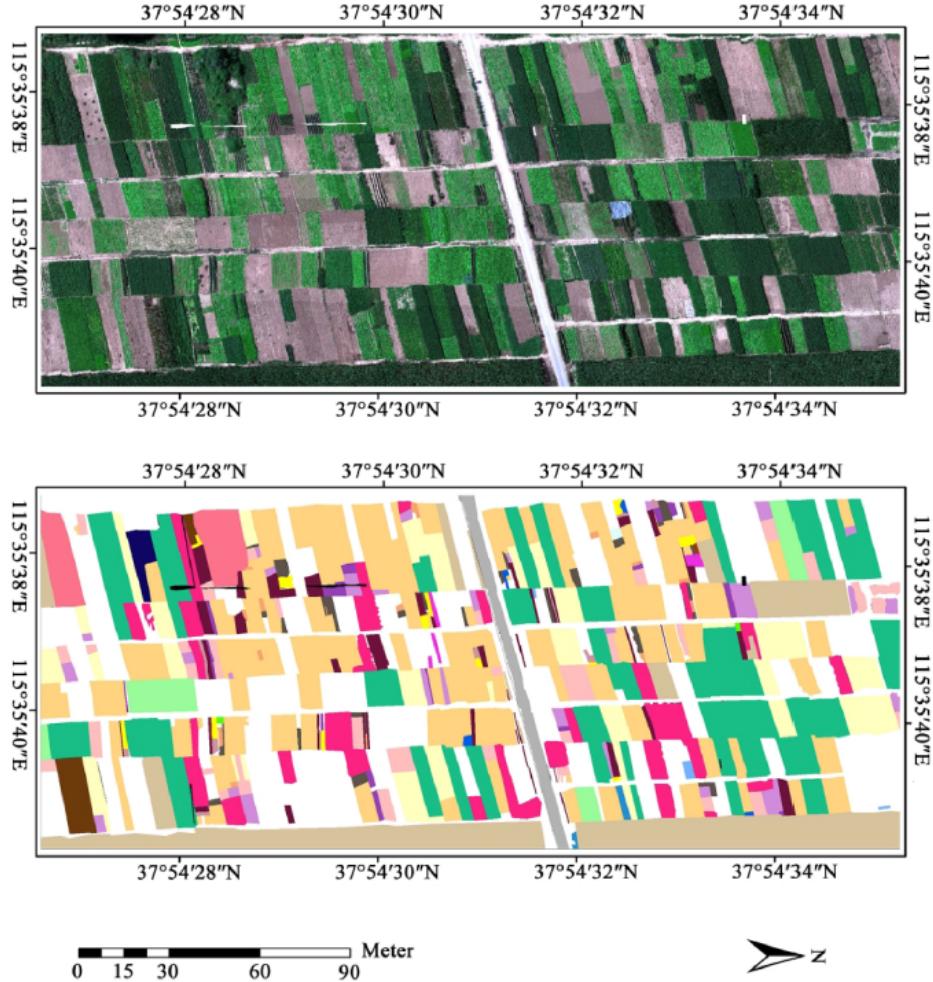


Figure 25: Example of Crop mapping [114].

One way of crop mapping is using a deep classifier, a simple task that may overfit the model [115]. However, the inference task on a scene has quadratic time complexity. It may take longer as the model classifies individual pixels in the image. In contrast, crop mapping using SS is a complex task that underfits a model without enough data. However, the model completes inference tasks in one operation, where the model classifies all the pixels simultaneously. Thus, we will investigate the SS solution considering the inference time when deployed.

## 5.1 Previous Work

U-Net is a versatile semantic segmentation network among deep networks with various data types and architecture [116, 117]. Multiple researchers have included transformers [114, 118] and attention networks [119, 120] in the U-NET. Figure 26 shows the U-Net-like Visual Transformer that Chen et al. [118] had proposed. Schlemper et al. [121] proposed an Attention-gated U-Net. Figure 27 shows their proposed U-Net. Sadly, these deep SS models cannot achieve their full potential when training from scratch using HS image datasets because there are various large RGB datasets but very limited large HS image datasets for various tasks [122].

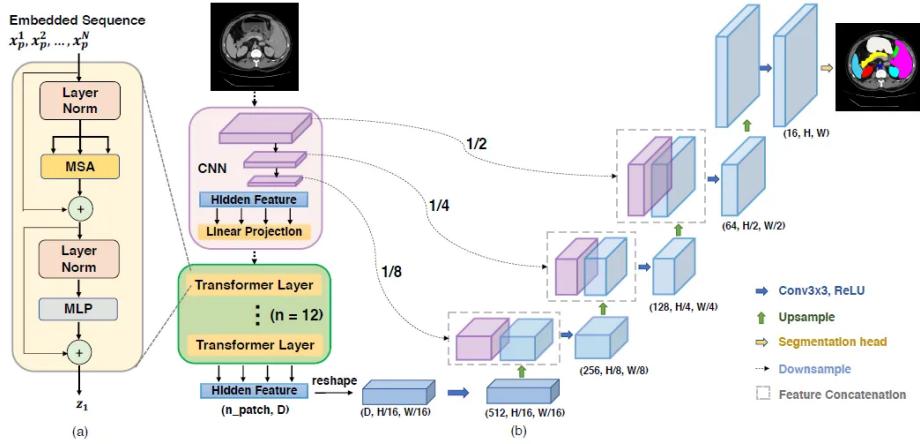


Figure 26: Overview of the TransUNet framework [118]: (a) schematic of the Transformer layer; (b) architecture of the TransUNet.

Compared to greyscale and RGB images, HS images contain hundreds of channels. Wang et al. [123] proposed a Channel Attention Module (CAM). Figure 28 shows a basic structure of their CAM. Leveraging this spectral information in the CAM can help differentiate objects better in our new U-Net. However, various HS sensors collect HS data differently, and the same sensor may produce a different number of usable channels because of noise, which hinders the way to domain adaptation. Thus, unmixing to select the most informative channels [37] ensures that our model always uses the same number of channels with the best spectral information.

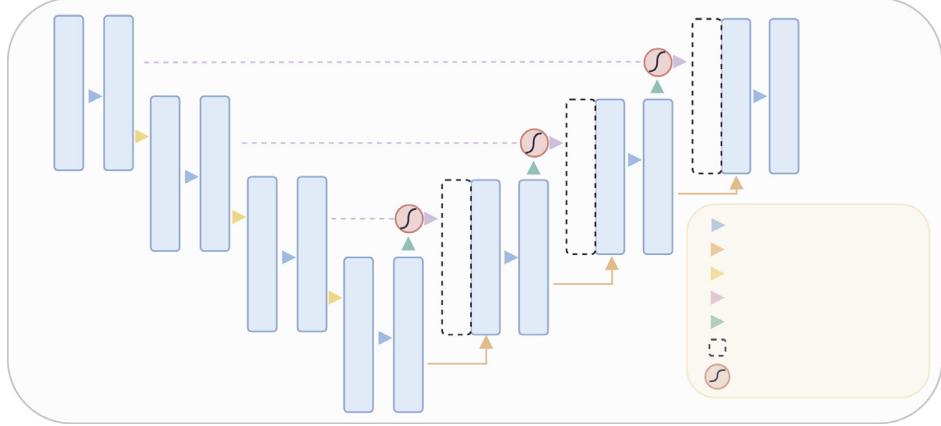


Figure 27: A block diagram of the Attention U-Net [121] segmentation model. Input image is progressively filtered and downsampled by factor of 2 at each scale in the encoding part of the network (e.g.  $H_4 = H_1 / 8$ ).  $N_c$  denotes the number of classes. Attention gates (AGs) filter the features propagated through the skip connections.

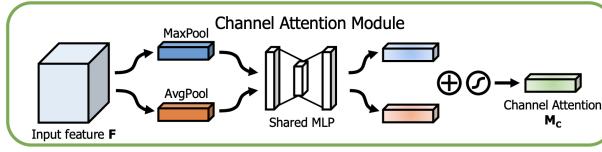


Figure 28: Channel Attention Module [123].

## 5.2 Preliminary Work

AiTR cannot use these deep semantic segmentation models, which are designed and trained with RGB images, directly for HS image SS. The main challenge here is the high dimensionality of HS images, which DL model cannot handle them effectively. One solution is updating and fine-tuning the input and final output layers. However, this can hamper their performance as middle layers still expect features from RGB. Another solution is end-to-end fine-tuning of existing DL models, which can compromise its general knowledge. Thus, I propose to use a channel attention network to utilize these hundreds of channels from HS images and extract crucial spectral information. A channel attention network is a neural network that learns to focus on vital channel information. By combining the spectral feature in the final layers of the DL model with the RGB image and fine-tuning it, we can expect to produce significantly

improved results.

Our proposal to combine an HS channel attention network with U-Net and transformers, thereby enhancing the Trans-U-Net [118] trained on a large RGB dataset, offers a promising avenue for advancing HS image SS. Compared to other models, we integrate RGB and HS images through the fine-tuning technique to create a versatile model. Using the U-Net model further simplifies the implementation and testing of our new solutions. Importantly, we design a separate HS CAM network to avoid modifying the entire pre-trained Trans-U-Net model, thereby preserving its general knowledge and stability for new tasks, providing a reliable foundation for our approach.

By leveraging task adaptation and fine-tuning, we improve the accuracy of U-Net in segmenting crops [124]. Although visual transformers provide better results than convolution neural networks, they are hard to train and need hundreds of thousands of images. However, there are no hundreds of thousands of annotated HS images available. Thus, domain adaptation is the key factor in successfully using visual transformers. This approach is convenient when the source and target domains have different tasks, and the target domain is not only related but also accessible during training. In our case, we will use pre-trained U-Net and Transformers trained on large RGB ImageNet datasets [125]. We will further fine-tune these models on the RGB and HS image dataset. Transfer learning [126] or fine-tune helps to enhance a model’s performance trained for one task and deployed for a different but related task. The original transformer cannot handle all hundreds of channels present in an HS image. In contrast, our model will be able to utilize the HS dataset.

### 5.3 Preliminary Results

Our evaluation investigated the model’s performance on large UAV HS agricultural imagery datasets [114]. Figure 29 shows results from some existing models. We have tested Trans-U-Net [118] on RGB images we created from the HS dataset as it was trained on the RGB dataset. We also trained U-Net [105] from scratch using the HS training images from this

dataset. Figure 29d shows how challenging it is to train from scratch. We plan to implement and investigate our proposed model further to improve its performance.

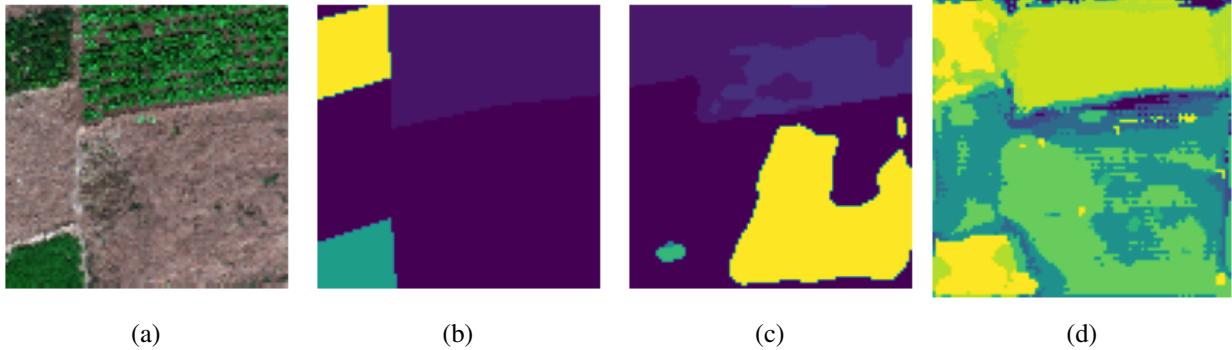


Figure 29: The visualization shows an example of semantic segmentation using an image from large UAV hyperspectral agricultural imagery datasets. (a) RGB, (b) Ground Truth,(c) Trans-U-Net [118] fine-tuned on this agricultural RGB dataset (dice coefficient 0.13), (d) U-Net [105] trained on this agricultural HS image dataset (dice coefficient 0.06).

#### 5.4 Further Research Plan

We will integrate the HS channel attention network into the Trans-U-Net to harness the spectral information from all those hundreds of channels. We will comprehensively analyze this model, focusing on complexity reduction without sacrificing performance. We will also incorporate dimensionality reduction techniques like unmixing to select the most informative channels [37], ensuring a constant number of channels with the best spectral information. This thorough approach will enable us to generalize our model across various datasets and examine the impact of fine-tuning on tiny datasets such as Indian Pines [127], which consists of just one image.

## **Chapter 6**

### **Proposed Work Timeline**

Table 9 outlines the comprehensive timeline for this proposal. I will continue investigating the impact of HS image processing on machine detection and digital RGB image processing on human visibility and understandability.

Anomaly detection in HS imaging is the leading task we should address in AiTR. Chapter 3 discusses the existing anomaly detection algorithms and our solution. Our stacking ensemble uses existing methods as base methods, reducing their variability for diverse scenes. We have published this work as a journal article in the MDPI Journal of Imaging. However, the hyperspectral sensor variability challenges our model's generalizability. Utilizing domain adaptation to enhance anomaly detection accuracy for unseen data shows great promise. By applying various domain adaptation methods within ensemble machine learning and deep learning frameworks, I, in collaboration with my advisors and research team, can develop a robust tool for detecting anomalies in HS images. Chapter 4 discusses the challenges we may encounter and presents solutions we will introduce to handle new unseen datasets and novelty anomalies. In addition, I plan to investigate this solution over two months, produce a journal article with my collaborators, and submit it to the MDPI Sensors Journal.

Semantic segmentation, a key tool in our research, can help with AiTR identification. It is crucial when investigating scenes with registered anomalous targets. We aim to identify these targets if they are our object of interest. We utilize a semantic segmentation model trained on an RGB dataset and apply task adaptation for HS images to achieve this. The use of the spectral channel attention module, which leverages information from the invisible spectral range, is a key component of our approach. This module significantly improves semantic segmentation accuracy for objects that are hard to separate from in the visible wavelength. Chapter 5 delves into the challenges we can face and presents the solution to handle the scarcity of large HS datasets and the complexity of hundreds of channels from HS images, which we have published as a

Deadline	Work
Complete	Investigate Greedy Ensemble Anomaly Detection (GE-AD)
Complete	Publish GE-AD to MDPI Journal of Imaging
Complete	Investigate Semantic segmentation on UAV HS Agricultural Imagery
Complete	Publish HS image Semantic Segmentation to SPIE D+CS conference
May 01, 2025	Start Generalized HS AD
May 31, 2025	Start Preprocess digital images
June 30, 2025	Start Impact of image processing
August 15, 2025	Submit Generalized HS AD to MDPI Sensors Journal
August 15, 2025	Work on preparing dissertation
August 30, 2025	Apply to Graduate
September 18, 2025	Defend dissertation
October 7, 2025	Submit Final copy of dissertation

Table 9: Timeline to complete my proposed disssertation works.

conference article to the SPIE D+CS conference.

Finally, I will tidy up all the chapters, update them with new findings, and submit the findings with my written dissertation.

## References

- [1] J. A. Ratches, Review of current aided/automatic target acquisition technology for military target acquisition tasks, *Optical Engineering* 50 (7) (2011) 072001–072001.
- [2] M. Harbaugh, Unmanned aerial systems (uas) for intelligence, surveillance, and reconnaissance (isr), State-of-the-Art-Report (SOAR), Defense Systems Information Analysis Center (DSIAC).  
URL <https://dsiac.org/wp-content/uploads/2018/05/UNMANNED-AERIAL-SYSTEMSUAS-FOR-INTELLIGENCE-SURVEILLANCE-AND-RECONNAISSANCE.pdf>
- [3] M. Patterson, A. Brescia, Operation of small sensor payloads on tactical sized unmanned air vehicles, *The Aeronautical Journal* 114 (1157) (2010) 427–436.
- [4] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, A. Abbas, Modern trends in hyperspectral image analysis: A review, *IEEE Access* 6 (2018) 14118–14129.
- [5] M. Paoletti, J. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: A review, *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (2019) 279–317.
- [6] G. Geography, Multispectral vs hyperspectral imagery explained, <https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>, accessed: 2024-11-23 (2014).  
URL <https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained/>
- [7] P. S. Teh, N. Zhang, A. B. J. Teoh, K. Chen, A survey on touch dynamics authentication in mobile devices, *Computers & Security* 59 (2016) 210–235.
- [8] A. Lasisi, R. Ghazali, T. Herawan, Application of real-valued negative selection algorithm to improve medical diagnosis, in: *Applied Computing in Medicine and Health*, Elsevier, 2016, pp. 231–243.
- [9] R. Nayak, U. C. Pati, S. K. Das, A comprehensive review on deep learning-based methods for video anomaly detection, *Image and Vision Computing* 106 (2021) 104078.
- [10] J. Liu, X. Wang, Plant diseases and pests detection based on deep learning: a review, *Plant Methods* 17 (2021) 1–18.
- [11] H. Su, Z. Wu, H. Zhang, Q. Du, Hyperspectral anomaly detection: A survey, *IEEE Geoscience and Remote Sensing Magazine* 10 (1) (2021) 64–90.
- [12] M. Hossain, M. Younis, A. Robinson, L. Wang, C. Preza, Greedy ensemble hyperspectral anomaly detection, *Journal of Imaging* 10 (6) (2024) 131.

- [13] M. Hossain, A. Robinson, L. Wang, C. Preza, Investigation of unsupervised and supervised hyperspectral anomaly detection, in: Applications of Machine Learning 2024, Vol. 13138, SPIE, 2024, pp. 251–261.
- [14] M. S. Younis, M. Hossain, A. L. Robinson, L. Wang, C. Preza, Hyperspectral unmixing-based anomaly detection, in: Computational Imaging VII, Vol. 12523, SPIE, 2023, p. 1252302.
- [15] T. P. Watson, K. McKenzie, A. Robinson, K. Renshaw, R. Driggers, E. L. Jacobs, J. Conroy, Evaluation of aerial real-time rx anomaly detection, in: Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXIX, Vol. 12519, SPIE, 2023, pp. 254–260.
- [16] M. K. Griffin, H.-h. K. Burke, Compensation of hyperspectral data for atmospheric effects, Lincoln laboratory journal 14 (1) (2003) 29–54.
- [17] G. P. Anderson, G. W. Felde, M. L. Hoke, A. J. Ratkowski, T. W. Cooley, J. H. Chetwynd Jr, J. Gardner, S. M. Adler-Golden, M. W. Matthew, A. Berk, et al., Modtran4-based atmospheric correction algorithm: Flaash (fast line-of-sight atmospheric analysis of spectral hypercubes), in: Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery VIII, Vol. 4725, SPIE, 2002, pp. 65–71.
- [18] T. Wang, T. Guan, F. Qiu, L. Liu, X. Zhang, H. Zeng, Q. Zhang, Evaluation of scale effects on uav-based hyperspectral imaging for remote sensing of vegetation, Remote Sensing 17 (6) (2025) 1080.
- [19] A. Maity, A. Pattanaik, S. Sagnika, S. Pani, A comparative study on approaches to speckle noise reduction in images, in: 2015 International Conference on Computational Intelligence and Networks, IEEE, 2015, pp. 148–155.
- [20] R. C. Gonzalez, R. E. Woods, Digital image processing, Pearson education india, 2009.
- [21] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Transactions on image processing 16 (8) (2007) 2080–2095.
- [22] Y. Mäkinen, L. Azzari, A. Foi, Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching, IEEE Transactions on Image Processing 29 (2020) 8339–8354.
- [23] C. C. Aggarwal, Outlier ensembles: position paper, ACM SIGKDD Explorations Newsletter 14 (2) (2013) 49–58.
- [24] A. Zimek, R. J. Campello, J. Sander, Ensembles for unsupervised outlier detection: challenges and research questions a position paper, Acm Sigkdd Explorations Newsletter 15 (1) (2014) 11–22.
- [25] T. Välikangas, T. Suomi, L. L. Elo, A systematic evaluation of normalization methods in quantitative label-free proteomics, Briefings in bioinformatics 19 (1) (2018) 1–11.

- [26] Y. Zhao, L. Wong, W. W. B. Goh, How to do quantile normalization correctly for gene expression data analyses, *Scientific reports* 10 (1) (2020) 15534.
- [27] R. Farmer, Data Transformation.  
URL <https://dataalltheway.com/posts/001-data-transformation>
- [28] S. C. Hicks, R. A. Irizarry, When to use quantile normalization?, *BioRxiv* (2014) 012203.
- [29] A. Ghadiri, Implementation of an automated image processing system for observing the activities of honey bees, Ph.D. thesis, Appalachian State University Boone, NC, USA (2013).
- [30] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, O. Chae, A dynamic histogram equalization for image contrast enhancement, *IEEE transactions on consumer electronics* 53 (2) (2007) 593–600.
- [31] D. T. Susetianingtias, S. Madenda, D. A. Rahayu, et al., Retinal microaneurysm detection using maximally stable external region algorithm, *International Journal on Advanced Science, Engineering and Information Technology* 6 (5) (2016) 644–648.
- [32] I. The MathWorks, Contrast limited adaptive histogram equalization,  
<https://www.mathworks.com/help/visionhdl/ug/contrast-adaptive-histogram-equalization.html>, accessed: 2024-12-23 (2024).  
URL <https://www.mathworks.com/help/visionhdl/ug/contrast-adaptive-histogram-equalization.html>
- [33] K. Zuiderveld, Contrast limited adaptive histogram equalization, Academic Press Professional, Inc., USA, 1994, Ch. 45, p. 474–485.
- [34] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on image processing* 15 (11) (2006) 3440–3451.
- [35] S. Athar, Z. Wang, A comprehensive performance evaluation of image quality assessment algorithms, *Ieee Access* 7 (2019) 140030–140070.
- [36] R. B. Lyngdoh, A. S. Sahadevan, T. Ahmad, P. S. Rathore, M. Mishra, P. K. Gupta, A. Misra, Avhyas: A free and open source qgis plugin for advanced hyperspectral image analysis, in: 2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), IEEE, 2021, pp. 71–76.
- [37] Q. Du, H. Yang, Similarity-based unsupervised band selection for hyperspectral image analysis, *IEEE geoscience and remote sensing letters* 5 (4) (2008) 564–568.
- [38] I. Biederman, Recognition-by-components: a theory of human image understanding., *Psychological review* 94 (2) (1987) 115.

- [39] M. Younis, Investigation of Hyperspectral Data Unmixing and Its Impact on the Task of Anomaly Detection, The University of Memphis, 2023.
- [40] A. Kostovska, C. Doerr, S. Džeroski, D. Kocev, P. Panov, T. Eftimov, Explainable model-specific algorithm selection for multi-label classification, in: 2022 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2022, pp. 39–46.
- [41] I. S. Reed, X. Yu, Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution, *IEEE Transactions on acoustics, speech, and signal processing* 38 (10) (1990) 1760–1770.
- [42] C.-I. Chang, S.-S. Chiang, Anomaly detection and classification for hyperspectral imagery, *IEEE Transactions on geoscience and remote sensing* 40 (6) (2002) 1314–1325.
- [43] P. C. Mahalanobis, On the generalized distance in statistics, *Sankhyā: The Indian Journal of Statistics, Series A* (2008-) 80 (2018) S1–S7.
- [44] H. Kwon, N. M. Nasrabadi, Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing* 43 (2) (2005) 388–397.
- [45] N. Acito, M. Diani, G. Corsini, Gaussian mixture model based approach to anomaly detection in multi/hyperspectral images, in: *Image and Signal Processing for Remote Sensing XI*, Vol. 5982, SPIE, 2005, pp. 209–217.
- [46] A. Schaum, Joint subspace detection of hyperspectral targets, in: 2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720), Vol. 3, IEEE, 2004, p. 1824 Vol.3.
- [47] M. J. Carlotto, A cluster-based approach for detecting man-made objects and changes in imagery, *IEEE Transactions on Geoscience and Remote Sensing* 43 (2) (2005) 374–387.
- [48] P. C. Hytla, R. C. Hardie, M. T. Eismann, J. Meola, Anomaly detection in hyperspectral imagery: comparison of methods using diurnal and seasonal data, *Journal of Applied Remote Sensing* 3 (1) (2009) 033546.
- [49] X. Kang, X. Zhang, S. Li, K. Li, J. Li, J. A. Benediktsson, Hyperspectral anomaly detection with attribute and edge-preserving filters, *IEEE Transactions on Geoscience and Remote Sensing* 55 (10) (2017) 5600–5611.
- [50] B. Du, R. Zhao, L. Zhang, L. Zhang, A spectral-spatial based local summation anomaly detection method for hyperspectral images, *Signal Processing* 124 (2016) 115–131.
- [51] K. Tan, Z. Hou, F. Wu, Q. Du, Y. Chen, Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation, *Remote sensing* 11 (11) (2019) 1318.
- [52] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 eighth ieee international conference on data mining, IEEE, 2008, pp. 413–422.

- [53] S. Li, K. Zhang, P. Duan, X. Kang, Hyperspectral anomaly detection with kernel isolation forest, *IEEE Transactions on Geoscience and Remote Sensing* 58 (1) (2019) 319–329.
- [54] Y. Lu, X. Zheng, H. Xin, H. Tang, R. Wang, F. Nie, Ensemble and random collaborative representation-based anomaly detector for hyperspectral imagery, *Signal Processing* 204 (2023) 108835.
- [55] N. Merrill, C. C. Olson, Unsupervised ensemble-kernel principal component analysis for hyperspectral anomaly detection, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 112–113.
- [56] X. Yang, X. Huang, M. Zhu, S. Xu, Y. Liu, Ensemble and random rx with multiple features anomaly detector for hyperspectral image, *IEEE Geoscience and Remote Sensing Letters* 19 (2022) 1–5.
- [57] S. Wang, W. Feng, Y. Quan, W. Bao, G. Dauphin, L. Gao, X. Zhong, M. Xing, Subfeature ensemble-based hyperspectral anomaly detection algorithm, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022) 5943–5952.
- [58] S. Fatemifar, M. Awais, A. Akbari, J. Kittler, A stacking ensemble for anomaly based client-specific face spoofing detection, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 1371–1375.
- [59] J. Nalepa, M. Myller, L. Tulczyjew, M. Kawulok, Deep ensembles for hyperspectral image data classification and unmixing, *Remote Sensing* 13 (20) (2021) 4133.
- [60] X. Zhang, H. Zhao, Hyperspectral-cube-based mobile face recognition: A comprehensive review, *Information Fusion* 74 (2021) 132–150.
- [61] G. Seni, J. Elder, *Ensemble methods in data mining: improving accuracy through combining predictions*, Morgan & Claypool Publishers, 2010.
- [62] C. C. Aggarwal, S. Sathe, Theoretical foundations and algorithms for outlier ensembles, *Acm sigkdd explorations newsletter* 17 (1) (2015) 24–47.
- [63] N. Keshava, J. F. Mustard, Spectral unmixing, *IEEE signal processing magazine* 19 (1) (2002) 44–57.
- [64] D. Hong, N. Yokoya, J. Chanussot, X. X. Zhu, An augmented linear mixing model to address spectral variability for hyperspectral unmixing, *IEEE Transactions on Image Processing* 28 (4) (2018) 1923–1938.
- [65] C.-I. Chang, Q. Du, Estimation of number of spectrally distinct signal sources in hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing* 42 (3) (2004) 608–619. doi:10.1109/TGRS.2003.819189.
- [66] M. E. Winter, N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, in: M. R. Descour, S. S. Shen (Eds.), *Imaging*

Spectrometry V, Vol. 3753, International Society for Optics and Photonics, SPIE, 1999, pp. 266 – 275. doi:10.1117/12.366289.  
URL <https://doi.org/10.1117/12.366289>

- [67] C.-I. Chang, An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis, *IEEE Transactions on Information Theory* 46 (5) (2000) 1927–1932. doi:10.1109/18.857802.
- [68] P. de Tchébychef, Average values, *Journal of Pure and Applied Mathematics* 12 (1867) 177–184.
- [69] B. Pavlyshenko, Using stacking approaches for machine learning models, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), IEEE, 2018, pp. 255–258.
- [70] L. Breiman, Stacked regressions, *Machine learning* 24 (1996) 49–64.
- [71] J. Sill, G. Takács, L. Mackey, D. Lin, Feature-weighted linear stacking, arXiv preprint arXiv:0911.0460.
- [72] D. R. Cox, The regression analysis of binary sequences, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 20 (2) (1958) 215–232.
- [73] J. R. Quinlan, *C4. 5: programs for machine learning*, Elsevier, 2014.
- [74] L. Breiman, Random forests, *Machine learning* 45 (1) (2001) 5–32. doi:10.1023/A:1010933404324.
- [75] T. Fushiki, Estimation of prediction error by using k-fold cross-validation, *Statistics and Computing* 21 (2011) 137–146.
- [76] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, S. Ridella, et al., The ‘k’ in k-fold cross validation., in: ESANN, 2012, pp. 441–446.
- [77] T.-T. Wong, P.-Y. Yeh, Reliable accuracy estimates from k-fold cross validation, *IEEE Transactions on Knowledge and Data Engineering* 32 (8) (2019) 1586–1594.
- [78] X. Kang, Airport-Beach-Urban (ABU) Datasets, <http://xudongkang.weebly.com/data-sets.html>, , accessed: 2023-04-23 (2017).
- [79] Y.-Q. Zhao, J. Yang, Hyperspectral image denoising via sparse representation and low-rank constraint, *IEEE Transactions on Geoscience and Remote Sensing* 53 (1) (2014) 296–308.
- [80] L. Zhu, G. Wen, Hyperspectral anomaly detection via background estimation and adaptive weighted sparse representation, *Remote Sensing* 10 (2) (2018) 272.

- [81] B. A. M Graña, MA Veganzons, Hyperspectral remote sensing scenes, University of the Basque Country, , accessed: 2024-01-13 (2001).  
URL [https://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)
- [82] W. Li, G. Wu, Q. Du, Transferred deep learning for anomaly detection in hyperspectral imagery, Git Hub, , accessed: 2024-01-13 (2019).  
URL [https://github.com/yousefan/CNNB/blob/master/HYDICE\\_urban.mat](https://github.com/yousefan/CNNB/blob/master/HYDICE_urban.mat)
- [83] N. J. P. Laboratory, Aviris - airborne visible / infrared imaging spectrometer, <https://aviris.jpl.nasa.gov/>, , accessed: 2024-01-23 (2006).  
URL <https://aviris.jpl.nasa.gov/>
- [84] X. Liu, M. Tanaka, M. Okutomi, Single-image noise level estimation for blind denoising, IEEE transactions on image processing 22 (12) (2013) 5226–5237.
- [85] D. B. Delbosque, Hypercube, spectral imagery analysis software,  
<https://www.erdc.usace.army.mil/media/fact-sheets/fact-sheet-article-view/article/610433/hypercube/>, , accessed: 2024-04-22 (July 2015).  
URL <https://www.erdc.usace.army.mil/media/fact-sheets/fact-sheet-article-view/article/610433/hypercube/>
- [86] R. Basedow, E. Zalewski, Characteristics of the hydice sensor, in: Summaries of the Fifth Annual JPL Airborne Earth Science Workshop. Volume 1: AVIRIS Workshop, Vol. 1, 1995, p. 9.
- [87] Resonon, Pika l - hyperspectral sensors - lightweight, compact vnir - benchtop systems, <https://resonon.com/Pika-L>, , accessed: 2023-09-13 (2022).  
URL <https://resonon.com/Pika-L>
- [88] J. A. Hanley, B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve., Radiology 143 (1) (1982) 29–36.
- [89] C.-I. Chang, Comprehensive analysis of receiver operating characteristic (roc) curves for hyperspectral anomaly detection, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–24. doi:10.1109/TGRS.2022.3211786.
- [90] r. rasharmaravindrasharma, Auc-roc curve, <https://www.geeksforgeeks.org/auc-roc-curve/>, , accessed: 2024-01-13 (January 2022).  
URL <https://www.geeksforgeeks.org/auc-roc-curve/>
- [91] N. Chinchor, Muc-4 evaluation metrics, in: Proceedings of the 4th Conference on Message Understanding, MUC4 '92, Association for Computational Linguistics, USA, 1992, p. 22–29. doi:10.3115/1072064.1072067.  
URL <https://doi.org/10.3115/1072064.1072067>

- [92] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets, *PLoS one* 10 (3) (2015) e0118432.
- [93] F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics Bulletin* 1 (6) (1945) 80–83.  
URL <http://www.jstor.org/stable/3001968>
- [94] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *Journal of the american statistical association* 32 (200) (1937) 675–701.
- [95] E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics* (1988) 837–845.
- [96] W. J. Krzanowski, D. J. Hand, ROC curves for continuous data, Chapman and Hall/CRC, 2009.
- [97] S. (<https://stats.stackexchange.com/users/22311/sycorax>), Statistical significance (p-value) for comparing two classifiers with respect to (mean) roc auc, sensitivity and specificity, Cross Validated, , uRL:<https://stats.stackexchange.com/q/358598> (version: 2018-08-01). arXiv:<https://stats.stackexchange.com/q/358598>.  
URL <https://stats.stackexchange.com/q/358598>
- [98] U. Muhammad, J. Laaksonen, D. Romaissa Beddiar, M. Oussalah, Domain generalization via ensemble stacking for face presentation attack detection, *International Journal of Computer Vision* 132 (12) (2024) 5759–5782.
- [99] Q. Wang, P. Bu, T. P. Breckon, Unifying unsupervised domain adaptation and zero-shot visual recognition, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
- [100] L. Meegahapola, H. Hassoune, D. Gatica-Perez, M3bat: Unsupervised domain adaptation for multimodal mobile sensing with multi-branch adversarial training, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8 (2) (2024) 1–30.
- [101] A. Pillai, S. K. Nepal, W. Wang, M. Nemesure, M. Heinz, G. Price, D. Lekkas, A. C. Collins, T. Griffin, B. Buck, S. M. Preum, T. Cohen, N. C. Jacobson, D. Ben-Zeev, A. Campbell, Investigating generalizability of speech-based suicidal ideation detection using mobile phones, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7 (4). doi:10.1145/3631452.  
URL <https://doi.org/10.1145/3631452>
- [102] M. Mancini, S. R. Bulo, B. Caputo, E. Ricci, Best sources forward: domain generalization through source-specific nets, in: 2018 25th IEEE international conference on image processing (ICIP), IEEE, 2018, pp. 1353–1357.

- [103] J. Ryu, G. Kwon, M.-H. Yang, J. Lim, Generalized convolutional forest networks for domain generalization and visual recognition, in: International conference on learning representations, 2019, p. 0.
- [104] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, S. Y. Philip, Generalizing to unseen domains: A survey on domain generalization, *IEEE transactions on knowledge and data engineering* 35 (8) (2022) 8052–8072.
- [105] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.
- [106] H. Wu, J. Zhang, K. Huang, K. Liang, Y. Yu, Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation, *arXiv preprint arXiv:1903.11816*.
- [107] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.
- [108] R. Strudel, R. Garcia, I. Laptev, C. Schmid, Segmenter: Transformer for semantic segmentation, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 7262–7272.
- [109] R. Ranftl, A. Bochkovskiy, V. Koltun, Vision transformers for dense prediction, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 12179–12188.
- [110] EOS, Crop map, EOS Data Analytics, Inc., , accessed: Nov 01, 2024 (2022).  
URL <https://eos.com/industries/agriculture/crop-map/>
- [111] I. Becker-Reshef, B. Barker, A. Whitcraft, P. Oliva, K. Mobley, C. Justice, R. Sahajpal, Crop type maps for operational global agricultural monitoring, *Scientific Data* 10 (1) (2023) 172.
- [112] M. Alami Machichi, I. E. mansouri, Y. Imani, O. Bourja, O. Lahlou, Y. Zennayi, F. Bourzeix, I. Hanadé Houmma, R. Hadria, Crop mapping using supervised machine learning and deep learning: a systematic literature review, *International Journal of Remote Sensing* 44 (8) (2023) 2717–2753.
- [113] L. Wan, H. Cen, J. Zhu, J. Zhang, Y. Zhu, D. Sun, X. Du, L. Zhai, H. Weng, Y. Li, et al., Grain yield prediction of rice using multi-temporal uav-based rgb and multispectral images and model transfer—a case study of small farmlands in the south of china, *Agricultural and Forest Meteorology* 291 (2020) 108096.
- [114] B. Niu, Q. Feng, B. Chen, C. Ou, Y. Liu, J. Yang, Hsi-transunet: A transformer based semantic segmentation model for crop mapping from uav hyperspectral imagery, *Computers and Electronics in Agriculture* 201 (2022) 107297.

- [115] R. Li, S. Zheng, C. Duan, Y. Yang, X. Wang, Classification of hyperspectral image based on double-branch dual-attention mechanism network, *Remote Sensing* 12 (3) (2020) 582.
- [116] Z. Luo, W. Yang, Y. Yuan, R. Gou, X. Li, Semantic segmentation of agricultural images: A survey, *Information Processing in Agriculture* 11 (2) (2024) 172–186. doi:<https://doi.org/10.1016/j.inpa.2023.02.001>.
- [117] S. Tian, Y. Dong, R. Feng, D. Liang, L. Wang, Mapping mountain glaciers using an improved u-net model with cse, *International Journal of Digital Earth* 15 (1) (2022) 463–477.
- [118] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306.
- [119] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999.
- [120] Z. Fan, K. Liu, J. Hou, F. Yan, Q. Zang, Jaunet: A u-shape network with jump attention for semantic segmentation of road scenes, *Applied Sciences* 13 (3) (2023) 1493.
- [121] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, *Medical image analysis* 53 (2019) 197–207.
- [122] M. Amiri, R. Brooks, H. Rivaz, Fine tuning u-net for ultrasound image segmentation: which layers?, in: *MICCAI Workshop on Domain Adaptation and Representation Transfer*, Springer, 2019, pp. 235–242.
- [123] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, Eca-net: Efficient channel attention for deep convolutional neural networks, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11534–11542.
- [124] X. Li, H. Xiong, H. Wang, Y. Rao, L. Liu, Z. Chen, J. Huan, Delta: Deep learning transfer using feature map with attention for convolutional networks, arXiv preprint arXiv:1901.09229.
- [125] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi:[10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- [126] S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Transactions on knowledge and data engineering* 22 (10) (2009) 1345–1359.
- [127] M. F. Baumgardner, L. L. Biehl, D. A. Landgrebe, 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, *Purdue University Research Repository* 10 (7) (2015) 991, , accessed: Nov 01, 2024.

**URL** <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>