Hyperspectral Unmixing-Based Anomaly Detection

Mohammed Younis[1], Mazharul Hossain[2], Aaron Robinson[1],
Lan Wang[2], and Chrysanthe Preza[1]

[1]Electrical and Computer Engineering Department
[2]Computer Science Department
The University of Memphis, Memphis, TN, USA

ABSTRACT

Research supporting improved anomaly detection performance benefits a wide range of technical applications. Thus, the definition of anomalies and the subsequent means to detect them are wide-ranging. This treatment presents an overview of the development of an anomaly detection approach based on spectral signatures obtained with hyperspectral unmixing. Anomaly Detection is a binary classification that does not require prior information about the anomaly. For example, an anomaly detector applied to hyperspectral imaging (HSI) would take a hyperspectral image with hundreds of channels as an input and output a two-dimensional image map of pixel intensities based on a threshold procedure applied to the probability of that pixel being an anomaly. There have been many advancements in the field of HSI Anomaly Detection. Our ensemble method algorithm, presented here, addresses some of the shortcomings of current state-of-the-art techniques. We present details about the extracted end-members and use them for effective anomaly detection. Our current ensemble method opens the path for future machine-learning processes. We evaluated our method on multiple datasets and reported the F1-macro score. We suggest that the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve should not be used in Hyperspectral anomaly detection as an evaluation metric.

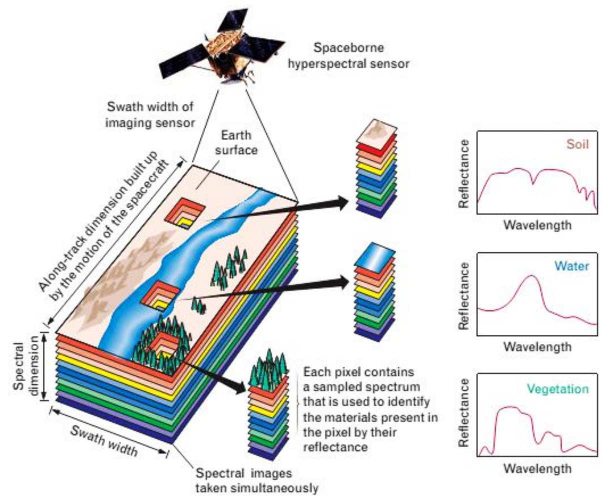
Keywords: Anomaly detection, spectral unmixing, endmembers, machine learning

1. INTRODUCTION

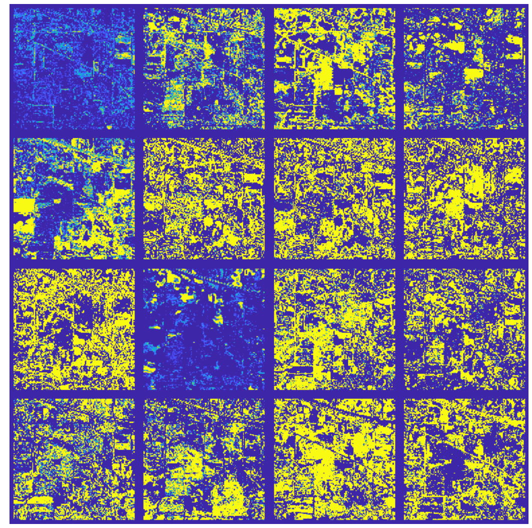
The electromagnetic spectrum of visible light is 380 nm to 700 nm, and the typical computer image uses RGB channels from this visible light spectrum. RGB images consist of Red, Green, and Blue channels. It is sometimes hard to distinguish similarly colored objects using eyes and computer vision in RGB images. Hyperspectral cameras can aid difficult computer vision tasks such as object discrimination because they can capture information from a wide range of bands extending from ultraviolet (350 nm) to infrared (2500 nm). A Hyperspectral Image contains data from the electromagnetic spectrum, hundreds of channels, and additional information per pixel, increasing the possibility for each object to possess a unique measured signature. We can use this unique signature to identify each object's material, making it possible to recognize any anomalous object in the image.[1] Anomaly Detection has immense potential in agriculture, metallurgy, surveillance, and versatile applications like quality control, process control, object sorting, and remote sensing. Materials reflect, absorb, and transmit the electromagnetic radiation emitted by the sun in unique ways. The measure of electromagnetic energy reflected or bounced back from a material is called the reflectance of the material.[1] Figure 1a demonstrates the difference in reflectance between different materials.

It is essential to talk about the endmembers in the field of hyperspectral imagery. Endmembers are the unique spectra that represent the material in a hyperspectral image. A pixel in a hyperspectral image is termed a pure pixel if it only contains/represents the spectral signature of one material or endmember. Conversely, the pixel's spectral signature can result from a linear or non-linear combination of more than one endmember, and in this case, it is called a mixed pixel.[2] Spectral unmixing is a process that decomposes the measured spectra into a collection of endmembers while indicating the proportion of each endmember in the pixel.[3] The spectral unmixing process starts with identifying/extracting the number of representative endmembers in the hyperspectral image; then an endmember estimating algorithm is used to identify those unique endmembers.

Further author information: Send correspondence to Chrysanthe Preza; E-mail: cpreza@memphis.edu



(a) HS image reflectance of soil, vegetation and water[1]



(b) Abundance map for Indian Pines

Figure 1: HSI used in remote sensing can result in information about the reflectance properties of materials and the abundance map, after spectral unmixing.

Once all endmembers are found, an algorithm to obtain the abundance map can be utilized.[3] The result of spectral unmixing is often called an abundance map, as shown in Figure 1b. For example, the Indian Pines dataset contains Hyperspectral images represented by 16 unique endmembers. This resulted in the abundance map shown in Figure 1b where the intensity of each pixel in the 16 individual squares indicates the presence of each endmember.

In this treatment, anomaly detection is defined as a classification problem with two classes that do not require previous knowledge about the anomalies and aim to find the abnormalities in the hyperspectral image.[4] An anomaly detection algorithm subsequently classifies each pixel in the HS image into an anomaly class or a background class. The issue with some anomaly detection methods is that they make Gaussian assumptions about the hyperspectral data when the algorithm is developed. Still, hyperspectral images captured in the natural world show strong non-linearity and non-gaussianity.[5] A Gaussian assumption means that anomalies tend to be the small and less frequently occurring objects in the image.[4] If we have an image of the sea, and only one ship in the middle of the water, then the Gaussian assumption will work, and the ship would be considered the anomaly. However, real-life scenarios are much more complicated than this, which is why many anomaly detectors fail at identifying anomalies.

This paper mentions a few state-of-the-art (SOTA) Anomaly Detectors (AD), and these methods do not always perform well. An ensemble model will be presented that combines some of these methods and performs better than each SOTA ADs. We considered the performance of each of these SOTA ADs and chose the best three based on their average F1 score on all airports of Airport–Beach–Urban (ABU) Datasets.[6]

The main contributions of this paper can be divided into three parts:

1. The design of a robust ensemble method that outperforms SOTA Anomaly detectors.
2. Integrating spectral unmixing and target detection as part of the ensemble anomaly detector.
3. Results demonstrating why Area Under Curve(AUC) is not the optimal evaluation metric for Anomaly Detection tasks.

2. RELATED WORKS

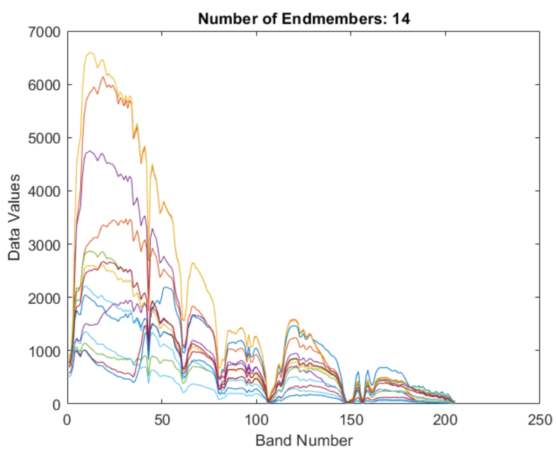
2.1 Hyperspectral Target Detection With Endmembers and Abundances

Higher-resolution multispectral images have successfully made whole-pixel classification possible. However, additional analysis of substances in the pixel was limited due to the limited number of spectral bands in multispectral images.[3] In addition, pixels may be a combination of multiple materials has created the need to unmix those materials.[3]

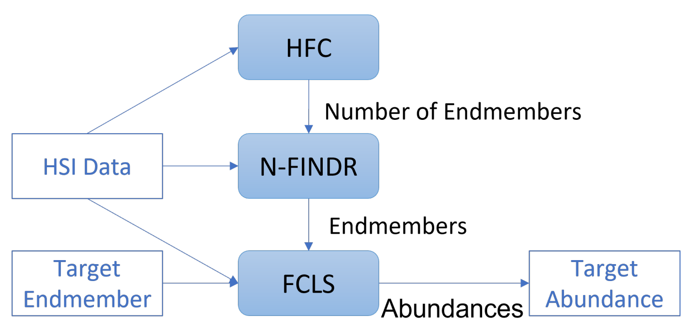
As mentioned above, hyperspectral (HS) images have significantly more bands than their multispectral counterparts. However, images in HSI also suffer from lower spatial resolution. As s result, an HS image pixel can be a pure pixel consisting of spectral characteristics of a single material, or a mixed pixel representing a mixture of two or more distinct materials has a higher probability of occurrence. This mixture could be due to the spectral sensor's limitation, resulting in a low spatial resolution, and mixed pixels can result from a homogeneous mixture of materials.[3]

Spectral unmixing decomposes each pixel into unique spectra called endmembers and their estimated contribution or abundances.[3] 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.[7]

Once the number of endmembers in the hyperspectral image is known, the N-FINDR algorithm can estimate those individual spectra .[8] Figure 2a displays the endmember estimated using the N-FINDER algorithm[8] on ABU-Airport-I hyperspectral image. In ABU-Airport-I, there are 14 endmembers, and thus there are 14 unique spectra in the endmembers graph.



(a) Endmembers collected from ABU-Airport-I



(b) Our method to get target abundance

Figure 2: Our complete method to get target abundance using target endmember. Our 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)[3] method could be applied to estimate the abundance map of the unique spectra in the hyperspectral image. This method unmixes an HS image into the fractional abundance of each material in each pixel. The area of this abundance map is the same as the input HS image and has the same number of abundance maps as the number of estimated end-members in the HS image. Each endmember contributes to the HS image structure in a corresponding and dedicated smaller square. An abundance map is shown in Figure 1b.

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.[9] 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. FCLS[3] assigns a probability distribution to each class as an abundance map. We have shown our process in Figure 2b. The class with the highest probability is normalized to 1 and classified as that pure material spectrum. We use this normalized target abundance for our work.

2.2 Hyperspectral Anomaly Detection

This section discusses the current state-of-the-art hyperspectral anomaly detection algorithms.

2.2.1 Reed-Xiaoli (RX) Anomaly Detector

The RX algorithm utilizes the Mahalanobis distance[10] to calculate the distance between the pixel under test and the background pixel. A pixel is considered an anomaly if the Mahalanobis distance exceeds a certain threshold. The background of the hyperspectral image is characterized by the covariance and mean of the hypercube.[11,12] Equation 1 below shows the square of the Mahalanobis distance:

$$D^2 = (\mathbf{x} - \mathbf{m})^T . C^{-1} . (\mathbf{x} - \mathbf{m}), \tag{1}$$

where, $\mathbf{x}$ is the vector of the observation, $\mathbf{m}$ is the vector of mean values of independent variables and C^{-1} is the inverse covariance matrix of the independent variables.

RX algorithm became popular due to its ability to detect anomalies without requiring labeled data. Since labeled HS images may be considered scarce, RX was widely adopted to overcome the shortage of labeled data in the hyperspectral field. There are multiple versions of RX have been introduced over time. Küçük, S., and Yüksel, S. E.[13] evaluated multiple versions of the RX algorithm, which we used in our consideration in selecting methods.

2.2.2 Hyperspectral Anomaly Detection With Attribute and Edge-Preserving Filters (AED)

Anomaly Detection With Attribute and Edge-Preserving Filters (AED) is a two-stage algorithm for anomaly detection in hyperspectral images. AED is based on two main ideas; the first states that anomalies usually appear small and have distinct reflectance signatures,[6] and the second idea states that pixels belonging to the same class tend to have a high correlation in the space domain.[6]

In AED, those pixels with unique signatures and small areas are first detected using an attribute filtering and boolean map-based fusion that generates the initial detection output.[6] Then this initial output is modified by using edge-preserving filters that utilize the spatial correlation between adjacent pixels; this leads to the reduction in false positive anomalies.[6] The structure of AED is shown in Figure 3 below.

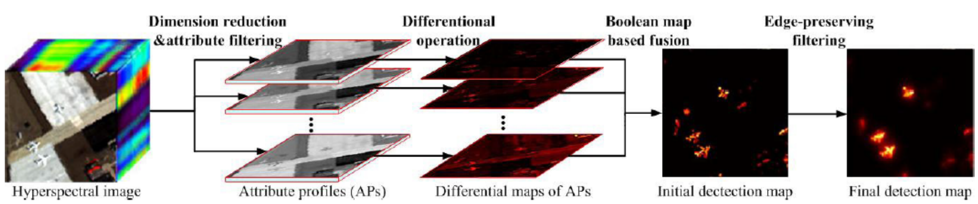


Figure 3: Schematic of AED Anomaly Detection[6]

2.2.3 Hyperspectral Anomaly Detection With Kernel Isolation Forest

Kernel Iforest is mainly based on the idea that anomalies are more prone to isolation in kernel space.[14] This algorithm works by first mapping the data into kernel space. Then a principal component analysis is applied to the data and the first few components are chosen. Then isolated samples are constructed with the use of IForest algorithm. The initial detection output is iteratively improved by using locally constructed IForest.[14] The general idea of Kernel Iforest is that if a pixel can be easily isolated in the kernel space, it is most likely to be an anomaly.[14] One of the advantages of using IForest is the fact that it is a good algorithm to detect outliers or anomalies based on the separability of the datapoint without the need for complex computations compared to other algorithms that rely on distance or density.[14]

Kernel IForest implementations operate upon the assumption that anomalies are points or objects that do not appear frequently and are distinguishable from other points or objects.[14] If a dataset contains normal and anomaly instances, the goal is to separate or isolate them by constructing a tree that splits the data using different attributes. The distance between the main node and the last leaf node is the depth of the split. Since this algorithm assumes that anomalies tend to be different than normal instances, the depth of the split for anomaly tends to be shorter than normal instances.[14]

2.2.4 Local Summation Unsupervised Nearest Regularized Subspace With An Outlier Removal Anomaly Detector (LSUNRSORAD)

The Local Summation Anomaly Detection (LSAD) algorithm has great performance. Still, because it obtains local spatial distributions for the neighboring pixels of the pixel under test (PUT) by using multiple sliding window filters, it becomes time-consuming and computationally expensive.[15] In addition, the background statistics could be contaminated with anomalies when the algorithm uses a single window, causing a high false alarm rate.[15]

In[15] the authors present a modified Local Summation Anomaly Detection called LSUNRSORAD.[15] The main idea of this algorithm is that each pixel in the background image can be represented by its neighbors.[15] In the original LSAD method, a correlation matrix represents the correlation between background pixels. During the calculation process, the matrix is inverted, causing the algorithm to be computationally expensive. The improved algorithm replaces matrix inversion with a linear combination using addition and multiplication.

LSUNRSORAD uses outlier removal as they affect the accuracy of the algorithm. Outliers are detected by setting two thresholds; the first being the maximum threshold and the second being the minimum threshold. A pixel with a value greater than the maximum threshold or smaller than the minimum threshold would be considered an outlier.[15]

3. METHODOLOGY

This paper presents an ensemble learning method that combines multiple anomaly detectors and the abundances obtained from hyperspectral unmixing to find and classify the target. The ultimate goal behind this hyperspectral unmixing-based ensemble method is to design an anomaly detector that is more accurate and robust when compared to the algorithms used individually.

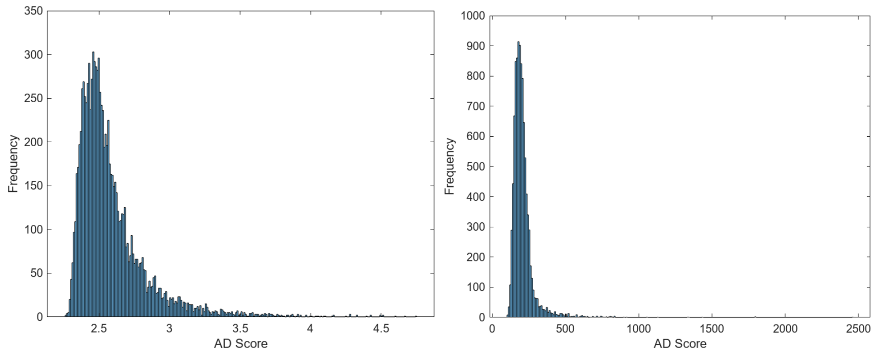
This method uses each algorithm individually to detect anomalies in the hyperspectral image. Then the output of each anomaly detector is standardized if it is not already a binary image. Finally, applying the ensemble method combines those results and produces the anomaly detection map.

3.1 Standardize the Anomaly Detector results with Binary

As we have to combine multiple Anomaly Detectors, it is important to standardize the results to reduce bias. Figure 4 shows the algorithms produce results in different scales and frequencies. The Kernel Isolation Forest (Figure 4a) has a score range of 0 to 5. The RX algorithm (Figure 4b) produces scores between 0 and 2500. Feeding these unmodified results into the ensemble-based system would reduce system accuracy; therefore, a solution to mitigate the impact of the scale mismatch must be identified.

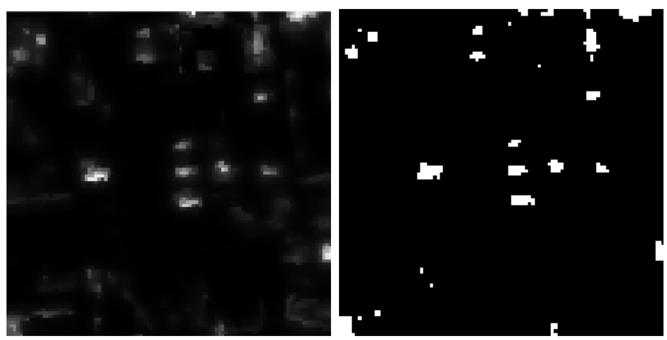
Figure 5 shows a comparison between the raw AED output shown in Figure 5a and the binarized (its values are either ones or zeros) AED output shown in Figure 5b. After observing both figures, it becomes obvious that Figure 5a is not user-friendly to visualize and understand the scene, whereas Figure 5b is easier to interpret and understand.

As most of the algorithms used do not output a binary image, there was a need to convert the algorithms' output into a binary image. Initial efforts focused on histograms of the algorithm output, with results indicating that the values produced followed a skewed normal distribution. We selected thresholds based on percentile values to account for the skewed nature of the data. Positive thresholds were then tested and applied to each pixel with the interpretation that pixel values equal to or higher than the threshold would be considered an anomaly; conversely, if the pixel intensity value were less than the threshold, it would be considered normal. Table 4 compares the results of two possible thresholds chosen due to their performance on the skewed data. By



(a) Anomalies using Kernel Isolation Forest Detector (b) Anomalies using Reed-Xiaoli Detector

Figure 4: Comparison between output range of Isolation Forest Detector and RX Detector using ABU-Airport-I



(a) Raw AUC = 0.9923 (b) Binarized AUC = 0.9128

Figure 5: Comparison between original AED output and Binarized AED output using ABU-Airport-I

taking the average F1 score for all of the algorithms and comparing the results of the images binarized using the 95$^{\text{th}}$ and 97$^{\text{th}}$ percentiles, the 97$^{\text{th}}$ percentile resulted in a higher average F1 score. Thus it was chosen as the threshold. The output of this binarizing algorithm is called an image map. This image map is a binary image with the same dimension as the hyperspectral image, and its values can only be 0 or 1. A zero value represents the background or the nonanomalous class, whereas a one represents an anomaly. Once the output of all algorithms has been converted into a binary image, we can use them in our ensemble method.

3.2 Hyperspectral Unmixing-based Voting Ensemble Anomaly Detector (HUE-AD)

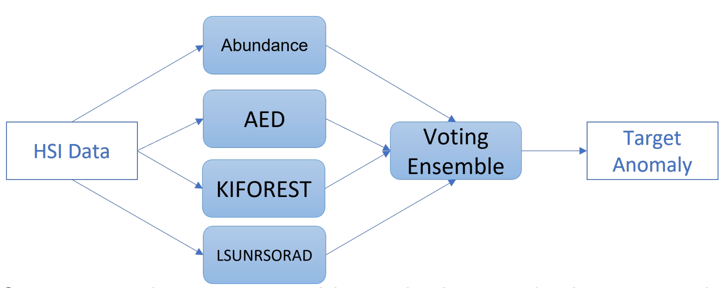


Figure 6: Our proposed voting ensemble method anomaly detector with abundance

The ensemble method combines the votes of every Anomaly Detector to output the final result image. Our ensemble method utilizes equal weights for each anomaly detector and the abundance-based procedure. The voting is designed to declare a pixel to be anomalous in case: if at least three of the four methods indicate the pixel is anomalous.

Like the image binarization procedure, the ensemble method utilizes a predefined threshold; if the voting result is higher than or equal to the threshold, then the pixel under test (PUT) is declared an anomaly. This approach to ensemble method design ensures that the anomaly detector outperforms individual anomaly detectors and produces a lower false alarm rate.

The voting process can be summarized using Equations (2) and (3)

$$totalPixelvote = IM_{AED} + IM_{KIforest} + IM_{LSUNRSORAD} + IM_{abundances} \tag{2}$$

where I is an integer value greater than zero and represents the weighted contribution of each algorithm output to the pixel vote. Equation (2) indicates that equal weight to applied to each method.
M is the algorithm output for the pixel and takes values that are either 0 or 1.
The first three methods were binarized using the percentile method. The abundance-based procedure output was normalized to 1 by taking the highest probability in the probability distribution of the abundance classes.

Then the thresholding process for each pixel can be summarized as follows:

$$\begin{aligned} if(totalPixelvote) &\geq (N-1) * I : pixel = 1 \\ elseif(totalPixelvote) &< (N-1) * I : pixel = 0 \end{aligned} \tag{3}$$

Where N is the number of methods used in the ensemble method.

4. EVALUATION

4.1 Datasets

For our evaluation, we have used Airport–Beach–Urban (ABU) Datasets.[6,16] The 100 $\times$ 100 patches were manually extracted from large images downloaded from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Web site.[17] Some features of these images are listed in Table 1.

Table 1: Some Features of the ABU Datasets Airport scenes[18]

Images	Captured place	Resolution	Sensor	Flight time	Band
ABU-Airport-I	Los Angels	7.1m	AVIRIS	11/09/2011	205
ABU-Airport-II	Los Angels	7.1m	AVIRIS	11/09/2011	205
ABU-Airport-III	Los Angels	7.1m	AVIRIS	11/09/2011	205
ABU-Airport-IV	Gulfport	3.4m	AVIRIS	07/07/2010	191

As shown in Table 1, most of the images are captured by the AVIRIS sensor except for ABU-Airport-IV, which is captured by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor. Moreover, due to the different heights of the flights, the spatial resolutions of the images are also different.

4.2 Looking Beyond AUC as an Excellent Evaluation Metric

Multiple metrics apply to hyperspectral anomaly detector evaluation; one of the most adopted ones is the Area Under Curve (AUC). However, after closely investigating the reliability of AUC to evaluate different anomaly detection algorithms, it was observed that a higher AUC score does not necessarily mean better performance.

The algorithmic results summarized in Figures 7 and 8 can be used to illustrate the issue. As described above, the binarized images indicate anomalous pixels with pixel intensity values of one. In both images the Ground Truth anomalous pixel positions are given in the right-most image. The other images contain the output of other algorithms with their respective AUCs indicated below. The figures demonstrate that having more false positives could mean a higher AUC score. This is readily apparent in Figure 8. Figure 8a has more false positives than Figure 8b, yet it has a much better AUC score. This demonstrates that consideration of AUC score alone is not recommended for Hyperspectral Anomaly Detection algorithmic evaluation.

These conclusions above can be supported by considering the characteristics of the expected imagery. Hyperspectral images have severe imbalances, with most of the pixels belonging to one class and a limited number

(a) AUC = 0.9350 (b) AUC = 0.9690 (c) AUC = 0.9696 (d) Ground Truth

Figure 7: A higher AUC score does not necessarily mean better performance. Subfigures 7b and 7c have more false positives compared with 7a, yet they have a higher ROC AUC score for the ABU-Airport-I image.

belonging to the other. Specifically, the HSI anomaly detector developed in this paper results in two classes; a minority class representing anomalies and a majority class representing non-anomalous objects. This is problematic because a small number of incorrect or correct predictions can lead to a significant change in the AUC score.[19] Figures 7 and 8 illustrate this and demonstrate that ROC AUC is not a reliable way to evaluate classification tasks when there is a class imbalance.

(a) AUC = 0.9484 (b) AUC = 0.8385 (c) Ground Truth

Figure 8: A higher AUC score does not necessarily mean better performance: 8a has more false positives than 8b, but the latter has a smaller ROC AUC score for the ABU-Airport-II image.

If we examine the Equations (4) and (5) used to create the ROC AUC shown below, we can also provide additional support the AUC issue.

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

$$FPR = \frac{FP}{TN + FP} \tag{5}$$

Note that in Equations (4) and (5) TPR is the True Positive Rate, FPR is the False Positive Rate, TP is the actual positives that were predicted as positives. TN represents actual negatives that were predicted as negatives. FP are actual negatives that were predicted as positives. FN are the actual positives that were predicted as negatives.

Examine equation 5, and note that FPR contains TN in the denominator. In most anomaly detection applications, TN is a very large number when compared to FP. The implication is that FPR would be a relatively small number and the corresponding AUC score would be high simply because AUC pays more attention to the majority class.

4.3 Using Classification metrics

Figure 5 shows a comparison between the raw AED output shown in Figure 5a and the binarized AED output shown in Figure 5b. It shows that binarized 5b is user-friendly and easier to visualize and understand the scene's implications. The image is cleaner and readily conveys the data indicating lower false positives did not result in a better AUC score. Therefore the need for a better evaluation method for Anomaly Detectors is clearly apparent.

Since hyperspectral anomaly detection is a binary classification task that does not need prior knowledge about what is defined to be an anomaly,[4] traditional classification evaluation metrics can be utilized to evaluate anomaly detectors as long as their output is a binary image map. Some of the algorithms we used do not output binary images. Thus we used 97^{th} percentile binarization method to convert the output of these algorithms into a binary image.

Once the output of all algorithms has been converted into a binary image, classical classification evaluation metrics could be applied to evaluate those methods. A good example of those metrics is the F1 score, which gives the overall performance by highlighting the trade-off between precision and recall.
The reason behind choosing precision and recall is that they make it possible to evaluate the performance of a classifier on minority classes.[20] This becomes obvious after looking at the equations for Precision and Recall shown in equations 6 and 7 :

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Where TP is the actual positives that were predicted as positives. FP are actual negatives that were predicted as positives. FN are actual positives that were predicted as negatives. This makes precision and recall better suitable for highly skewed data where ROC provides an extremely optimistic evaluation for the performance.[21]

The F1 score is a harmonic mean of precision and recall. It is a more balanced method to measure model performance.

$$F1\_score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{8}$$

Because of class imbalance, the general F1 score will not give us the perfect insight into algorithmic performance. With the need for different metrics understood, we used the macro-averaged F1 score (or F1-macro score). It is computed using the arithmetic mean (aka unweighted mean) of all the per-class F1 scores. Here, N is the number of classes, in our case, two.

$$F1\_macro = \frac{\sum_{n=1}^{N} F1\_score_n}{N} \tag{9}$$

4.4 Performance Comparison

Table 2: F1 Score comparison between individual algorithms and Proposed Ensemble method

Anomaly Detector	ABU-Airport-I	ABU-Airport-II	ABU-Airport-III	ABU-Airport-IV
Binarized AED	0.78	0.65	0.71	0.61
Binarized KIFOREST	0.68	0.68	0.70	0.63
Binarized LSUNRSORAD	0.69	0.61	0.72	0.57
Binarized Abundances	0.67	0.75	0.76	**0.74**
HUE-AD	**0.79**	**0.77**	**0.77**	0.67

Table 2 summarizes the performance of individual methods and our ensemble method. The data indicates the hyperspectral unmixing-based ensemble method performs better in the majority of the cases. We have

Table 3: F1 Score comparison between Other Algorithms

Anomaly Detector	ABU-Airport-I	ABU-Airport-II	ABU-Airport-III	ABU-Airport-IV
Binarized CSD[22]	0.59	0.59	0.65	0.59
Binarized FCBAD[23]	0.61	0.60	0.61	0.62
Binarized GMRX[24]	0.59	0.57	0.55	0.61
Binarized KRX[25]	0.56	0.66	0.67	0.60
Binarized RX	0.58	0.59	0.65	0.60

considered some other methods as the input for our ensemble methods. However, they did not perform as well as the methods used in the ensemble method developed in this treatment. Their results are summarized in Table 3 for completeness.

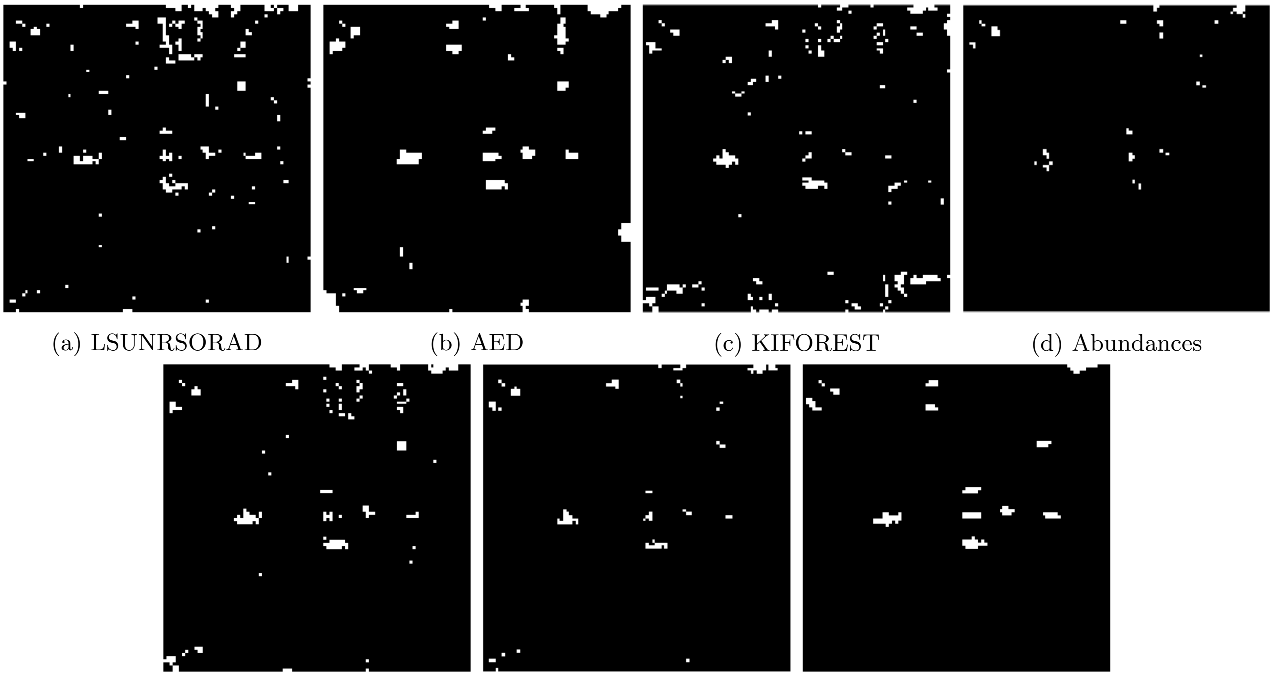
This paper mentions a few state-of-the-art (SOTA) anomaly detectors, and these methods do not always perform well. Therefore we developed and present an ensemble model that performs better than each of those SOTA anomaly detectors Individually. Those SOTA anomaly detectors were chosen based on the performance recorded in Tables 2 and 3. Table 3 contains a performance comparison between other Anomaly Detectors that were not chosen to be part of the ensemble method. By comparing the algorithms in Table 2 and Table 3, it is clear that the chosen algorithms outperform those in Table 3. After looking at Table 3, one might wonder why KRX was not chosen, eventhough it performes well. The reason is that KRX is computationally expensive and requires a large amount of memory depending on the size of the hyperspectral image. This has led to choosing LSUNRSORAD over it, as both have comparable F1 scores.

Table 4: F1 Score comparison between 95^{th} percentile and 97^{th} percentile

Anomaly Detector	ABU-Airport-I	ABU-Airport-II	ABU-Airport-III	ABU-Airport-IV
AED with 95^{th} percentile	0.78	0.65	0.71	0.63
AED with 97^{th} percentile	0.78	0.65	0.71	0.61
KIFOREST with 95^{th} percentile	0.65	0.65	0.68	0.61
KIFOREST with 97^{th} percentile	0.68	0.68	0.70	0.63
LSUNROD with 95^{th} percentile	0.65	0.58	0.69	0.56
LSUNROD with 97^{th} percentile	0.69	0.61	0.72	0.57
KRX with 95^{th} percentile	0.56	0.67	0.64	0.67
KRX with 97^{th} percentile	0.56	0.66	0.67	0.60

4.5 Visual Comparison

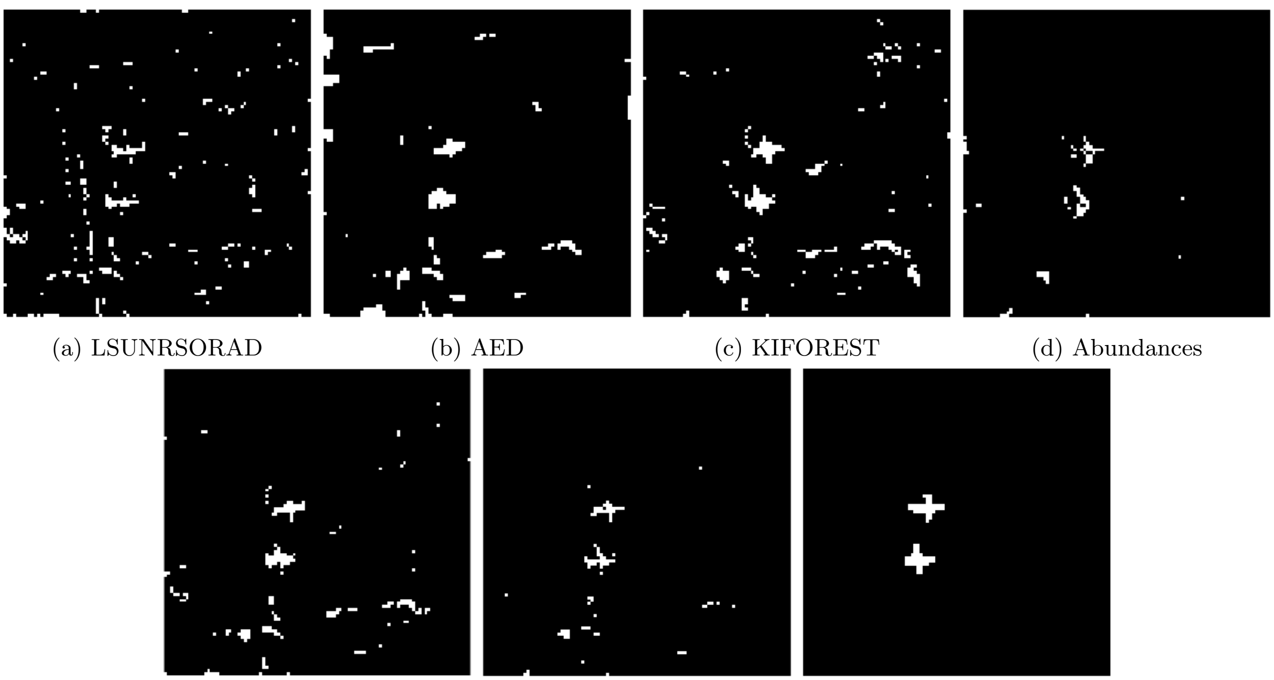
The images in Figures 9 and 10 highlight the visual differences. Figures 9 and 10 both demonstrate how the unmixing-based ensemble method outperforms each of the SOTA Anomaly Detectors individually.



(a) LSUNRSORAD (b) AED (c) KIFOREST (d) Abundances

(e) Ours w/out abundances (f) Ours with abundances (g) Ground truth

Figure 9: Comparison between different methods and our proposed ensemble method using ABU-Airport-I



(a) LSUNRSORAD (b) AED (c) KIFOREST (d) Abundances

(e) Ours w/out abundances (f) Ours with abundances (g) Ground truth

Figure 10: Comparison between different methods and our proposed ensemble method using ABU-Airport-II

5. CONCLUSION

This paper describes a hyperspectral unmixing-based anomaly detector approach that classifies anomalous pixels without prior knowledge. This ensemble anomaly detector method works with abundance (target spectra) and performs better than other state-of-the-art methods. The results demonstrate that including hyperspectral unmixing in the ensemble improves algorithmic performance, stabilizes, and produces robust results beyond individual implementations. Additionally, this paper presents results that do not support the use of AUC ROC as an evaluation method for anomaly detectors because, in the case of anomaly detection, there is an imbalance between the two classes, and AUC does not focus on the minority class, which in this case, is the anomaly class. This paper recommends using the F1-macro score, however, in order to use the F1 score, the output of the algorithms needs to have discrete or binary values. There are many well-performing algorithms for anomaly detection tasks, but many do not output a binary image map. This paper utilizes the 97^{th} percentile thresholding method to convert the output of anomaly detectors into a binary image and then evaluates the performance of all algorithms using the F1 score.

6. ACKNOWLEDGEMENTS

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-21-2-0294. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., and Abbas, A., "Modern trends in hyperspectral image analysis: A review," *Ieee Access* **6**, 14118–14129 (2018).

[2] Racetin, I. and Krtalić, A., "Systematic review of anomaly detection in hyperspectral remote sensing applications," *Applied Sciences* **11**(11), 4878 (2021).

[3] Keshava, N. and Mustard, J. F., "Spectral unmixing," *IEEE signal processing magazine* **19**(1), 44–57 (2002).

[4] Raza Shah, N., Maud, A. R. M., Bhatti, F. A., Ali, M. K., Khurshid, K., Maqsood, M., and Amin, M., "Hyperspectral anomaly detection: a performance comparison of existing techniques," *International Journal of Digital Earth* **15**(1), 2078–2125 (2022).

[5] Zhao, R., Shi, Z., Zou, Z., and Zhang, Z., "Ensemble-based cascaded constrained energy minimization for hyperspectral target detection," *Remote Sensing* **11**(11), 1310 (2019).

[6] Kang, X., Zhang, X., Li, S., Li, K., Li, J., and Benediktsson, J. A., "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Transactions on Geoscience and Remote Sensing* **55**(10), 5600–5611 (2017).

[7] Chang, C.-I. and Du, Q., "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **42**(3), 608–619 (2004).

[8] Winter, M. E., "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in [*Imaging Spectrometry V*], Descour, M. R. and Shen, S. S., eds., **3753**, 266 – 275, International Society for Optics and Photonics, SPIE (1999).

[9] Chang, C.-I., "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Transactions on Information Theory* **46**(5), 1927–1932 (2000).

[10] Chandra, M. P. et al., "On the generalised distance in statistics," in [*Proceedings of the National Institute of Sciences of India*], **2**(1), 49–55 (1936).

[11] Reed, I. S. and Yu, X., "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE transactions on acoustics, speech, and signal processing* **38**(10), 1760–1770 (1990).

[12] Chang, C.-I. and Chiang, S.-S., "Anomaly detection and classification for hyperspectral imagery," *IEEE transactions on geoscience and remote sensing* **40**(6), 1314–1325 (2002).

[13] Küçük, S. and Yüksel, S. E., "Comparison of rx-based anomaly detectors on synthetic and real hyperspectral data," in [*2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*], 1–4, IEEE (2015).

[14] Li, S., Zhang, K., Duan, P., and Kang, X., "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Transactions on Geoscience and Remote Sensing* **58**(1), 319–329 (2019).

[15] Tan, K., Hou, Z., Wu, F., Du, Q., and Chen, Y., "Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation," *Remote sensing* **11**(11), 1318 (2019).

[16] Kang, X., "Airport-Beach-Urban (ABU) Datasets," (2017).

[17] "AVIRIS - Airborne Visible / Infrared Imaging Spectrometer."

[18] Andika, F., Rizkinia, M., and Okuda, M., "A hyperspectral anomaly detection algorithm based on morphological profile and attribute filter with band selection and automatic determination of maximum area," *Remote Sensing* **12**(20), 3387 (2020).

[19] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., and Herrera, F., [*Learning from imbalanced data sets*], vol. 10, Springer (2018).

[20] Ma, Y. and He, H., [*Imbalanced learning: foundations, algorithms, and applications*], John Wiley & Sons (2013).

[21] Branco, P., Torgo, L., and Ribeiro, R. P., "A survey of predictive modeling on imbalanced domains," *ACM computing surveys (CSUR)* **49**(2), 1–50 (2016).

[22] Schaum, A., "Joint subspace detection of hyperspectral targets," in [*2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*], **3**, IEEE (2004).

[23] Hytla, P. C., Hardie, R. C., Eismann, M. T., and Meola, J., "Anomaly detection in hyperspectral imagery: comparison of methods using diurnal and seasonal data," *Journal of Applied Remote Sensing* **3**(1), 033546 (2009).

[24] Acito, N., Diani, M., and Corsini, G., "Gaussian mixture model based approach to anomaly detection in multi/hyperspectral images," in [*Image and Signal Processing for Remote Sensing XI*], **5982**, 209–217, SPIE (2005).

[25] Kwon, H. and Nasrabadi, N. M., "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing* **43**(2), 388–397 (2005).