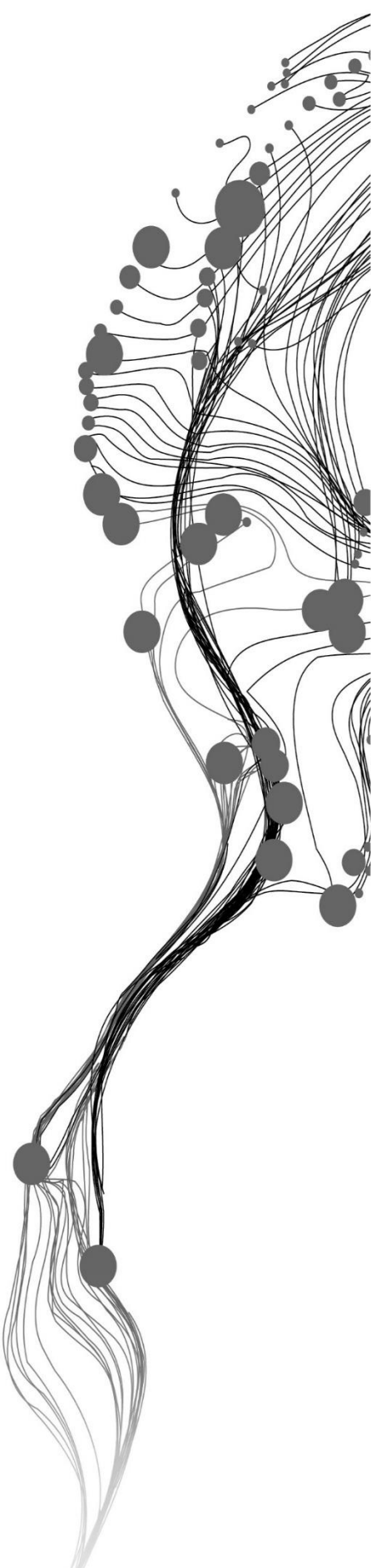


CRITICAL NODES FOR FOOD SECURITY: DETECTING FOOD STORAGE FACILITIES USING REMOTE SENSING IMAGES

EMMANUEL TOSIN SALAMI
JULY, 2024

SUPERVISORS:
Dr, C, Paris
Dr, Y, Dou



CRITICAL NODES FOR FOOD SECURITY: DETECTING FOOD STORAGE FACILITIES USING REMOTE SENSING IMAGES

EMMANUEL TOSIN SALAMI

Enschede, The Netherlands, JULY 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Resources Management

SUPERVISORS:

Dr, C, Paris

Dr, Y, Dou

THESIS ASSESSMENT BOARD:

Prof.Dr.ir. L.L.J.M. Willemen

Prof. Dr. Shaohua Wang (External Examiner, Aerospace Information Research Institute, Chinese Academy of Sciences)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the faculty.

ABSTRACT

The rising global population, projected to reach 9.1 billion by 2050, demands a 70% increase in food production. Currently, 1.3 billion tons of food are wasted annually due to post-harvest losses, particularly in developing countries. Despite the numerous research of silos as a food storage facility for food security, little has been done on the automatic detection of the silo's structures using earth observation data. This study aims to enhance food security by addressing the lack of information on the spatial coverage of grain storage facilities, especially silos. By employing multi-resolution optical earth observation data, namely High Resolution (HR) Sentinel-2 satellite data and Very High Resolution (VHR) aerial images from the National Agricultural Imagery Program (NAIP), this research develops innovative methods for the detection and mapping of silos. The research explores machine learning algorithms to automatically identify the silo's locations leveraging on their specific properties in terms of object's size, shape, and spectral signature. This study focuses on the United States as a case study due to its extensive grain production and storage capabilities. The research integrates multiresolution earth observation data, and machine learning algorithms to detect silos, aiming to contribute to sustainable development goals by reducing post-harvest losses and informing policymaking for food security.

Keywords: Food security; Image Analysis; Land cover; Silo detection; Unmixing, Sentinel 2; Earth Observation data; Very High Resolution (VHR) aerial images

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to the Almighty God for providing me with the strength, knowledge, and perseverance to complete this thesis. His blessings have guided me through every step of this journey.

I am profoundly grateful to my parents Mr. and Mrs. Salami and my sister Ibukun Coco for their unwavering support, encouragement, and love throughout my academic journey. Their belief in me has been a constant source of motivation.

I would like to extend my heartfelt thanks to my supervisors, Dr. Claudia Paris, and Dr. Yue Dou, for their invaluable guidance, patience, and support. Their expertise and insights have been instrumental in the successful completion of this thesis.

Additionally, I would like to express my appreciation to the entire faculty and staff of the ITC at the University of Twente. Their dedication and commitment to providing a conducive learning environment have greatly contributed to my academic and personal growth.

I am also deeply thankful to my friends Prince Boateng, Lawrence Lemgo, and Charlie Shawa for their support on my academic journey. Your inspiration and companionship have made this experience truly memorable.

Finally, I would like to acknowledge all my colleagues who have supported me in several ways during this journey. Your encouragement has made this experience truly memorable.

Thank you all.

TABLE OF CONTENTS

1.	Introduction	1
1.1.	Background and problem statement	1
1.2.	Research objectives.....	3
1.3.	Research questions.....	3
2.	Study area and dataset	5
2.1.	Study area	5
2.2.	Overall Methodology.....	6
2.3.	Dataset.....	7
2.4.	Specific methods	11
3.	Results	16
3.1.	Selection of the best season for built-up area detection.....	16
3.2.	Mapping the target built-up areas in agricultural land.....	17
3.3.	Refining detection results from the unmixing approach	23
4.	Discussion.....	33
5.1.	Mapping built-up areas within agricultural land.....	33
5.2.	Creating silo location maps with vhr images.....	33
5.3.	Scaling up the approach for large-scale detection	34
5.	Limitations, recommendations, and conclusion.....	35
6.1.	Limitation	35
6.2.	Recommendation.....	35
6.3.	Conclusion.....	36
6.	Ethical consideration.....	37
7.	Appendix	40

LIST OF FIGURES

Figure 1. Study area map	5
Figure 2. Research methodology flowchart.....	6
Figure 3. Research workflow.....	7
Figure 4. National agricultural imagery program images	8
Figure 5. Isolated and non-isolated silo image.....	8
Figure 6. VHR aerial images of the 3 counties.....	9
Figure 7. Spatial distribution of training data.....	9
Figure 8. Four seasonal sentinel-2 composite.....	12
Figure 9. UnFuSen2 application breakdown.....	13
Figure 10. Unmixing output for four season	16
Figure 11. Unmixing accuracy distribution per season.....	16
Figure 12-14. Unmixing results in county level	17
Figure 15-17. Comparison of NLCD map and unmixing results.....	19
Figure 18-20. Selected unmixing areas for RF and MLC	20
Figure 21-25. MLC map.....	23
Figure 26. Spatial distribution of MLC testing region.....	25
Figure 27. Random forest result for silo detection per county.....	27
Figure 28. Spatial location of built-up area detecting of the 3 counties	27
Figure 29-35. Random forest classification for silo location map.....	28
Figure 36. Spatial distribution of RF testing region	31
Figure 37-38. Comparison of RF and MLC base on accuracy.....	31

LIST OF TABLES

Table 1. Population, area, and proportion of main crop in study area	6
Table 2. Overview of data sources and uses.....	10
Table 4. Correctly and incorrectly aligned silos to urban.....	17
Table 5. Accuracy for selected unmixing	22
Table 5. Accuracy of MLC.....	25
Table 6. Correctly and incorrectly detected silo as built-up.....	26
Table 8. RF accuracy.....	30
Table 9. Unmixing accuracy for the best season.....	41

LIST OF ABBREVIATIONS

ASD	Absolute spectral difference
MLC	Maximum likelihood classification
RF	Random forest
NAIP	National agriculture imagery program
NLCD	National landcover database
UnFuSen2	Unmixing-based Sentinel-2

1. INTRODUCTION

1.1. Background and Problem Statement

Addressing the food requirements of an exponentially growing global population is becoming a significant concern for humanity. According to FAO, (2009), the global population will increase by 34% with about 9.1 billion people by 2050, and food production must rise by 70% to suit the demands of such a big population. Notwithstanding this circumstance, 1.3 billion tons of food are wasted globally each year, and 48 million people can be fed each year around the world if these losses are reduced (International Institute of Information Technology, 2016). This seems difficult to achieve because of post-harvesting losses experienced globally. One main contributor to food insecurity particularly in Africa is post-harvest losses (Tefera, 2012a). Throughout every stage of cultivation and after-harvest, encompassing reaping, managing, storing, processing, marketing, and final distribution, there is a wide range of both quantitative and qualitative food wastage (losses), which vary in magnitude (Prusky, 2011). Food storage facilities are identified as "critical nodes" in the food supply chain. These nodes are critical in reducing food wastage and enhancing food security, especially in regions prone to post-harvest losses (Rosalia et al, 2019). In various parts of the world, post-harvest losses have a notable impact and result in substantial economic losses. By decreasing post-harvest losses, it becomes probable to attain sustainability by balancing the economic, social, and environmental aspects (Raut et al., 2018). In 2015, the United Nations incorporated the reduction of harvest and after-harvest losses as a Sustainable Millennium Development Goal as established in SDG 12 target 12.3, this inclusion has sparked greater interest among researchers and policymakers to address these losses, aiming to ensure food security (Arends-Kuenning et al., 2022).

Numerous studies have emphasized the importance of improving storage facilities to decrease postharvest losses (Tefera, 2012b). According to Kumar & Kalita, (2017) the absence of proper storage techniques in developing countries (e.g., Nigeria) can result in as much as 50% to 60% losses of cereal grains during the storage phase due to technical inefficiencies. By providing suitable storage conditions, farmers can mitigate the risks of pests, insects, rodents, and fungal infestations that can cause spoilage and deterioration of stored crops (Befikadu,2018). In China, using steel silos has the potential to enhance the rice supply, save land, fertilizer, and water, lower carbon emissions, and meet the requirements of approximately 1.39 million people (Luo et al., 2021.). Effective storage methods, such as the usage of metal silos can help maintain the quality and nutritional value of the harvested crops for longer durations (Tefera et al., 2011).

When referring to key points or locations relating to maintaining and improving food security, "critical nodes" are important as they help prevent the breakdown of stability and efficiency within the supply chain in the distribution and storage of food. The storage facilities, such as silos, are particularly critical because they are vital in reducing post-harvest losses and ensuring the constant, sustainable availability of food (Sartori et al, 2015). However, there is limited knowledge regarding the spatial location and coverage of these storage facilities, leading to a missed opportunity in creating effective food security policies. According to (Grain Silos and Storage System Global Market Report, 2023), North America is the largest region with grain silos as of 2022. Other regions include Asia-Pacific, South America, Western Europe, Eastern Europe, Middle East, and Africa. The United States topped the list with the most silos, follows by China and Brazil, (Chen et al., 2022). However, there is no information of the intensive use of this storage facility and its database in Africa and some other developed countries. Considering the present lack of concrete information regarding the spatial coverage and capacity of food storage facilities (silos), it is crucial to initiate thorough research that validates and addresses this critical gap. This investigation aims to bring out information relating to the spatial distribution and coverage of these critical nodes by focusing on the

detection and mapping of these facilities using advanced remote sensing technologies. Such an effort is key in understanding and addressing the gaps within the global food supply chain, particularly in areas where data on these facilities are minimal or non-existent.

This study introduces a new approach to identify and analyse these critical nodes using earth observation data, including Sentinel-2 and Very High-resolution Imagery (VHR). Because silos are circular in shape and can be misinterpreted to other features on ground (such as oil tanks), focus of this research will be on pixel base image analysis, where the reflectance value of the pixel will be taken in consideration to provide a better result in delineating silos. It will also include environmental features, taking into consideration the surrounding croplands which could influence the location of silos for storage facilities.

The research will focus on The United States of America (USA) as a case study due to following reasons. First, the USA is one of the top suppliers of grains in the world with an estimation of about 7,400,678 stocks (i.e., large storage of grains measure in bushels) of corn, 108,769 stocks of sorghum, 42,909 stocks of oats, 88,722 stocks of barley, 945918 stocks of wheat, and 1,686,632 stocks of soybeans as of March 2023 (United States Department of Agriculture (USDA), 2023). Second, with approximately 13,580 on-farm and 11,822,820 off-farm storage facilities (silos) strategically distributed across the USA for storing these enormous quantities of grains, the United States stands as an ideal location for this research. Also, this rich resource landscape provides a unique opportunity to gain deeper insight into the spatial proximity of crop farms to each silo. Such insight helps conduct meaningful spatial analyses. The investigation primarily focuses on the United States due to its rich availability of data, which is essential for assessing the feasibility of the study. Recognizing the parallel challenges in developing countries, particularly in regions with significant food insecurity, it is relevant to note that these areas often feature grain storage silos similar to those in the U.S. This similarity provides a critical opportunity to apply findings from the U.S. context to enhance silo efficiency in developing regions. By doing so, the study can contribute to strategies that improve food storage and security in environments that are markedly different yet functionally comparable to the study area.

Remote sensing is a valuable tool for enhancing food security, but there is a noticeable shortage of research focused on utilizing remote sensing imagery and techniques for silo detection. While most research concentrates on aspects such as crop extent, yield, and recognizing crop conditions, there is currently limited attention given to post-harvest estimation which can be improved using silos. However, there have been some studies conducted on the detection of circular tanks, which share similarities with silos. One of the studies was conducted by Tadros et al., (2020). In this study, the authors initially used circular detection algorithms for identifying tanks but faced issues with false detections. To address this, they implemented clustering algorithms and acknowledged the method's limitations, particularly in recall rate. They further concluded that further research is needed to improve accuracy and reliability in tank detection.

Another study in this context was conducted by (Wang et al., 2018). The research introduces a new method for detecting oil tanks using a Goal-driven knowledge saliency model, incorporating circular feature maps to improve saliency. The authors used a boosting classifier to generate a global saliency map, refined with initial saliency results. Xia et al., (2018) also gave important research on oil tank detection. Here the authors offered an innovative approach for detecting circular oil tanks in high-resolution remote sensing images using deep learning. They used the Selective Search algorithm to identify targets and trained the CaffeNet network within the deep learning Caffe framework as a feature removal classifier. The method successfully identified and marked oil tanks, showing effectiveness in various backgrounds with improved detection and reduced false alarms. Tadros et al., (2020) use salient object extraction and oil tank shape classification method to detect circular-shaped object (oil tank) in low resolution satellite images. Zalpour et al., (2020) present an innovative approach to detect oil tanks using deep features, merging convolutional neural network and histogram of oriented gradients to achieve superior detection accuracy. Although these methods yield effective results in detecting oil tanks, they must solve other challenges when applied to silo detection.

Adapting methods designed for detecting oil tanks to silo detection presents several real-world challenges. First, unlike oil tanks, silos are often found amid complex backgrounds, making accurate detection harder. In the analysis of agricultural and industrial landscapes using remote sensing techniques, the characterization of various storage facilities, such as silos and oil tanks, is critical. Silos, primarily used for storing bulk materials like grains, generally exhibit larger dimensions, typically ranging from 10 to 30 meters in diameter. In contrast, oil tanks, which are utilized for the storage of liquid substances, often vary in size from 5 to 25 meters, reflecting their construction to accommodate different volumes and pressure requirements (Vogt & Gerding, 2017).

Due to the varying sizes and resolutions of silos, utilizing the oil tank detection method can result to inaccuracy of the outcome of detected silos. Balancing false alarms and missed detections is crucial to prevent silos to be misclassified as oil tanks or other circular objects. Overcoming these obstacles demands careful dataset curation, model tuning, and real-world testing. While metal silos bear a circular shape and possess structural similarities with oil tanks, it is crucial to recognize that they serve distinct purposes. Therefore, to accurately map silos, the contextual analysis of the surrounding areas is important. That is, to identify built-up areas within the large agricultural land areas. This calls for a more practical approach to balance the high-resolution imagery demand and large spatial extent.

According to Tong et al., (2023) high-quality satellite imagery such as Google and QuickBird offer detailed spatial data, which is crucial for categorizing land cover, especially when analysing man-made structures like silos, buildings, and parking lots around agricultural environments. This research will create a practical approach and workflow to locate silos by first employing Sentinel 2 imagery to identify the target built-up areas, and then using Random Forest and Maximum Likelihood Classifier on high-resolution satellite imagery to map silos within these areas.

The research will follow two steps in detecting silos from remote sensing images. The first one is using the Unfused Sentinel-2 approach, which classifies land cover on the bases of their endmember with focus on the urban band. The second step is using machine learning techniques such as Random Forest and Maximum Likelihood Classifier to detect silos using VHR for a more accurate delineation of silos; and identifying which agricultural cropland is closer to silos location. The advantage of machine learning includes its accuracy, scalability in terms of its application to large data, its adaptivity to different environmental condition, and reduction of false positives. Using ML algorithms and geospatial analysis in detecting these silos will provide critical insights for stakeholders to make informed decisions in addressing food insecurity. The overall aim of the suggested framework in this study is to support spatial identification of silos in other food insecure regions. Although the United States is the case study of the research, the real impact of the thesis will help understanding the availability and distribution of storage facilities around agricultural locations of over food-insecure regions. This will help in minimizing food wastage and enhancing food security.

1.2. Research Objectives

The research objectives are:

- a) To develop an approach that can be used to automatically identify built-up areas within agricultural land that may include silos, using high spatial resolution Sentinel-2 data.
- b) To determine the presence of silos within the identified built-up areas using Very High-Resolution (VHR) images at the local scale.

1.3. Research Questions

The research questions are:

- a) What is the optimal season to detect built-up areas located in agricultural regions that may have silos using high spatial resolution Sentinel-2 data?
- b) Which unsupervised approach can be used for automatically detecting built-up areas located in agricultural regions?
- c) What is the optimal machine learning classification model to determine the presence of silos within the identified built-up areas using VHR images?

2. DATA AND METHOD

2.1. Study Area.

The study area for this research is in the United States of America. This is because of its vast land and varied topography which presents a diverse agricultural field. Within this broad landscape, the Midwest region stands out as the country's grain belt, producing a huge portion of America's staple crops. At the heart of this study is a detailed exploration of grain silos within Iowa, Minnesota, and Illinois three pivotal states in the Midwest known for their agricultural output. The choice of the U.S. as the primary location for this research is influenced by the extensive use of silos throughout the nation. These structures are indispensable for the post-harvest storage of grain and other agricultural commodities, allowing large volumes of produce to be securely stored until they are either shipped to markets or used as seeds for the next growing season. The presence of silos is not just a testament to the U.S.'s ability to manage large-scale agricultural production; it also offers a wealth of opportunities to investigate various storage methods, preservation techniques, and the logistical aspects of modern agriculture. Focusing on Iowa, Minnesota, and Illinois, the study aims to delve into the agricultural prominence of these regions and the critical role that grain silos play there.

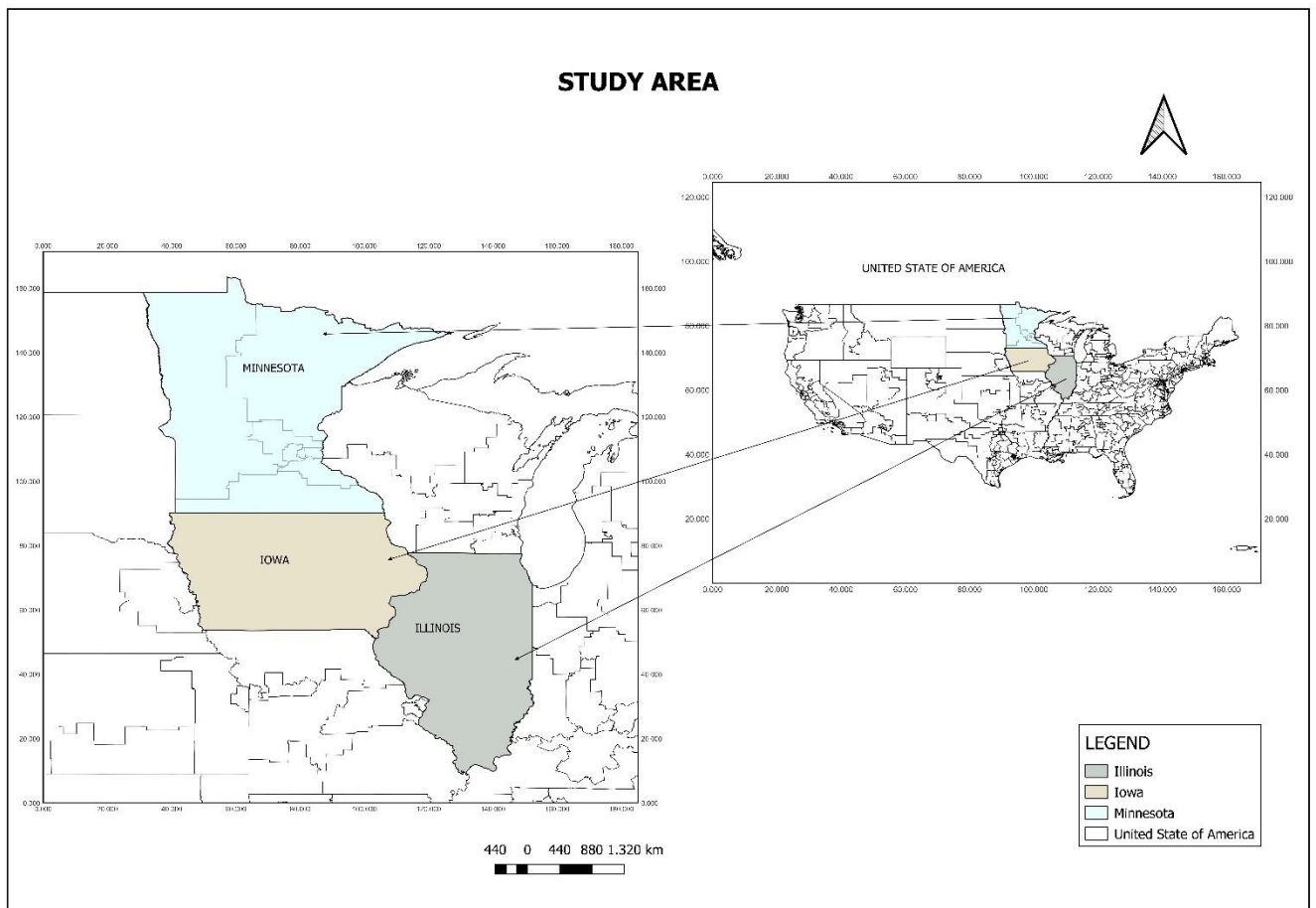


Figure 1. Map of the study area showing the main areas of focus in the United States of America which are Minnesota, Iowa, and Illinois.

Table 1. The population, area in square kilometer (sq km), proportion of agricultural areas, and main crops of the study area reported per state. This shows how presence of these crops influences the abundance of silos in the study area.

State	Population (2023)	Area (sq km)	Prop. of Agric. Areas	Main Crops
Iowa	3,203,345	145,746	85%	Corn, Soybeans
Illinois	12,477,595	143,742	75%	Corn, Soybeans
Minnesota	5,722,897	206,175	51%	Soybeans, Corn, Wheat

2.2. Overall Methodology

This study utilized and integrated different methods and dataset to detect silos across-scales. This includes using Sentinel-2, a 10-meter resolution image at large scale to identify potential target areas for silos with an automatic detection; and machine learning method on very high-resolution images to detect silos at a farm and county level.

The image below shows the workflow of the research methodology. Two general steps are taken. The purpose of the first step is to detect built-up areas through Sentinel-2 imagery across the 37 selected areas, and on three counties (Mower, Story, and Dewitt) outside the 37 areas. To do so, first, the 37 areas were used for determining the qualitative assessment of the best season to detect built-up area based on band corresponding to endmembers of urban to compare which season silos are more visible. Silos in the area were delineated and used to assess the quantitative performance of this season. The next step involves the use of RF and MLC to detect silos on these areas after built-up areas have been identified. RF and MLC classifiers were evaluated and compared on National Agriculture Imagery Program (NAIP) imagery to generate a silo location map. Using the NAIP data as training and testing data for the algorithms.

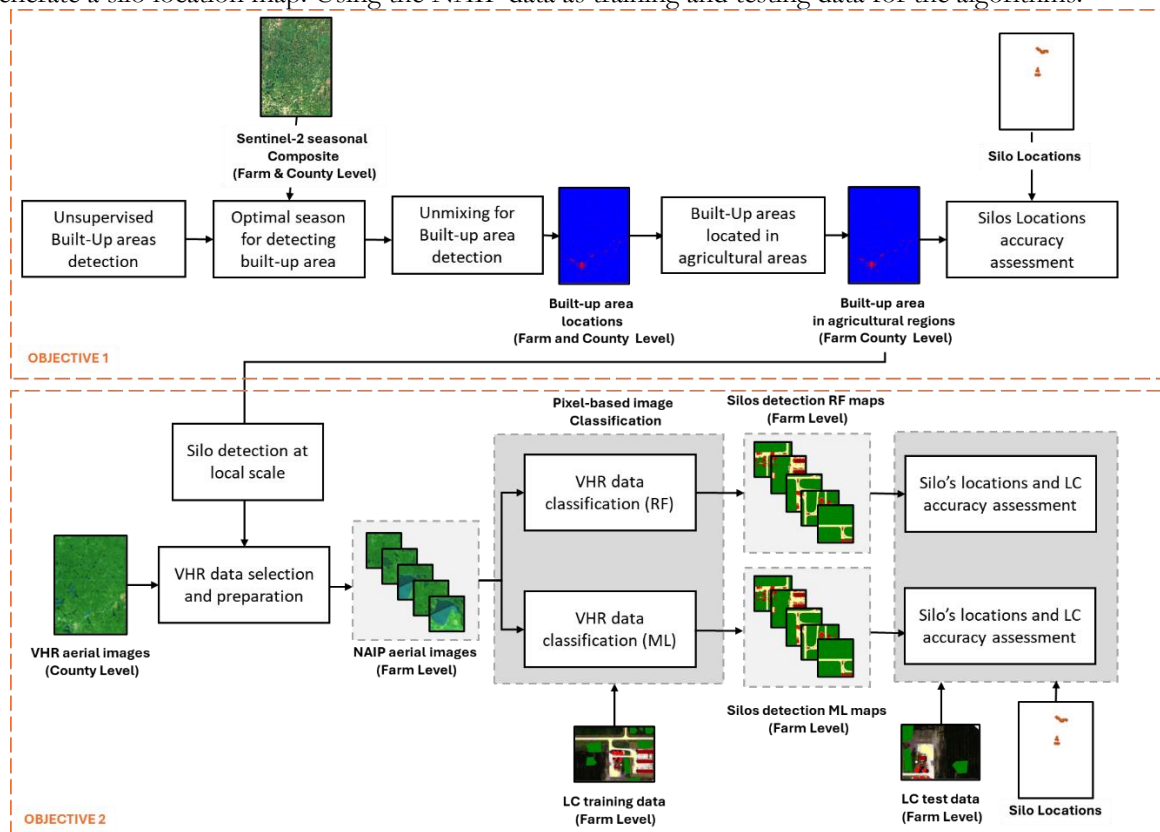


Figure 2. Research methodological flowchart.

The spatial levels used in this research are the farm and county level. The 37 selected areas are all on farm level. The counties used are Mower in Minnesota, Story in Iowa, and Dewitt in Illinois. Both identifying of built-up areas and detection of silos were performed on the farm and county level.

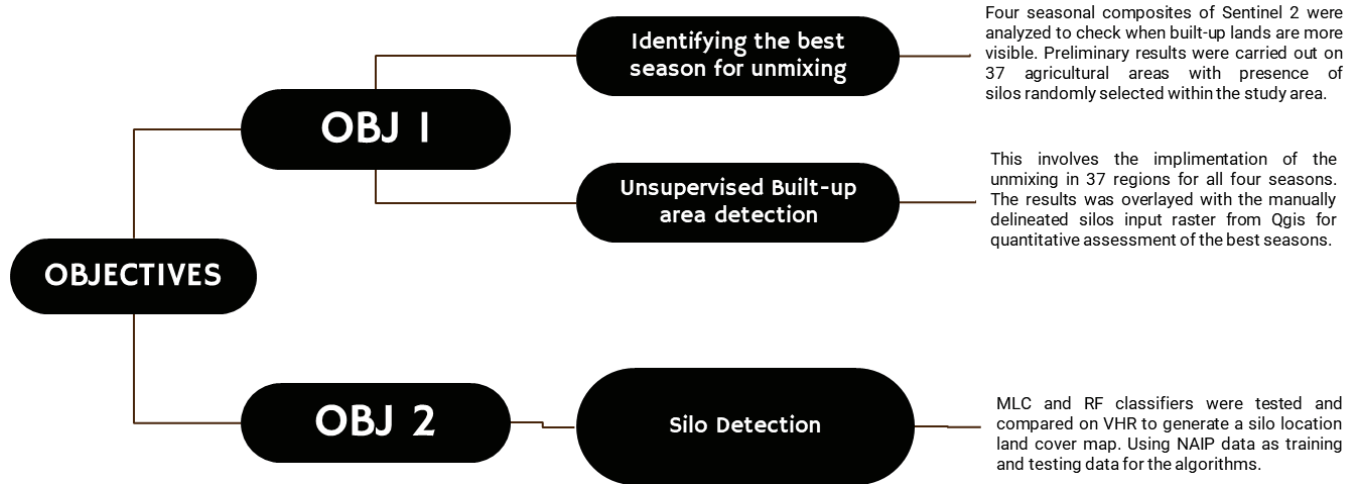


Figure 3. Research workflow.

2.3. Dataset

The study utilizes Sentinel-2 satellite data, VHR images and NAIP aerial images, the silo's location data along with GIS boundary data from Iowa, Minnesota, and Illinois. These data include temporal and spatial details essential for training and validating the silos detection results obtained in the considered study area. All selected data sets are freely accessible and consistent over time, with no personal data included.

To accurately establish the farm sites for both the unmixing and subsequent classification processes utilizing Random Forest (RF) and Maximum Likelihood Classification (MLC) methods, a 200-meter buffer was established around each silo identification number in the whole study area, using the NAIP imagery. The delineated silo sample of all three states was used here. This creates a loop around all the silos in GEE, where 37 random areas are selected. This careful integration of geospatial buffers and identification markers facilitated the selection of 37 optimal locations for detailed classification, thereby enhancing the precision and efficacy of the study's analytical outcomes. The image was also used for the unmixing built-up detection. The images spanned from January 1, 2019, to August 30, 2023, a period chosen for its stability in providing consistent classification results irrespective of the variations in the training sample sizes. This approach generated a focused training set that enhanced the learning process of the algorithm, using high-resolution imagery from the NAIP as a base.



Figure 4. NAIP image used for silo detection.

These show 20 out of the 37 randomly selected NAIP images used for maximum likelihood and random forest classification for mapping the location of silos. The spatial resolution of the NAIP image covers the entire United States. These areas are also used for qualitative and quantitative evaluation of the optimal season for the automatic detection of built-up lands.



Figure 5. Image of isolated and non-isolated silos

Silos have different spatial configurations too. As shown in Figure 5, some silos are isolated on a farm (A), while others can form complex topological networks that integrate storage, processing, and distribution of grain and other farm produce (B).



Figure 6. VHR images of Mower County (Minnesota), Story County (Iowa), and Dewitt County (Illinois)

These (Figure 6) are the 3 county areas outside the 37 areas, where the detection of built-up was made in county level.

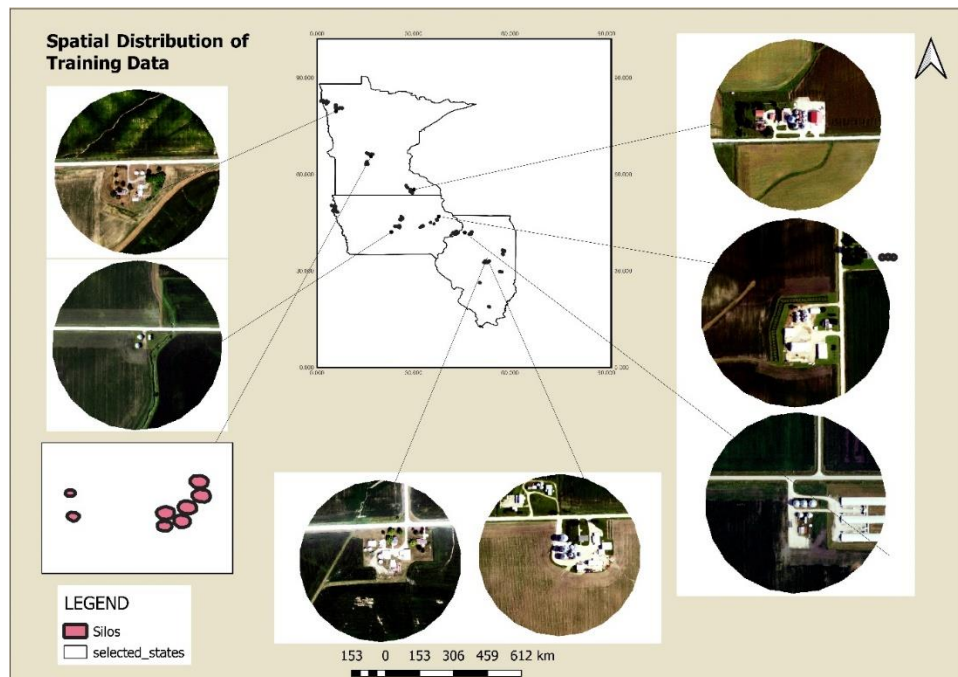


Figure 7: The spatial distribution of the training region. This also contain 186 delineated silo sample in the 37 regions.

Figure 6 illustrates the geographical locations of some training images used in this research. It consists of NAIP images of the silo location. This images also serves as a raster layer for training and testing the classification. Additionally, the figure displays the delineated silo samples which were also used for quantitative assessment of best season for detecting built-up areas. The delineated silos not only served to establish a 20-meter buffer around the agricultural area to generate the 37 random areas, but also enhanced the focus on local silos identified through RF and MLC.

Table 2. Overview of the data sources used in the considered study and their properties.

Data Source	Type of Data	Spatial Resolution	Temporal Resolution	Data Availability	Use in study
Sentinel 2	EO	10m.	5 days.	Free access. Google Earth Engine	Create built-up location in agricultural land
Iowa Select County Boundary	GIS	N/A	N/A	Iowa County Boundaries Iowa County Boundaries geodata	Training and testing area selection in Iowa State
Minnesota Select County Boundary	GIS	N/A	N/A	County Boundaries, Minnesota - Resources - Minnesota Geospatial Commons (mn.gov)	Training and testing area selection in Minnesota State
Illinois Selected County Boundary	GIS	N/A	N/A	Illinois State Boundary clearinghouse.isgs.illinois.edu	Training and testing area selection in Illinois State
National Agricultural Imagery Program (NAIP) Satellite Image	EO	1.2m	N/A	QGIS	Use to generate classification maps
United State Department of Agric. (USDA) Cropland Data	EO	30m	16 days	https://developers.google.com/earth-engine/datasets/catalog/USDA_NASS_CDL	Use to determine cropland proximity to silo location
Training Data (Manually Delineated Data of All Class)	EO	1m	N/A	QGIS and National agriculture imagery program https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ	Training and testing model for classification and accuracy assessment
NLCD DATA	EO	30m	N/A	https://www.mrlc.gov/data	Assessing the built-up area detection on county level

2.3.1. Earth Observation Data Preparation

Both Sentinel 2 and NAIP Imagery were used in this step. The Sentinel-2 data selection was made to reduce cloud coverage to less than 5% through cloud masking, which minimizes presence of cloud in the image. The Sentinel-2 was used to achieve the first objective of identify the target built-up area. The NAIP was acquired from GEE catalogue. There is also selection of the NAIP by date from 2019 till August 30, 2023, which makes its temporally significant for this study. The NAIP was included to achieve the second objective which is to determine the presence of silos within the identified built-up area. The training data of the silo location was manually delineated and derived by leveraging VHR data.

2.3.2. Sample Data

A total of 186 silos were delineated across 37 selected areas spanning the study region: 65 silos were identified in Minnesota, 76 in Iowa, and 45 in Illinois. The delineation process involved the use of high-resolution aerial imagery sourced from the NAIP and QGIS to accurately identify and outline each silo.

2.4. Specific methods

2.4.1. Season selection for detecting built-up areas in Sentinel-2 images.

To ascertain the best season for the accurate detection of silos and areas probable to be silos utilizing Sentinel-2 images, this study conducted a qualitative and quantitative assessment to compare composite satellite images from various months. The study focusses on thirty-seven (37) randomly selected areas within the study area. To achieve the qualitative assessment, the application of the unmixing was done for four (4) different seasons, with focus only on the urban (built-up) class, while using the threshold of 0.75. The manually delineated silo raster was created for each of the 37 randomly selected areas and intersected with the unmixing (urban) result. This is conducted to determine the best season that provides a good image where the delineated silos fall on the urban unmixing class. Also, the quantitative assessment was achieved by creating a binary confusion matrix which summarizes the results obtained per season in the whole area. The aim is to identify a period where artificial surfaces, such as metal silos, are most discernible and distinguishable from natural landscapes. This is important because areas covered by vegetation provide a better contrast than bare soil which can be confused with artificial surfaces. The most effective seasonal timing for the EO data can sharply improve the results obtained when applying the unmixing technique, thus effectively and for ensuring a higher accuracy in detecting areas characterized by artificial structures which are probable to silos. This analysis is essential, as seasonal changes can significantly affect the visibility of the artificial surfaces due to agricultural activities, and vegetation cover which can appear to be man-made structures in leave off or dry season. The chosen season shows the optimal lighting conditions, thus minimizing the risk of misclassification and enhancing the reliability of remote sensing data for the detection and mapping of silo structures.



Figure 8: Qualitative example of the four seasonal Sentinel-2 composites for a specific silo's location.

2.4.2. Built-up area detection

UnFuSen2 stands for "*Unmixing-based Sentinel-2 Image Fusion*" (Xu & Somers, 2021), designed to enhance Sentinel-2 imagery by merging high spatial detail with rich spectral data. This technique generates images with superior spatial and spectral resolution, useful for identifying built-up land. The unmixing was first applied on farm level (37 selected areas) to detect the built-up area in a local scale. It was also applied to the county level outside the 37 farm areas, to get a view of the research performance in a larger scale, and to achieve the generalization of the method and its applicability in other geographical location, demonstrating its utility for scaling up in similar studies. This method was implemented using Google Earth Engine on Sentinel 2 image, which allows efficient processing and analysis. Validation was conducted by comparing the unmixing results with ground truth images from the NAIP dataset and manually delineated silo training samples. The Unmixing methodology breakdown includes.

- **Endmember Identification:** The process begins by using spectral matching to identify unique features (endmembers) like silos in high-resolution bands of Sentinel-2 images. Absolute Spectral Difference (ASD) calculations help set a spectral uniqueness threshold, aiding feature differentiation.
- **Mixing Equations Construction:** Next, mixing equations connect the reflectance of 20-meter pixels to corresponding 10-meter pixels. This involves identifying similar coarse pixels through a geographical search, using ASD to ensure accuracy in feature matching.
- **Constraints Application:** Constraints are applied to the mixing equations to maintain realistic reflectance values. These include an Equality Constraint that models the target pixel's reflectance as a weighted average and a Regression Model Constraint that ensures spectral realism.
- **Solution Refinement:** The final step involves refining the reflectance values using a constrained least squares solution, ensuring accuracy for practical applications like silo detection.

The formular for this assessment;

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Where True positive is number of silo pixels correctly identified as urban, False Positives is number of silo which are not classified as urban.

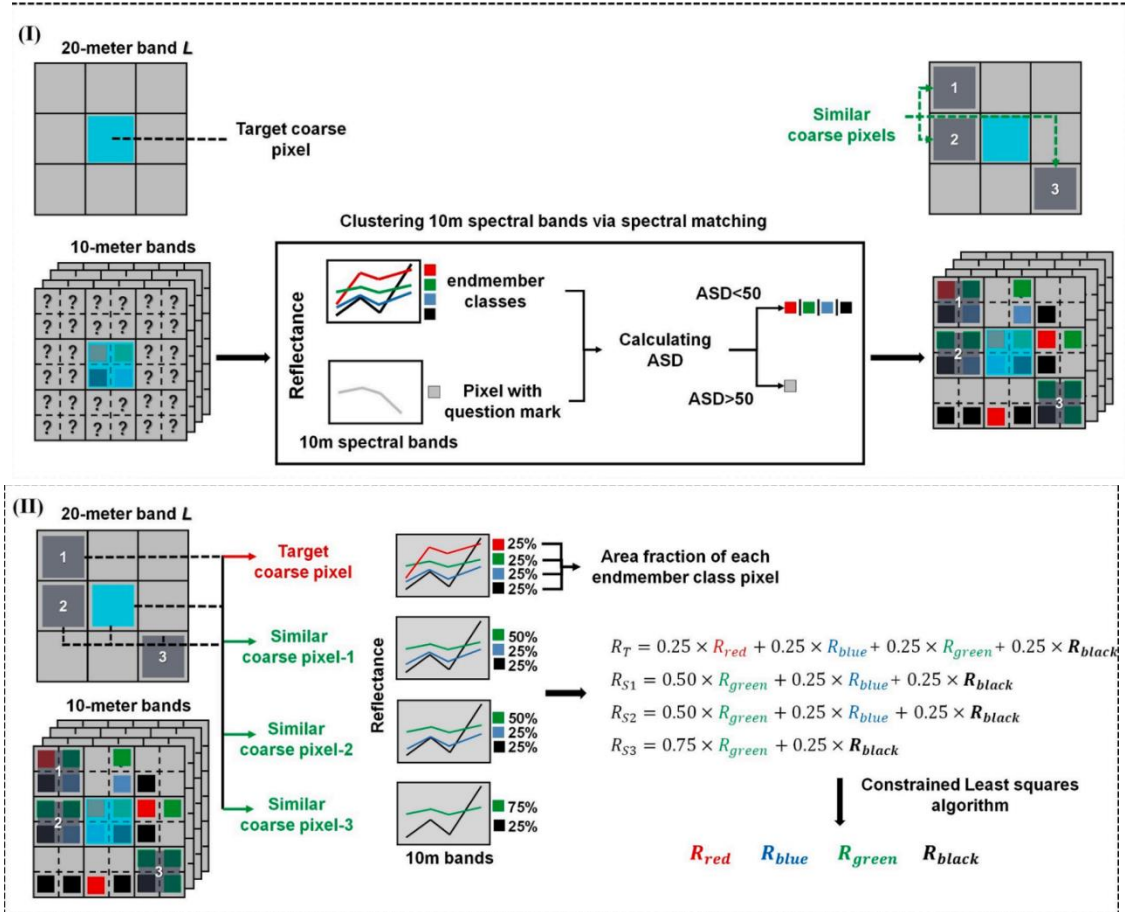


Figure 9. Source: (Xu & Somers, 2021). The Unfusen2 application breakdown.

The image shows step-by-step process to construct the set of equations needed to mix a designated 20-meter pixel from Sentinel-2 imagery (referred to as the target coarse pixel) involves: (I) identifying comparable coarse pixels through spectral comparison; (II) creating the set of mixing equations using the identified similar coarse pixels.

2.4.3. Silo Detection at Local Scale

After the areas probable to be silos have been identified, silos were detected on a farm level and county level for accurate delineated from other classes using the NAIP imagery. For this purpose, two methods were used: the Maximum Likelihood Classifier and Random Forest.

2.4.4. Maximum Likelihood Classification

The Maximum Likelihood Classifier was used for this objective. The Maximum Likelihood Classifier is a statistical technique used in landcover classification. It predicts each pixel's class by estimating chances based on the pixel's features and comparing them to known categories (Lillesand et al. 2014). It is effective for categorizing land types in remote sensing images, using a normal distribution model for accuracy (Ali et al., 2018). Pixels are allocated to the class with the highest likelihood, making it crucial to choose training samples so that each training class adheres to a Gaussian distribution (Mishra et al., 2017).

The study employs Maximum Likelihood Classification (MLC) to identify silos through land cover classification within a research area divided into four primary categories: Silo, Building, Vegetation, and Bare Land. This technique refines the accuracy of land cover classification. Utilizing the robust dataset from the National Agricultural Imagery Program (NAIP), the analysis focused on the 37 areas initially examined, selected randomly to encompass diverse land cover features critical for assessing the algorithm's performance in complex environments. Unlike unmixing methods that focus solely on urban bands, MLC considers all surrounding features of a silo, enhancing the differentiation from adjacent areas.

The training and subsequent testing of the model were carried out on the NAIP dataset using the Envi GIS application, where only polygon data were used to train the MLC algorithm. For sampling, a stratified random sampling method was used to ensure a comprehensive and representative selection of data points across all categories, thereby enhancing the reliability and validity of the classification results. Although MLC is a good method for feature detection and land cover classification, it also has limitations. A key limitation of the MLC is its assumption that the statistical distribution of each class in feature space follows a normal (Gaussian) distribution. This assumption often does not hold true in real-world scenarios, leading to inaccuracies in classification, especially in heterogeneous or complex landscapes where data distributions can significantly deviate from the normal.

The following equation is used to calculate Maximum Likelihood Classifier:

$$\text{A data point } x \text{ is assigned to class } \omega_i \text{ if for all } j \neq i: P(x|\omega_i) > P(x|\omega_j)$$

Where $P(x|\omega_i) \sim N(\mu_i, C_i)$, x is the data point being classified, ω_i is the Class label I , $P(x|\omega_i)$ is the conditional probability of data point x given that it belongs to class ω_i , $N(\mu_i, C_i)$ is the normal (Gaussian) distribution of class ω_i , characterized by mean μ_i and covariance matrix C_i . μ_i is the mean vector of class ω_i . C_i is the covariance matrix of class ω_i .

This rule states that a data point x is classified into the class ω_i for which the probability $P(x|\omega_i)$, under the Gaussian distribution with mean μ_i and covariance matrix C_i , is highest compared to all other classes (Guan, Y. 2010). MLC will be applied to 37 farm area.

2.4.5. Random Forest Classification

This study utilizes Random Forest Classifier (RF) for silo detection. The Random Forest classifier, used in landcover classification, is a robust machine learning technique. It builds multiple decision trees, using different subsets of data, and merges their results for improved accuracy and control over overfitting (Pal, 2005). Each tree in the forest votes for a class, and the majority vote decides the final class. This method effectively handles large datasets with multiple features, making it ideal for classifying diverse landcover types based on satellite or aerial imagery.

For this study, the NAIP with no cloud interference was selected. The imagery, characterized by high-resolution visuals, was specifically chosen to enhance silo cover estimation accuracy. These images were accessed through the NAIP catalogue in GEE. Additionally, to enhance the precision and reliability of the

landcover analysis, a 2021 agricultural land cover map with a 30-meter spatial resolution was utilized. This map plays an essential role in accurately identifying various landcover types situated near the silos, ensuring a more comprehensive understanding of the surrounding agricultural landscape. Four landcover classes will be put into consideration in this analysis which are: Silo, Building, Vegetation and Bare land.

The random forest approach involves creating an ensemble of trees, thereby enhancing classification accuracy and robustness. Each tree in the random forest is constructed using a randomly selected subset of the training data (bootstrap sample) and a random subset of features. The final classification of a data point is determined by the majority vote across all trees in the forest, which aggregates their individual predictions. This ensemble decision process can be mathematically expressed as:

$$RF(x) = \text{mode} \{ T_1(x), T_2(x), \dots, T_n(x) \}$$

Here, $RF(x)$ represents the final classification decision by the random forest for a data point x , $T_i(x)$ denotes the classification output of the i -th decision tree, and n is the total number of trees in the forest. Each tree contributes a class prediction based on its constructed decision rules, tailored to the specific subset of features and data it was trained on. Rf will be applied to 37 farm area and 3 counties.

2.4.6. Comparison of MLC and RF results based on accuracy.

Both the results obtained from Maximum Likelihood Classifier and Random Forest Classifier were compared based on accuracy to determine which is better for delineating silos from other artificial structures. Misclassification of class will lead to a higher false positive.

The formula used to calculate the overall accuracy of RF and MLC per area is.

$$\text{Overall Accuracy} = \frac{\text{Sum of Diagonal Elements (True Positives)}}{\text{Total Number of Pixels (All Elements)}}$$

Where Sum of Diagonal Elements (True Positives); are the correctly classified instances for each class in the confusion matrix. Each element on the diagonal of the confusion matrix represents the number of times a particular class was correctly predicted.

Total Number of Pixels (All Elements); are total number of instances or pixels considered in the classification, which includes both correctly and incorrectly classified instances. It is the sum of all the elements in the confusion matrix.

The individual class accuracy was calculated using the formula.

$$\text{Class } i = \frac{\text{True Positives for Class } i}{\text{Total Ground Truth for Class } i}$$

Where Class i ; represents the accuracy metric for a specific class i .

True Positives for Class i ; are the instances correctly classified as class i . It is the number of times class i was correctly identified by the classification model.

Total Ground Truth for Class i ; represent the total number of actual instances of class i in the dataset, which includes both correctly and incorrectly classified instances.

3. RESULTS

3.1. Selection of the best season

The unmixing was used for mapping built-up areas within agricultural land through downscaling the high spatial resolution of Sentinel-2 images.

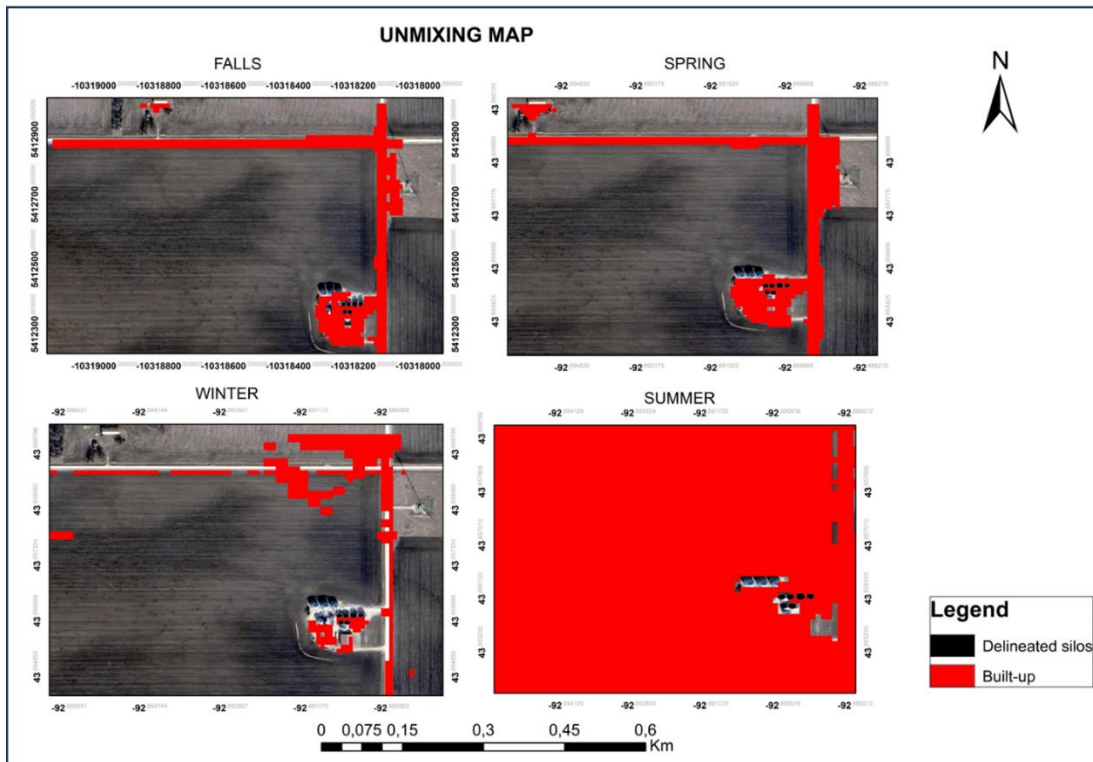


Figure 10: The unmixing output of 4 areas amongst the 37 areas used for qualitative evaluation of the best season through the unfunsen2 approach.

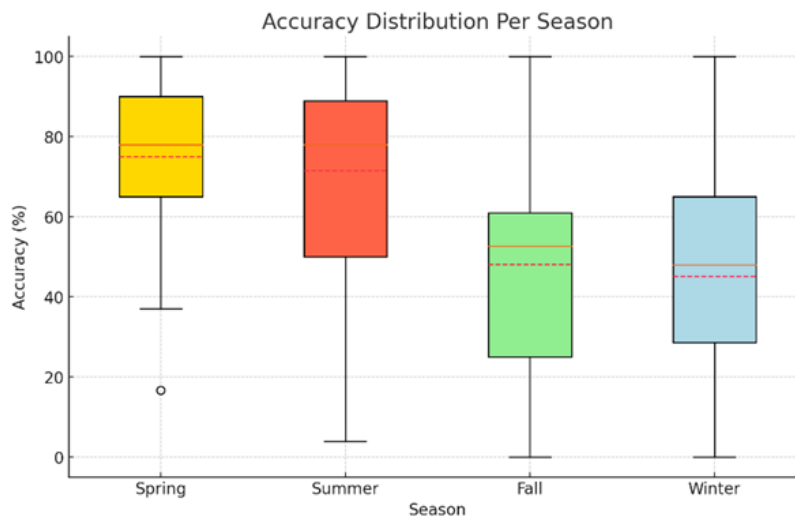


Figure 11: Box plot showing the accuracy distribution of the 37 areas per season

The Spring season was the best for applying the unmixing based on the qualitative and quantitative assessment of the 37 local areas. It was the best season because of the improved visibility and image quality during the time, and difference vegetation species can be distinguished easily from bare and other built-up land. It was used for further analysis in this research.

3.2. Mapping the target built-up areas in agricultural regions

In the next step, unmixing was applied at county level. The Dewitt County in Illinois, the Story County of Iowa, and the Mower County of Minnesota were selected for this analysis.

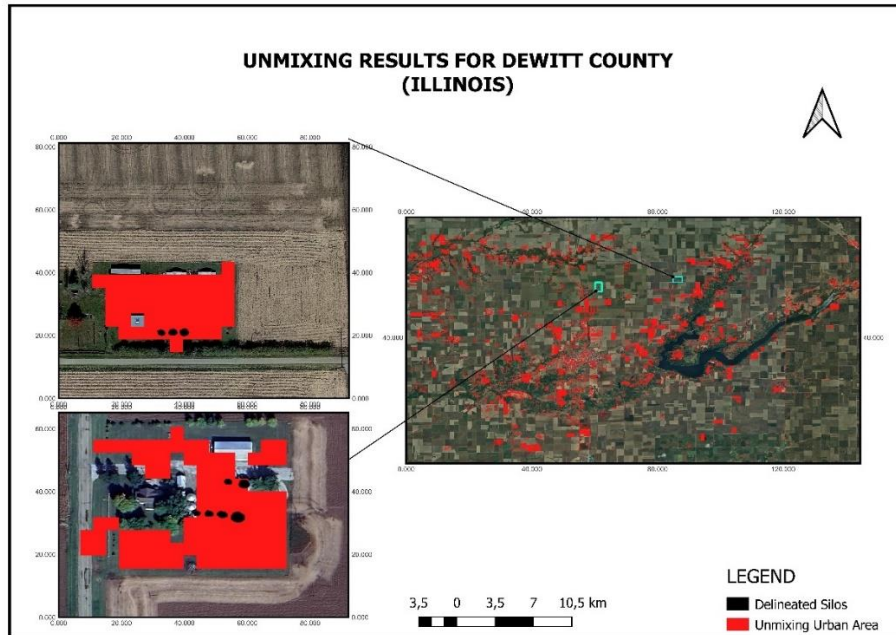


Figure 12: the unmixing results for Dewitt County of Illinois.

The unmixing analysis for DeWitt County, Illinois, as depicted in the provided map, demonstrates how built-up land is identified in a county level. This unmixing result is outside the 37-farm area. The central map highlights the built-up (red areas) interspersed with agricultural fields, while detailed insets offer a closer look at specific regions where built-up areas and silos (black) are identified. Out of 48 known silos, the technique successfully identified 33, achieving an accuracy rate of approximately 68.75%, though 15 silos were missed. This study indicates that while the UNFUSEN2 technique is promising for rural built-up area detection, further refinement and integration was done through MLC and RF to improve its accuracy and reliability.

Table 3 shows the total number of silos that are correctly and incorrectly detected by the unmixing as urban in the 3 counties (one per state)

County	Number Of Silos in Built-Up Areas	
	Correct detection	Missed detection
Dewitt (Illinois)	33	15
Mower (Minnesota)	55	29
Story (Iowa)	47	30

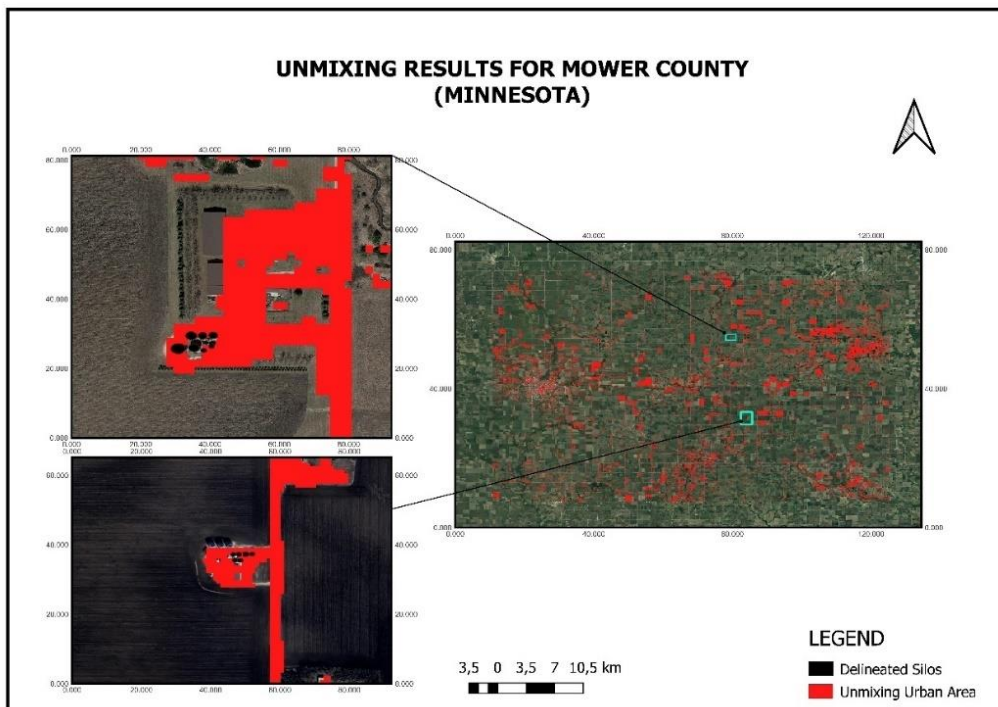


Figure 13: The unmixing results for Mower County of Minnesota.

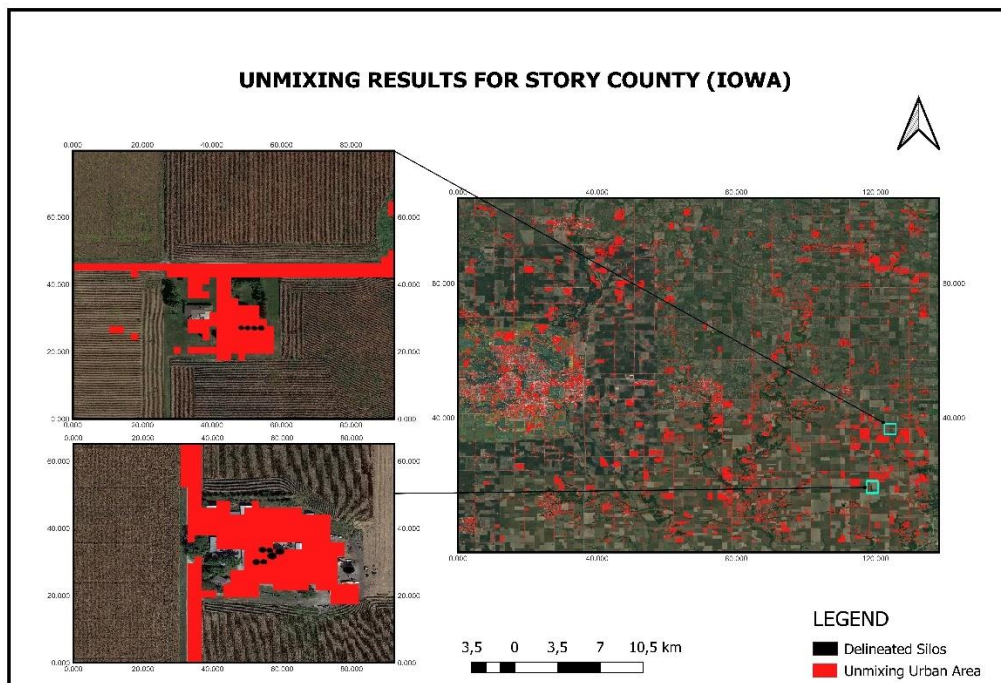


Figure 14: Shows the delineated Silo unmixing results for Story County of Iowa.

The Mower County located in Minnesota, and Story County, Iowa are the two selected counties where the unmixing was performed to identify built-up areas within agricultural lands with a specific focus on locating silos which are outside the 37 area. For Mower County, out of the 84 known silos in the area, the method successfully identified 55, yielding an accuracy rate of approximately 65.5%. However, 29 silos were not correctly identified, highlighting a significant limitation in the unmixing process. Story county analyzed 77 silos; 47 were correctly identified within built-up areas, whereas 30 were not. Additionally, in relation to Dewitt County, the detected urban areas also extend erroneously into vegetative regions, indicating a spectral

overlap that affects the accuracy of the unmixing results. This misclassification suggests that the spectral signatures of built-up areas and certain types of vegetation are not sufficiently distinct in Sentinel-2, which necessitate further refinement of the unmixing algorithm.

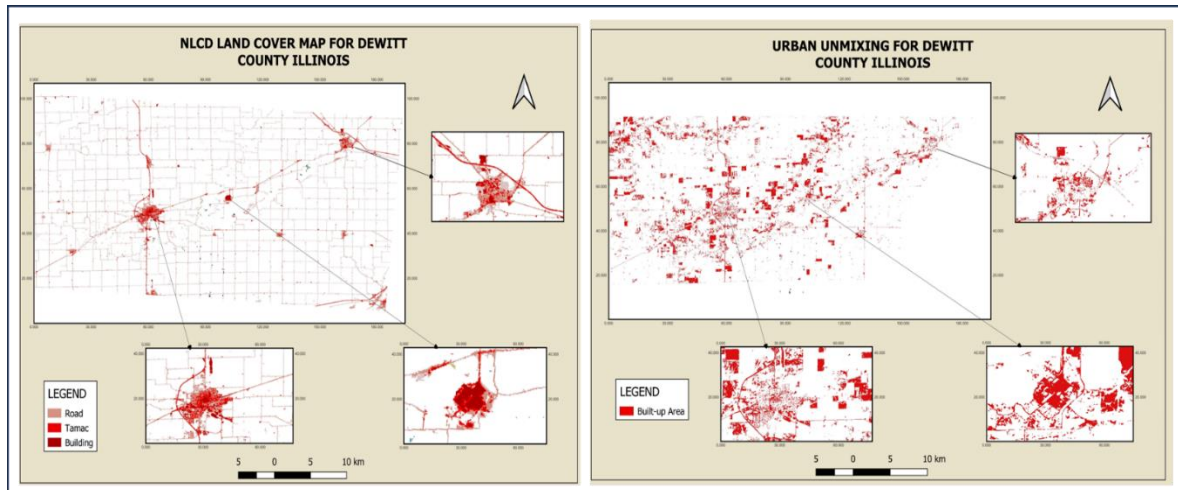


Figure 15: The comparison between the NLCD map and the unmixing result of Dewitt County, Illinois

Figure 15 to 17 show the comparison of unmixing results from three counties with the existing National Land Cover Database (NLCD) map. The NLCD map provides a comprehensive overview of major infrastructural elements such as roads and buildings, highlighting significant urban centers and transportation networks. However, the unmixing results, indicated in red, offer a similar granular view by detecting smaller and more dispersed built-up areas within agricultural lands, which the NLCD map also captures. This unmixing technique is crucial for identifying subtle patterns of urbanization and infrastructure development in rural settings. The added value of the unmixing results to the NLCD includes; its ability to distinguish between mixed pixels, identify more detailed land cover classes, and provide up-to-date temporal snapshots, making it particularly beneficial for specialized applications such as built-up area detection which can be applied to rural development and agricultural land management. This results in a more comprehensive understanding of land cover changes and patterns compared to the coarser resolution and fixed classes of the NLCD.

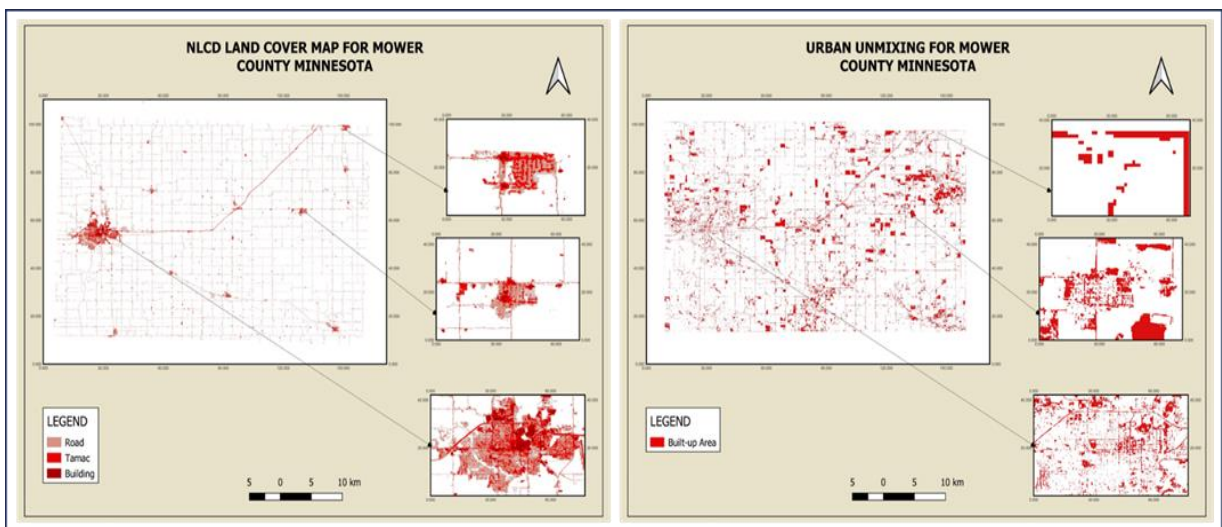


Figure 16: The comparison between the nlcd map and the unmixing result of Mower County, Minnesota

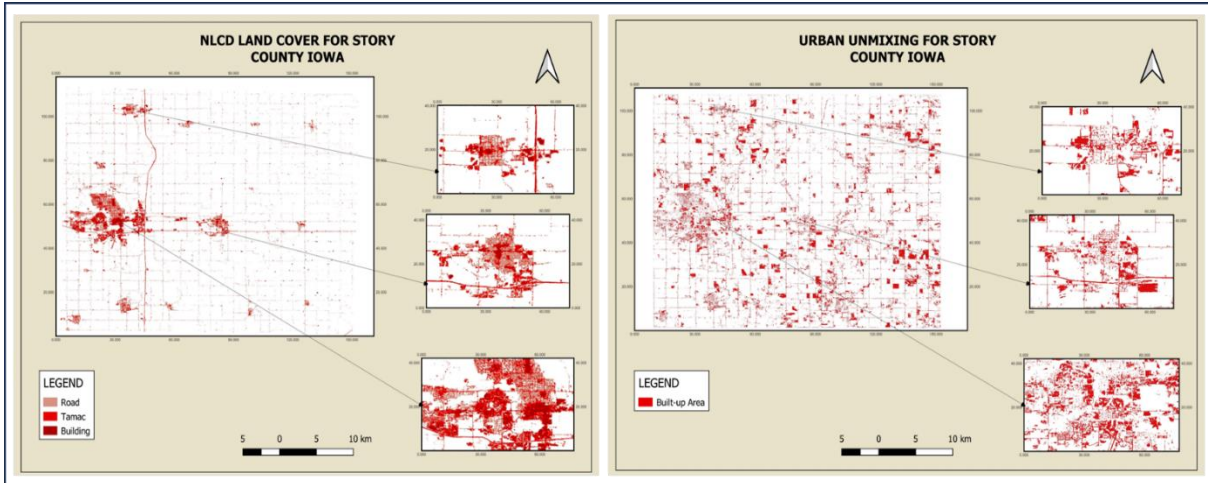


Figure 17: The comparison between the NLCD map and the unmixing result of Story County, Iowa.

The Unmixing was further applied to 37 random selected areas. These selected areas or the results of unmixing from these selected areas (built-up areas) are used as input for a refined classification using random forest and maximum likelihood classifier to disentangle silos from built-up areas.

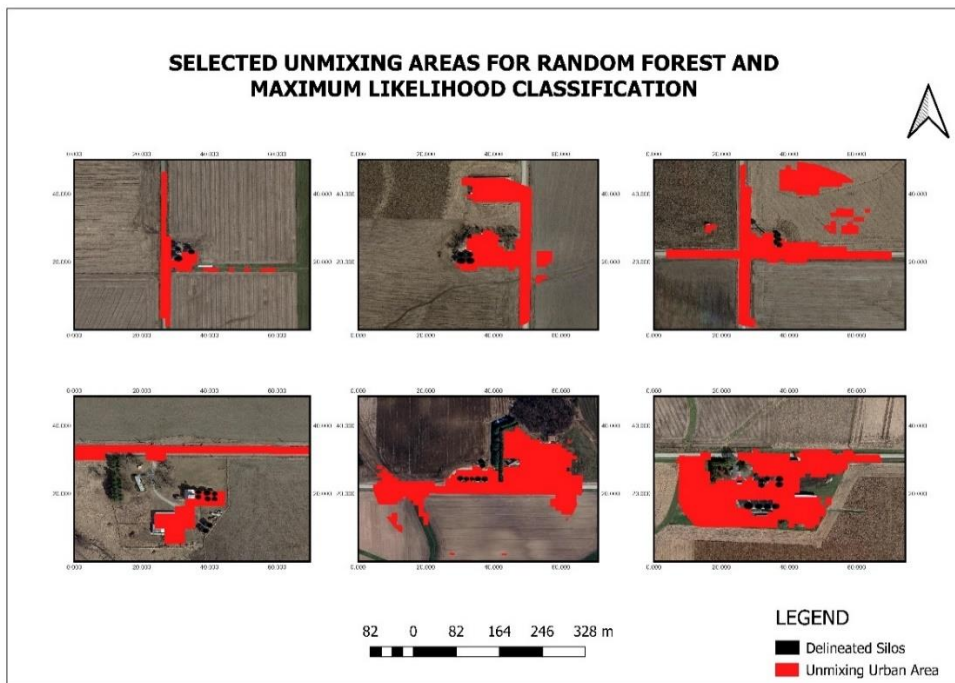


Figure 18: Six (6) out of the 37 selected unmixing areas for random forest and maximum likelihood classification.

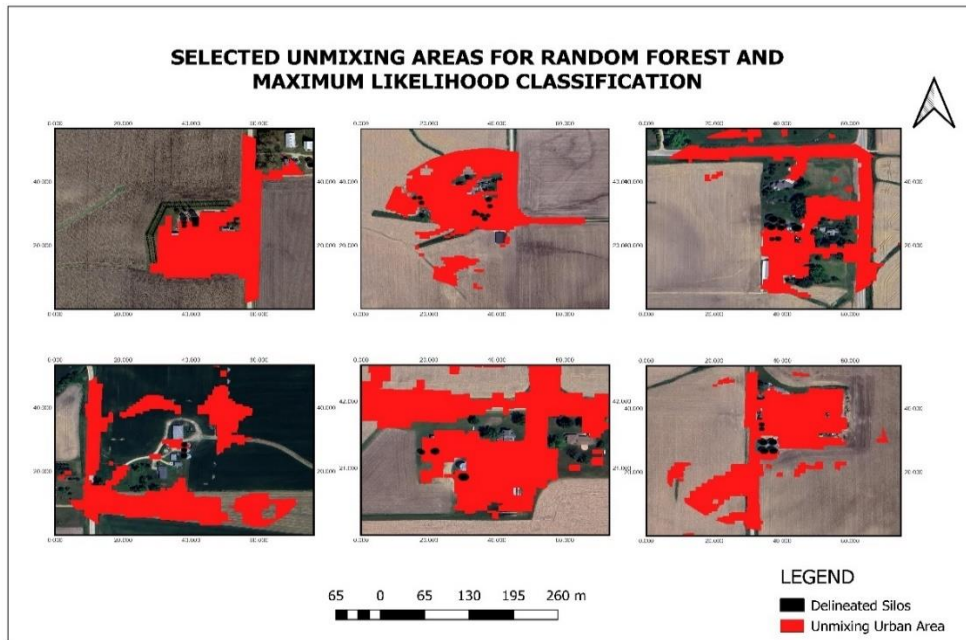


Figure 19: Six (6) out of the 37 selected unmixing areas for random forest and maximum likelihood classification.

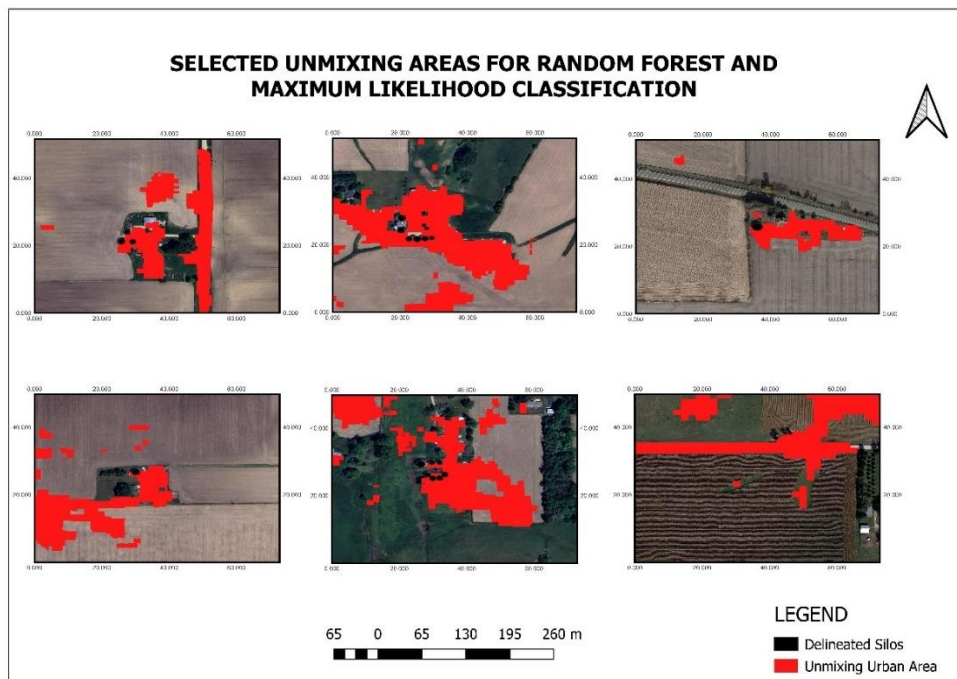


Figure 20: Six (6) out of the 37 selected unmixing areas for random forest and maximum likelihood classification.

Figure 18 through 20 show the results of the selected 37 areas (farm level) from the 200 meter buffer. This is displayed in a rectangular format and its different from the unmixing county level result.

Table 4: show the unmixing accuracy for all 37 areas.

S/N	True Positives	False Positives	Accuracy
1	249	241	50,82%
2	371	521	41,59%
3	116	148	43,94%
4	218	125	63,56%
5	302	422	41,71%
6	513	1255	29,02%
7	264	428	38,15%
8	468	483	49,21%
9	1104	31	97,27%
10	1036	105	90,80%
11	268	772	25,77%
12	82	223	26,89%
13	178	18	90,82%
14	293	828	26,14%
15	165	153	51,89%
16	532	187	73,99%
17	122	129	48,61%
18	83	65	56,08%
19	433	306	58,59%
20	79	92	46,20%
21	121	170	41,58%
22	170	89	65,64%
23	9	220	3,93%
24	64	39	62,14%
25	241	469	33,94%
26	305	20	93,85%
27	72	549	11,59%
28	354	904	28,14%
29	590	323	64,62%
30	21	540	3,74%
31	160	384	29,41%
32	13	166	7,26%
33	473	154	75,44%
34	1098	747	59,51%
35	539	490	52,38%
36	442	18	96,09%
37	838	1123	42,73%

3.3. Refining detection results from the unmixing approach by using Very High-Resolution imagery to identify silos in built-up areas.

3.3.1. Maximum Likelihood Classification

Figure 21 to 25 show the results of MLC as well as the geographical location of some of the testing regions.

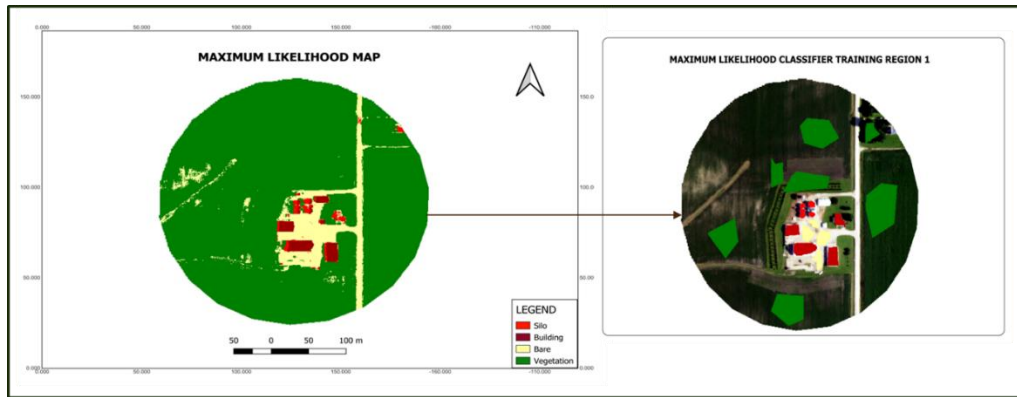


Figure 21: shows the Maximum likelihood classification and VHR image for the training 1.

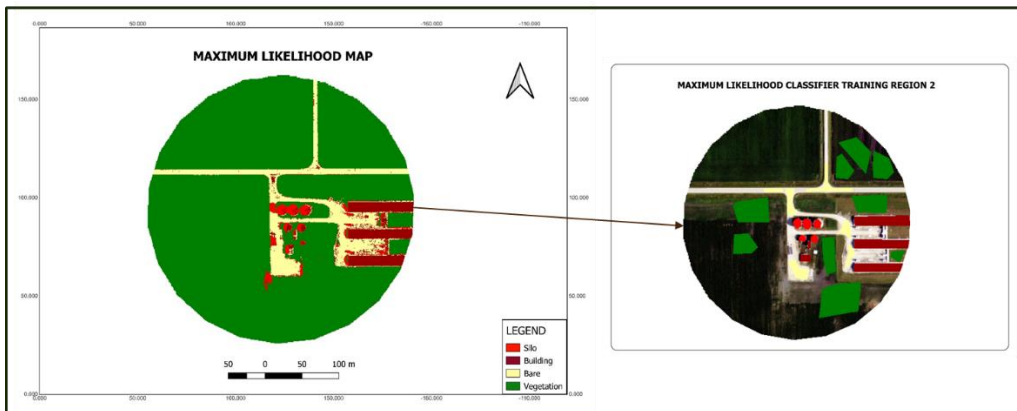


Figure 22: shows the Maximum likelihood classification and VHR image for training region 2.

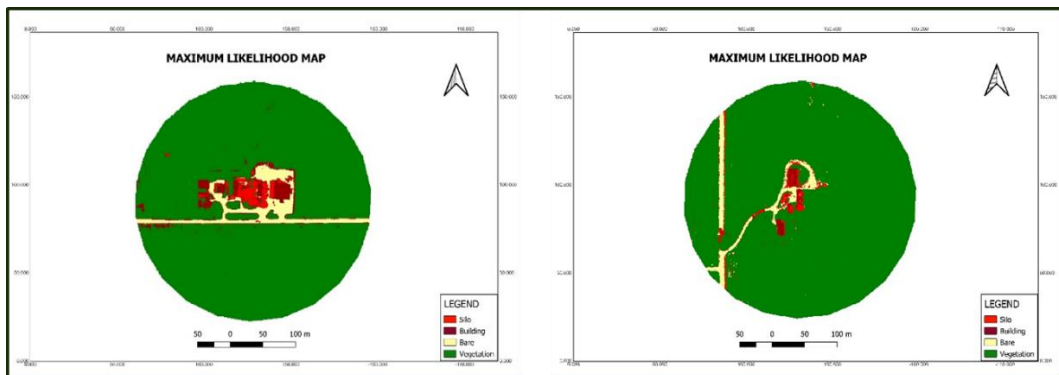


Figure 23: Maximum likelihood classification for testing region 1 and 2.

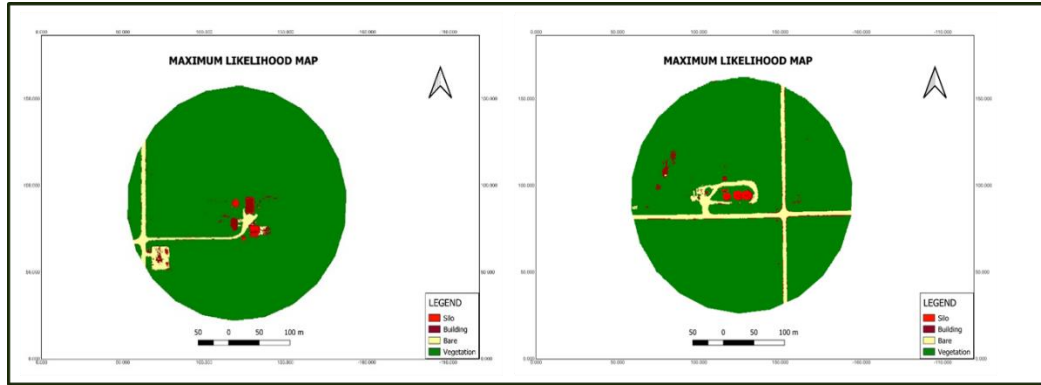


Figure 24: Maximum likelihood classification for testing region 3 and 4

Figures 24 to 26 present the outcomes of Maximum Likelihood Classification (MLC) applied to testing regions. The objective is to demonstrate the method's generalizability and its applicability across various geographic locations without the necessity for additional training samples specific to those areas. The results are favorable, illustrating that the method effectively delineates various land uses, particularly the silo locations, which are accurately identified and well-defined within the testing landscape. This suggests the robustness of MLC in adapting to diverse environments, making it a reliable tool for geographic analysis in varied settings.

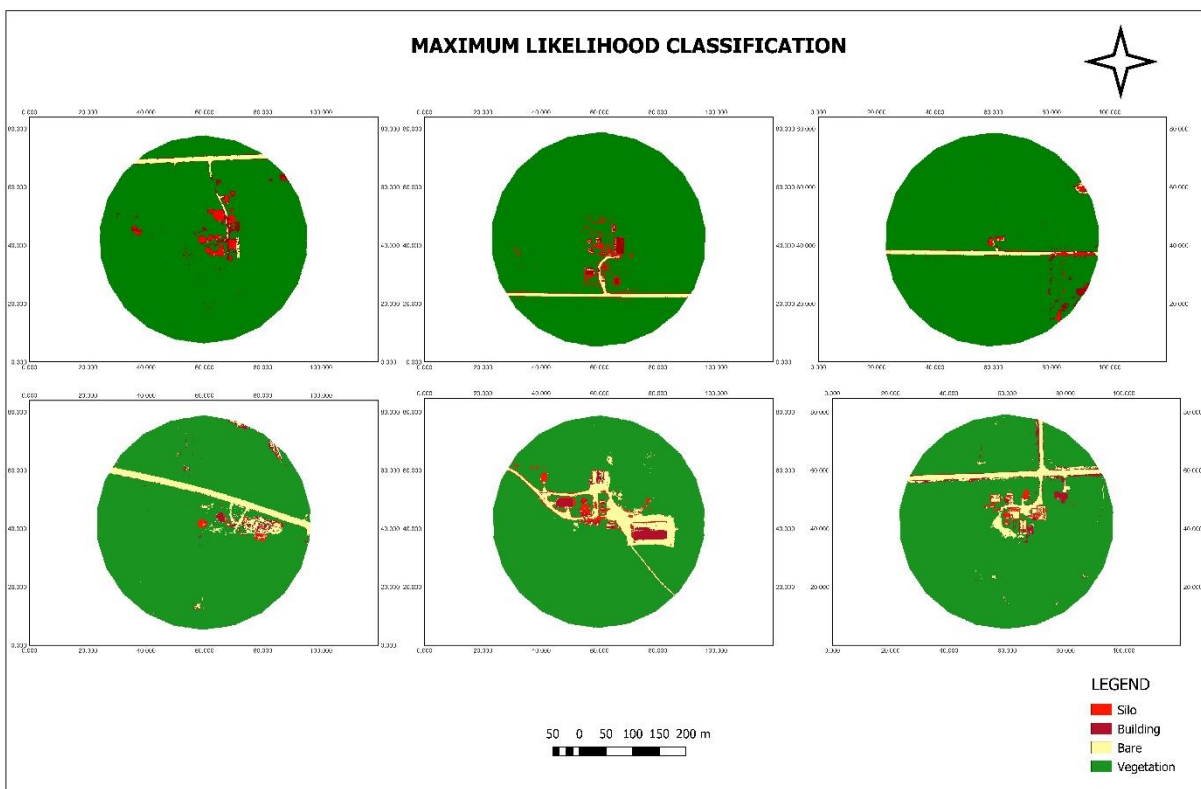


Figure 25: Maximum likelihood classification for some testing region

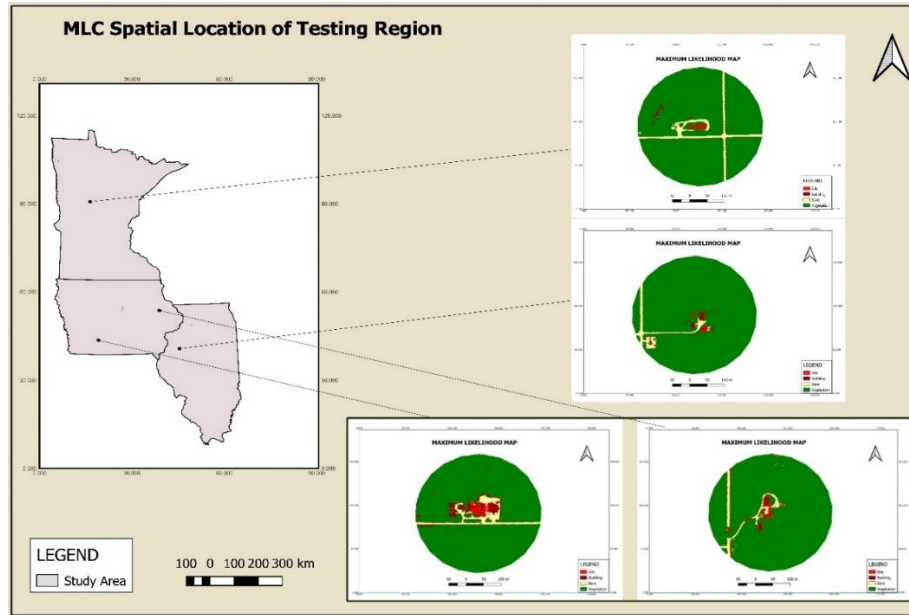


Figure 26. Spatial distribution of 4 out of 29 MLC testing regions.

Table 5: shows the for accuracy of the MLC.

S/N	REGION	OVERALL ACCURACY	SILO ACCURACY	BUILDING ACCURACY	BARE ACCURACY	VEGETATION ACCURACY
1	Area 1	84.71%	79.71%	80.71%	83.21%	85.71%
2	Area 2	85.71%	80.71%	81.71%	84.21%	86.71%
3	Area 3	85.16%	80.16%	81.16%	83.66%	86.16%
4	Area 4	84.97%	79.97%	80.97%	83.47%	85.97%
5	Area 5	79.96%	74.96%	75.96%	78.46%	80.96%
6	Area 6	79.89%	74.89%	75.89%	78.39%	80.89%
7	Area 7	82.88%	77.88%	78.88%	81.38%	83.88%
8	Area 8	82.28%	77.28%	78.28%	80.78%	83.28%
9	Area 9	74.16%	69.16%	70.16%	72.66%	75.16%
10	Area 10	74.87%	69.87%	70.87%	73.37%	75.87%
11	Area 11	85.36%	80.36%	81.36%	83.86%	86.36%
12	Area 12	83.00%	78.00%	79.00%	81.50%	84.00%
13	Area 13	77.50%	72.50%	73.50%	76.00%	78.50%
14	Area 14	80.04%	75.04%	76.04%	78.54%	81.04%
15	Area 15	80.00%	75.00%	76.00%	78.50%	81.00%
16	Area 16	81.31%	76.31%	77.31%	79.81%	82.31%
17	Area 17	84.24%	79.24%	80.24%	82.74%	85.24%
18	Area 18	82.07%	77.07%	78.07%	80.57%	83.07%
19	Area 19	75.33%	70.33%	71.33%	73.83%	76.33%
20	Area 20	79.78%	74.78%	75.78%	78.28%	80.78%
21	Area 21	76.25%	71.25%	72.25%	74.75%	77.25%
22	Area 22	80.00%	75.00%	76.00%	78.50%	81.00%
23	Area 23	76.32%	71.32%	72.32%	74.82%	77.32%
24	Area 24	68.36%	63.36%	64.36%	66.86%	69.36%
25	Area 25	65.31%	60.31%	61.31%	63.81%	66.31%
26	Area 26	69.04%	64.04%	65.04%	67.54%	70.04%
27	Area 27	74.59%	69.59%	70.59%	73.09%	75.59%
28	Area 28	81.27%	76.27%	77.27%	79.77%	82.27%
29	Area 29	78.88%	73.88%	74.88%	77.38%	79.88%

30	Area 30	68.09%	63.09%	64.09%	66.59%	69.09%
31	Area 31	77.48%	72.48%	73.48%	75.98%	78.48%
32	Area 32	67.43%	62.43%	63.43%	65.93%	68.43%
33	Area 33	82.05%	77.05%	78.05%	80.55%	83.05%
34	Area 34	81.33%	76.33%	77.33%	79.83%	82.33%
35	Area 35	83.99%	78.99%	79.99%	82.49%	84.99%
36	Area 36	60.21%	55.21%	56.21%	58.71%	61.21%
37	Area 37	65.45%	60.45%	61.45%	63.95%	66.45%

The table displays accuracy percentages for 37 regions, with the highest accuracy in Area 2 (training region) at 85.71% and the lowest in Area 36 (testing region) at 60.21%. Most regions have accuracies between 70% and 85%, with numerous areas exceeding 80%, such as Area 1, Area 3, and Area 11. However, several regions fall below 70%, including Area 24, Area 25, and Area 32, indicating areas that may need targeted improvements. However, the data highlights a significant variation in accuracy across regions, suggesting potential for enhancement in lower-performing areas. The lower accuracy in some testing areas (too many bare land and surrounding houses) points to potential issues with pixel misclassification and the influence of mixed pixels, especially between similar land cover types like silos and buildings.

3.3.2. Random Forest Classification

Table 6: it shows the number of correctly detected and missed detected silos as built-up in the 37 randomly selected regions spread across the three states.

State	Correct detection	Missed detection
Unmixing		
Iowa	50	26
Illinois	25	20
Minnesota	33	32
Random Forest		
Iowa	71	5
Illinois	35	10
Minnesota	58	7

The analysis encompassed the detection of silos within 37 built-up areas across Minnesota, Iowa, and Illinois using a test set of silos. In Minnesota, out of 65 silos, the unmixing process identified 33 within urban areas, with the Random Forest classifier correcting this to detect 58 correctly, though it missed 7 and misclassified 5 additional structures. Iowa's results from 76 total silos showed a strong detection performance, with unmixing identifying 50 within built-up areas and Random Forest improving this to 71, with only 5 misclassified. In Illinois, of 45 silos, 25 were detected within urban areas by unmixing, with Random Forest adjusting this figure to 35 correctly identified, though 10 were missed and some misclassifications occurred. These results suggest that the classification alone may lead to better results. However, it is computationally very heavy to apply it on the whole study area (Iowa, Illinois, and Minnesota), the introduction of the unmixing allows to sharply reduce the area to investigate but lead to low accuracy.

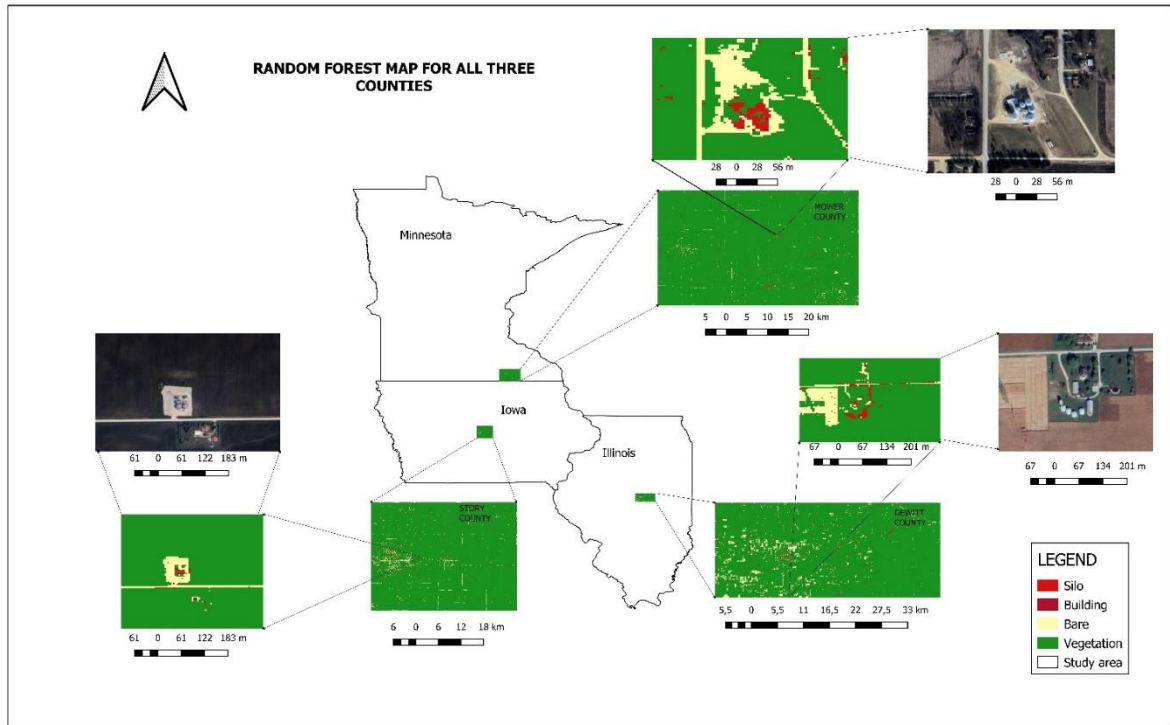


Figure 27: Example of random forest result for silo detection in the three counties

Figure 27 illustrates the application of the Random Forest algorithm at the county level. This figure highlights various urban areas, which are also identified through the unmixing process as shown in figure 28. Notably, the results demonstrate an improvement in the performance of the Random Forest algorithm in identifying these urban areas.

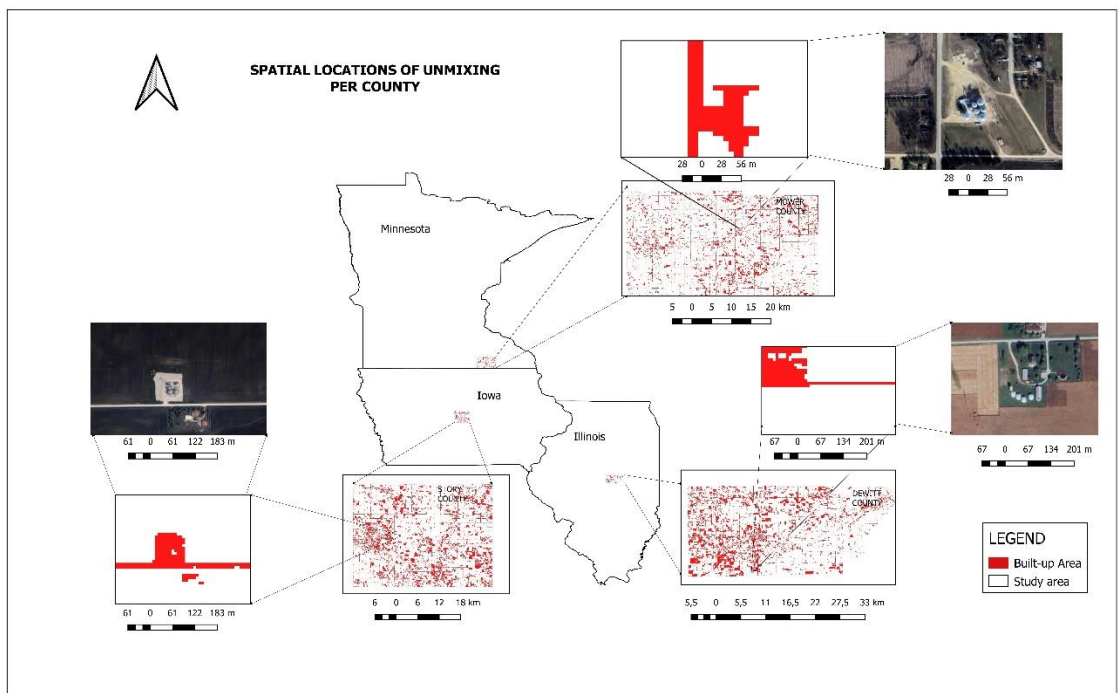


Figure 28. Spatial location of the 3 counties unmixing result

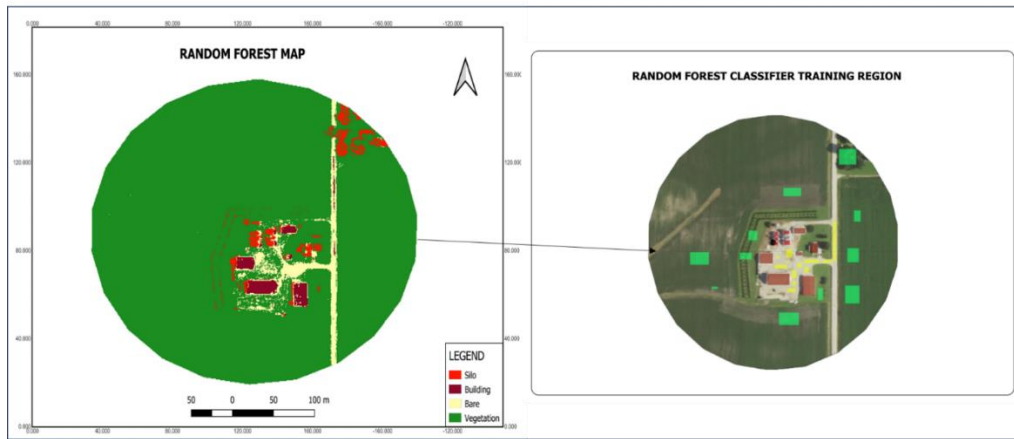


Figure 29: shows the Random Forest classification and VHR image for the training area.



Figure 30: shows the Random Forest classification training areas.

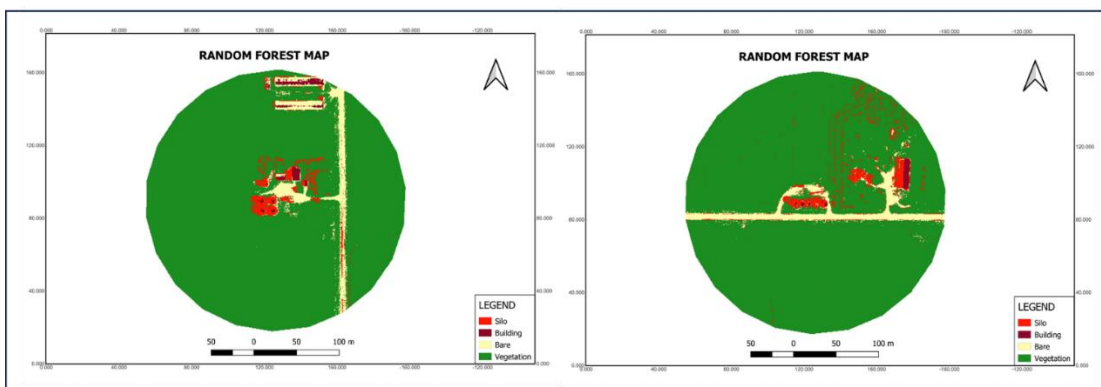


Figure 31: shows the Random Forest classification testing areas.

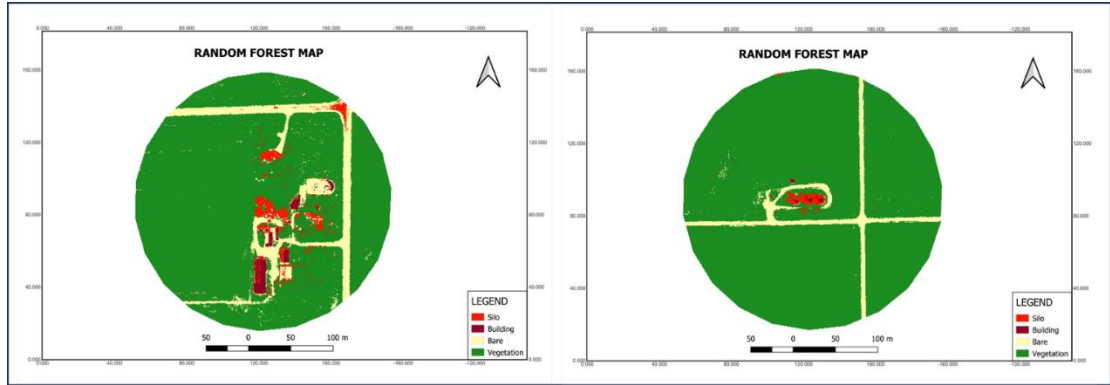


Figure 32: shows the Random Forest classification testing areas.

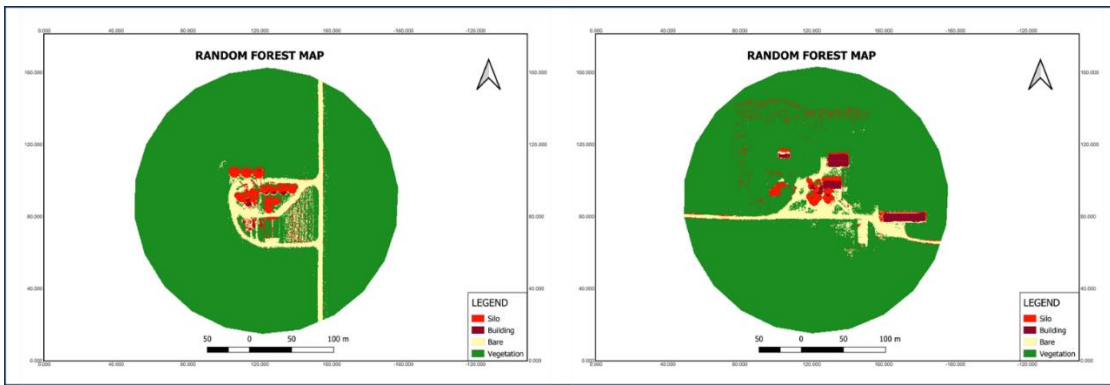


Figure 33: shows the Random Forest classification testing areas.

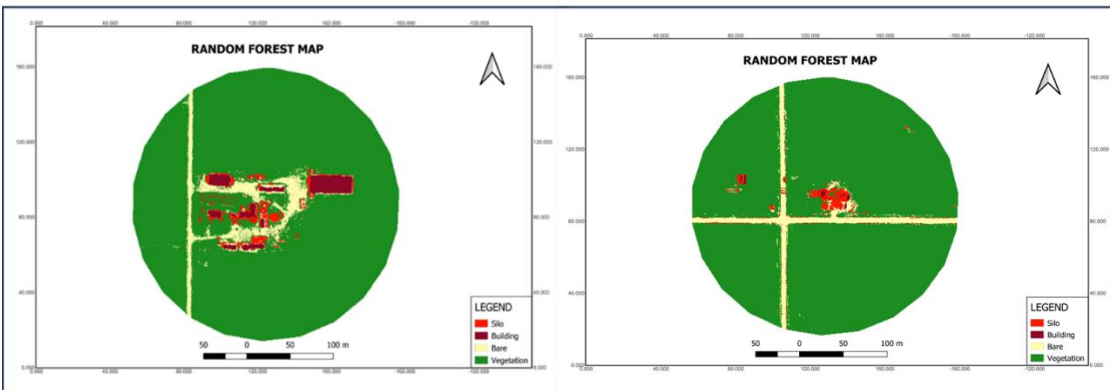


Figure 34: shows the Random Forest classification testing areas.

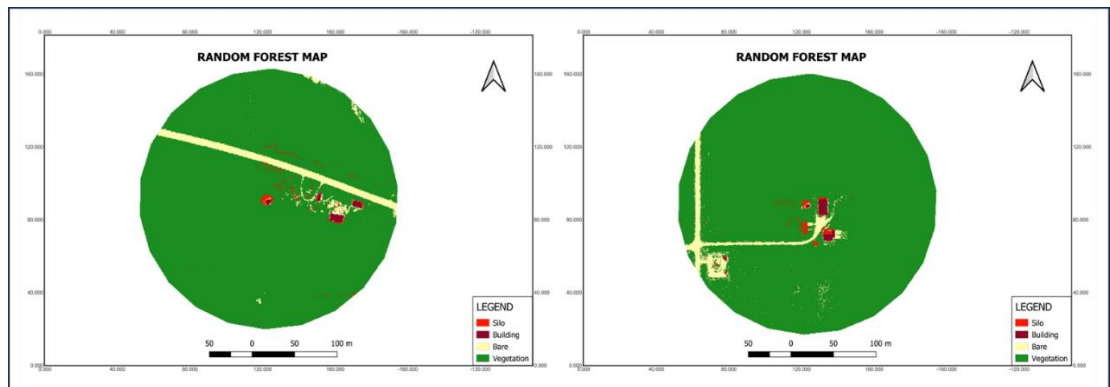


Figure 35: shows the Random Forest classification testing areas.

The Random Forest algorithm's performance was visually represented through a series of figures (Figures 29-35) that showcased the classification output on both training and testing areas. These figures illustrate how effectively the algorithm was able to discriminate between different land cover types within the high-resolution imagery.

Table 7: shows the accuracy of Random Forest.

S/N	REGION	OVERALL ACCURACY	SILO ACCURACY	BUILDING ACCURACY	BARE ACCURACY	VEGETATION ACCURACY
1	Area 1	83.99%	64.99%	70.99%	79.99%	86.99%
2	Area 2	84.55%	65.55%	71.55%	80.55%	87.55%
3	Area 3	85.74%	66.74%	72.74%	81.74%	88.74%
4	Area 4	83.04%	64.04%	70.04%	79.04%	86.04%
5	Area 5	82.39%	63.39%	69.39%	78.39%	85.39%
6	Area 6	81.90%	62.90%	68.90%	77.90%	84.90%
7	Area 7	81.55%	62.55%	68.55%	77.55%	84.55%
8	Area 8	81.77%	62.77%	68.77%	77.77%	84.77%
9	Area 9	81.95%	62.95%	68.95%	77.95%	84.95%
10	Area 10	83.68%	64.68%	70.68%	79.68%	86.68%
11	Area 11	76.63%	57.63%	63.63%	72.63%	79.63%
12	Area 12	77.51%	58.51%	64.51%	73.51%	80.51%
13	Area 13	80.70%	61.70%	67.70%	76.70%	83.70%
14	Area 14	80.63%	61.63%	67.63%	76.63%	83.63%
15	Area 15	77.48%	58.48%	64.48%	73.48%	80.48%
16	Area 16	79.70%	60.70%	66.70%	75.70%	82.70%
17	Area 17	74.51%	55.51%	61.51%	70.51%	77.51%
18	Area 18	74.28%	55.28%	61.28%	70.28%	77.28%
19	Area 19	80.78%	61.78%	67.78%	76.78%	83.78%
20	Area 20	76.63%	57.63%	63.63%	72.63%	79.63%
21	Area 21	81.23%	62.23%	68.23%	77.23%	84.23%
22	Area 22	75.44%	56.44%	62.44%	71.44%	78.44%
23	Area 23	68.00%	49.00%	55.00%	64.00%	71.00%
24	Area 24	81.82%	62.82%	68.82%	77.82%	84.82%
25	Area 25	80.44%	61.44%	67.44%	76.44%	83.44%
26	Area 26	81.61%	62.61%	68.61%	77.61%	84.61%
27	Area 27	82.11%	63.11%	69.11%	78.11%	85.11%
28	Area 28	81.27%	62.27%	68.27%	77.27%	84.27%
29	Area 29	81.68%	62.68%	68.68%	77.68%	84.68%
30	Area 30	77.69%	58.69%	64.69%	73.69%	80.69%
31	Area 31	74.39%	55.39%	61.39%	70.39%	77.39%
32	Area 32	81.39%	62.39%	68.39%	77.39%	84.39%
33	Area 33	80.86%	61.86%	67.86%	76.86%	83.86%
34	Area 34	77.66%	58.66%	64.66%	73.66%	80.66%
35	Area 35	80.27%	61.27%	67.27%	76.27%	83.27%
36	Area 36	71.65%	54.65%	60.65%	69.65%	76.65%
37	Area 37	75.34%	58.34%	64.34%	73.34%	80.34%

Table 7 presents the accuracy percentages of a Random Forest model across 37 regions, labeled as Area 1 through Area 37. The training region is from Area 1 to 8, while the testing region starts from area 9 till 38. The highest accuracy recorded is 85.74% in Area 3, while the lowest is 68.00% in Area 23. Most regions have accuracy above 75%, with several exceeding 80%, such as Areas 1, 2, 3, 4, 5, and 10. However, some areas, like Areas 17, 18, 23, and 36, have accuracies below 75%. This data highlights the varying performance of the model, with a noticeable concentration of high accuracies and a few regions with significantly lower performance.

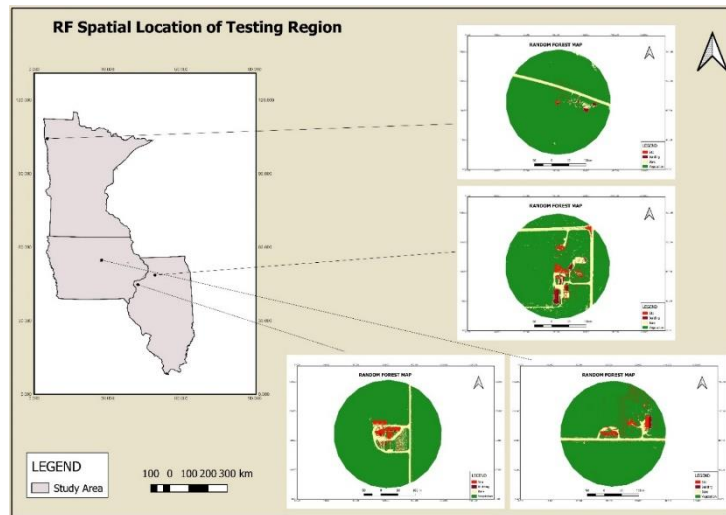


Figure 36. Spatial distribution of some the testing regions used for RF.

3.3.3. Comparison MLC and RF Results Base on Accuracy

In this section, the effectiveness of MLC and RF in classifying the silos will be presented and discussed.

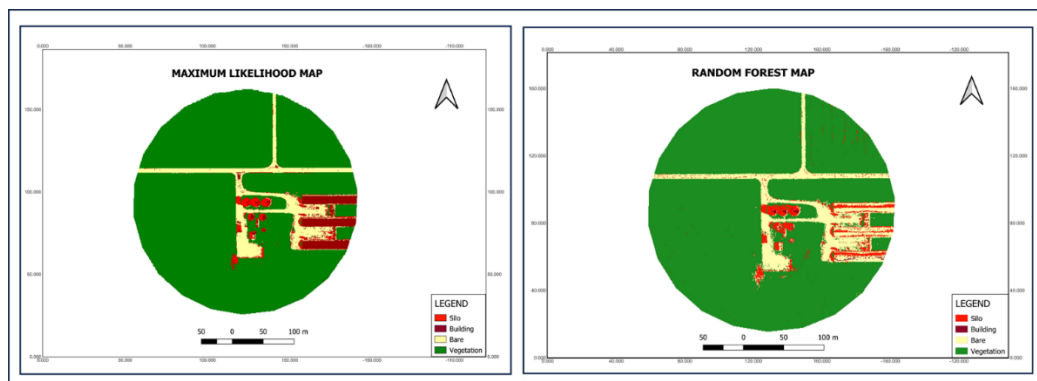


Figure 37: MLC map with an accuracy of 85.16% and RF map of the same area with the accuracy of 76.63%

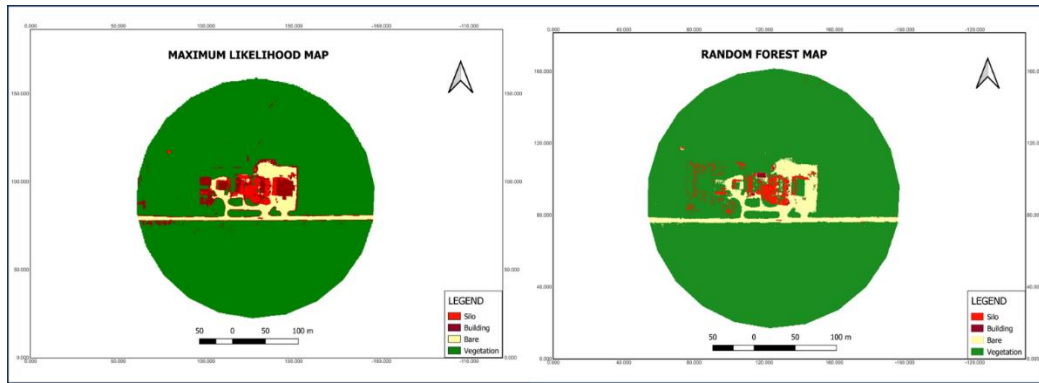


Figure 38. MLC map with an accuracy of 82.11% and RF map of the same area with the accuracy of 74.51%

Visually, the MLC maps display a more coherent and accurate representation of the classes (silo, bare, building, and vegetation) compared to the RF maps. The delineation of silos and buildings is notably clearer in the MLC outputs, with fewer misclassifications and mixed pixel issues, which are more prevalent in the RF maps. This visual clarity in MLC outputs is particularly evident in the areas surrounding the silos, where RF appears to struggle with distinguishing between buildings and silo structures due to similar reflectance properties.

The findings from the accuracy assessment and visual inspection indicate that MLC provides higher classification accuracy and better visual representation than RF. The superiority of MLC could be attributed to its statistical approach that probabilistically evaluates the likelihood of each pixel belonging to a particular class based on the training data, leading to more precise classification results. In contrast, the RF method, while robust and effective in many scenarios, may require change in raster image for classification, further tuning and more representative training data to achieve similar levels of precision and visual quality as MLC.

4. DISCUSSION

4.1. Mapping of Built-Up Areas Within Agricultural Land Using Unfusion Sentinel 2 Approach

The application of the unmixing was introduced to identify built-up areas around agricultural land which are also probable to be silo location. In this study (**Fig. 13**), we found that unmixing enhances Sentinel-2-based urban land cover mapping around agricultural land in county and farm level. However, the unmixing algorithm is not suitable for direct detection and delineation of silos. Also, it is not good enough in detecting built-up areas around the selected agricultural land despite the selection of the best season (spring) for its application and the moderation of threshold used. Other agriculture lands (forest, grasses) were misclassified with the urban band of the unmixing, making the accuracy lower as compared to RF and MLC (**Table 4**). It's important to note that unmixing downscaling results are less effective than other methods when handling areas with similar spectral properties (Xu & Somers, 2021). The results suggested that the use of RF and MLC can lead it higher accuracy, but it is computationally very heavy when applying it to the whole study area. This leads to the introduction of unmixing which allows to sharply reduce the area to investigate (built-up areas) but eventually leads to poor accuracy.

Random forest and MLC can refine the built-up classification, hence better single out silos compared to the unmixing approach, as shown by the results (**Table 5 and 7**).

The application of these classifiers to the selected areas demonstrated an improvement in the specificity of the classification, as shown in **Section 3.3**. These methods provided a deeper layer of analysis, which correctly identifies silos than the unmixing classifier. This enhances the discrimination between different urban materials and agricultural backgrounds.

4.2. Creating Silo Location Maps Using Very High-Resolution Imagery

4.2.1. Evaluating the Efficacy and Challenges of Maximum Likelihood Classification for Land Cover Analysis

The application of Maximum Likelihood Classification (MLC) on land cover classification demonstrated good performance across diverse environments (Liang et al., 2022). To achieve generalization of this research, the MLC was tested in regions that are independent from the training sets. (**Figure 21 and 22**). While the training phases showed high accuracy, testing phases across various areas revealed inconsistencies. These issues emphasize the complexities involved in applying MLC to areas beyond the scope of the training dataset. While the algorithm was performed on silos detection to refine the unmixing classification, it was met with several shortcomings. This includes the inability to classify the specific shapes of silos due to pixel misrepresentation. The spectral properties of some silos in this research are very identical to building. This often confuses the classifier in several instances and in turn leads to lower performance and accuracy (**Section 3.3.1, Table 5**). The study's findings suggest that while MLC is effective under specific conditions, its adaptability needs augmentation to ensure broader applicability. Otukey et al, (2010) confirm that Maximum Likelihood Classification (MLC) excels in familiar settings but struggles with spectral similarity in diverse environments, necessitating enhanced models for broader applicability. These variations highlight the MLC's challenges in universally applying the model to unfamiliar environments.

4.2.2. Effectiveness of Random Forest Classification of Silo and Other Properties Using Google Earth Engine

RF has played a pivotal role in getting a better classification accuracy by using the VHR imagery over the years (Hayes et al., 2014). However, RF has yet to be fully explore in remote sensing community, which includes the detection of circular structures like silos (Kulkarni et al, 2016). It was found that the NAIP imagery (1 meter resolution) which is available for free in GEE can provide a good accuracy with the help of proper time filtering. The image exists from 2007 till date, however, time ranges before 2013 provide a poor classification due to poor image quality.

Like the MLC, the RF also classified vegetation class correctly, while other built-up areas like bare, silo and building were misclassified due to their similar spectral properties. Unlike other VHR, the NAIP has a very poor spectral resolution, and this study was not able to create alternatives for this. The RF give a moderately high accuracy in all 37 regions (**Table 7**), with the training region having a higher accuracy than the testing region. This precision underlines the effectiveness of combining point and polygon data types, enhancing the model's capability to generalize across diverse spatial scales and environmental conditions. Nevertheless, applying the Random Forest algorithm to detect silos in Minnesota, Iowa, and Illinois provides a nuanced view of the algorithm's potential and its limitations. The RF Has also do well to identify silos correct in the selected states as compared to the unmixing but is not the same when compared to MLC (**Table 5**). These findings illustrate significant enhancements in detection accuracy, yet they also reveal the challenges of consistent performance across varied landscapes.

4.3. Scaling Up the Approach for Large-Scale Detection

Scaling up this research approach to automatically detect built-up areas within agricultural lands on a large scale presents several challenges. Firstly, processing high-resolution imagery over extensive geographic areas requires significant computational power. Handling large datasets from Sentinel-2 or NAIP imagery demands robust computational infrastructure, such as high-performance servers or cloud-based solutions. Another challenge lies in the algorithm's ability to adapt to different landscapes and spectral characteristics. The spectral signatures of built-up areas, silos, and vegetation can vary greatly across different regions due to factors like climate, soil type, and agricultural practices. To ensure the method works well on a larger scale, we need to improve the algorithm's generalization capabilities. This can be done by expanding the training dataset to cover a broader range of geographic areas and land cover types, exposing the model to various scenarios during training. Additionally, developing adaptive algorithms that can adjust to local spectral variations will enhance accuracy and robustness when applied to large-scale applications.

5. LIMITATIONS, RECOMMENDATIONS AND CONCLUSIONS

5.1. Limitations

The primary limitation of the unmixing technique is its difficulty in accurately distinguishing between built-up areas and vegetative cover, particularly due to seasonal changes in the images that affect the temporal properties of these areas. Additionally, the unmixing threshold, set at 0.75, yields better classification in some regions but performs poorly in other areas due to factors like the density of buildings and types of vegetation. This misclassification often results in transition of built-up areas to vegetative cover, as seen in the analysis of DeWitt, Mower, and Story County, where urban areas transit into vegetative regions. Additionally, the resolution of Sentinel-2 imagery, while high, may not be sufficient to distinguish small and dispersed structures like silos from other built-up features. The accuracy rates for identifying silos in DeWitt, Mower, and Story Counties indicate significant room for improvement. Another limitation is the dependency on the quality and timing of the satellite imagery; seasonal variations can affect the visibility and distinguishability of different land cover types. These limitations underscore the need for algorithm refinement, data integration, and continuous validation through ground-truthing exercises to improve the unmixing technique's performance. There is also some level of inconsistencies in MLC and RF Performance. While MLC displayed higher accuracy in controlled environments, its performance varied significantly under testing conditions, indicating challenges in generalizing the model. Similarly, RF showed promise but was inconsistent across different terrains and classes, which highlights the potential issues in model robustness and training data representativeness. Lastly, there is spectral limitation in the NAIP imagery: The NAIP imagery used has a poor spectral resolution leading to class misrepresentation.

5.2. Recommendations

To improve the unmixing algorithm for large-scale detection of built-up areas within agricultural lands, several enhancements are recommended. Firstly, refining the algorithm to better differentiate between built-up areas and vegetative cover is crucial, potentially through advanced spectral unmixing techniques or incorporating machine learning models that can distinguish subtle spectral variations. Integrating additional data sources like LiDAR and multispectral imagery can provide structural information and more detailed spectral data, improving accuracy. Further seasonal analysis, particularly in spring, can optimize timing due to better visibility and distinguishability of land cover types. Regular ground-truthing exercises are essential for validating remote sensing results and refining detection methods, addressing discrepancies and misclassification issues. Finally, exploring sophisticated classifiers like deep learning and ensemble methods can handle complex data patterns, enhancing the identification of built-up areas and silos, reducing misclassification rates, and improving overall accuracy.

5.3. Conclusion

This research aimed to identify silos at large scale while balancing accuracy and computational burden. Although, VHR data can lead to accurate silo detection, but applying classifiers at the state or county level is computationally intensive. To mitigate this, the research use Sentinel-2 data and unsupervised methods to initially identify areas where buildings are surrounded by agricultural land, thus narrowing the focus to those regions and reducing computational demands.

This research has explored the effectiveness and challenges of various classification techniques and silo detection methodologies using Sentinel-2 imagery, NAIP imagery, unmixing approach, MLC, and RF. Each method has demonstrated strengths in certain contexts but also faces significant limitations that can impact their practical application. Sentinel-2 imagery provides broad coverage and high temporal resolution but struggles with spectral resolution issues that can lead to misclassification. NAIP imagery offers high spatial resolution, enhancing the identification of small-scale features like silos, but lacks the spectral depth of Sentinel-2. MLC, with its statistical approach, excels in well-defined, homogenous areas but falters in diverse landscapes with mixed pixels. RF is robust and handles complex data patterns effectively yet requires substantial training data and computational resources for large-scale application.

The findings underscore the need for enhancing image processing techniques to preserve critical image features, which is essential for accurate classification and detection. Techniques that can maintain the integrity of spectral and spatial information during preprocessing are crucial. Additionally, refining algorithm calibration is necessary to ensure robust performance across different preprocessing outputs and varying environmental conditions. Ultimately, the study demonstrates that while the unmixing technique is promising, further refinement and integration with advanced classifiers and additional data sources are necessary to achieve higher accuracy and reliability.

6. ETHICAL CONSIDERATION

This research utilizes open-source geospatial data for areas within the United States such as Iowa, Minnesota, and Illinois, which does not have any ethical concern. The study adheres strictly to the ethical principles and guidelines outlined by the University of Twente's Research Ethics Policy, which emphasized on the use and handling of data. The ethical considerations of this thesis focus primarily on the responsible use of machine learning algorithms, ensuring that these tools do not perpetuate existing biases and that their application is transparent and accountable. Rigorous validation processes, including comparisons with ground-truth data, are employed to confirm the accuracy and reliability of the algorithms. By maintaining high ethical standards in these areas, the research aims to contribute positively to the scientific community while upholding the integrity of the data and respecting societal norms.

LIST OF REFERENCES

- Ali, M. Z., Qazi, W., & Aslam, N. (2018). A comparative study of ALOS-2 PALSAR and landsat-8 imagery for land cover classification using maximum likelihood classifier. *Egyptian Journal of Remote Sensing and Space Science*, 21, S29–S35. <https://doi.org/10.1016/j.ejrs.2018.03.003>
- Arends-Kuenning, M., Garcias, M., Kamei, A., Shikida, P. F. A., & Romani, G. E. (2022). Factors associated with harvest and postharvest loss among soybean farmers in Western Paraná State, Brazil. *Food Policy*, 112. <https://doi.org/10.1016/j.foodpol.2022.102363>
- Chen, G., Hou, J., & Liu, C. (2022). A Scientometric Review of Grain Storage Technology in the Past 15 Years (2007–2022) Based on Knowledge Graph and Visualization. *Foods*, 11(23). <https://doi.org/10.3390/foods11233836>
- FAO. (2009). *How to Feed the World in 2050*. http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf
- Grain Silos and Storage System Global Market Report. (2023). *Grain Silos and Storage System Global Market Report, Forecast 2032* <https://www.thebusinessresearchcompany.com/report/grain-silos-and-storage-system-global-market-report>. <https://www.facebook.com/>
- Hayes, M. M., Miller, S. N., & Murphy, M. A. (2014). High-resolution landcover classification using random forest. *Remote Sensing Letters*, 5(2), 112–121. <https://doi.org/10.1080/2150704X.2014.882526>
- International Institute of Information Technology (Pune, I., Institute of Electrical and Electronics Engineers. Pune Section, & Institute of Electrical and Electronics Engineers. (n.d.). *International Conference on Automatic Control & Dynamic Optimization Techniques (ICACDOT 2016) : 9th & 10th September 2016*.
- Kulkarni et al. (2016). *Random Forest Algorithm for Land Cover Classification - Random Forest Algorithm for Land Cover Classification*.
- Kumar, D., & Kalita, P. (2017). Reducing postharvest losses during storage of grain crops to strengthen food security in developing countries. *Foods*, 6(1), 1–22. <https://doi.org/10.3390/foods6010008>
- Liang, F., Zhang, X., Li, H., Yu, H., Lin, Q., Jiang, M., & Zhang, J. (2022). Land Use Classification Based on Maximum Likelihood Method. *Smart Innovation, Systems and Technologies*, 253, 133–139. https://doi.org/10.1007/978-981-16-5036-9_15
- Mishra, V. N., Prasad, R., Kumar, P., Gupta, D. K., & Srivastava, P. K. (2017). Dual-polarimetric C-band SAR data for land use/land cover classification by incorporating textural information. *Environmental Earth Sciences*, 76(1). <https://doi.org/10.1007/s12665-016-6341-7>
- Otukei et al. (2010). *Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms - 1-s2.0-S0303243409001135-main*.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Prusky, D. (2011). Reduction of the incidence of postharvest quality losses, and future prospects. *Food Security*, 3(4), 463–474. <https://doi.org/10.1007/s12571-011-0147-y>
- Raut, R. D., Gardas, B. B., Kharat, M., & Narkhede, B. (2018). Modeling the drivers of post-harvest losses – MCDM approach. *Computers and Electronics in Agriculture*, 154, 426–433. <https://doi.org/10.1016/j.compag.2018.09.035>
- Rosalia et al. (2019). *The contribution of Urban Food Policies toward food security in developing and developed countries_ A network analysis approach - 1-s2.0-S2210670718323576-main*.

- Sartori et al. (2015). *Connected we stand_ A network perspective on trade and global food security - 1-s2.0-S0306919215001207-main*.
- Tadros, A., Drouyer, S., Von Gioi, R. G., & Carvalho, L. (2020). Oil Tank Detection in Satellite Images via a Contrario Clustering. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2233–2236. <https://doi.org/10.1109/IGARSS39084.2020.9323249>
- Tefera, T. (2012a). Post-harvest losses in African maize in the face of increasing food shortage. *Food Security*, 4(2), 267–277. <https://doi.org/10.1007/s12571-012-0182-3>
- Tefera, T. (2012b). Post-harvest losses in African maize in the face of increasing food shortage. *Food Security*, 4(2), 267–277. <https://doi.org/10.1007/s12571-012-0182-3>
- Tefera, T., Kanampiu, F., De Groot, H., Hellin, J., Mugo, S., Kimenju, S., Beyene, Y., Boddupalli, P. M., Shiferaw, B., & Banziger, M. (2011). The metal silo: An effective grain storage technology for reducing post-harvest insect and pathogen losses in maize while improving smallholder farmers' food security in developing countries. In *Crop Protection* (Vol. 30, Issue 3, pp. 240–245). <https://doi.org/10.1016/j.cropro.2010.11.015>
- Tong, X. Y., Xia, G. S., & Zhu, X. X. (2023). Enabling country-scale land cover mapping with meter-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 178–196. <https://doi.org/10.1016/j.isprsjprs.2022.12.011>
- United States Department of Agriculture (USDA). (2023). *Grain Stocks*.
- Vogt, M., & Gerding, M. (2017). Silo and Tank Vision: Applications, Challenges, and Technical Solutions for Radar Measurement of Liquids and Bulk Solids in Tanks and Silos. *IEEE Microwave Magazine*, 18(6), 38–51. <https://doi.org/10.1109/MMM.2017.2711978>
- Wang, W., Zhao, D., & Jiang, Z. (2018). Oil tank detection via target-driven learning saliency model. *Proceedings - 4th Asian Conference on Pattern Recognition, ACPR 2017*, 126–131. <https://doi.org/10.1109/ACPR.2017.70>
- Xia, X., RongFeng, Y., & Kun, Y. (n.d.). *Oil Tank Extraction in High-resolution Remote Sensing Images based on Deep Learning*.
- Xu, F., & Somers, B. (2021). Unmixing-based Sentinel-2 downscaling for urban land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, 133–154. <https://doi.org/10.1016/j.isprsjprs.2020.11.009>

7. APPENDIX

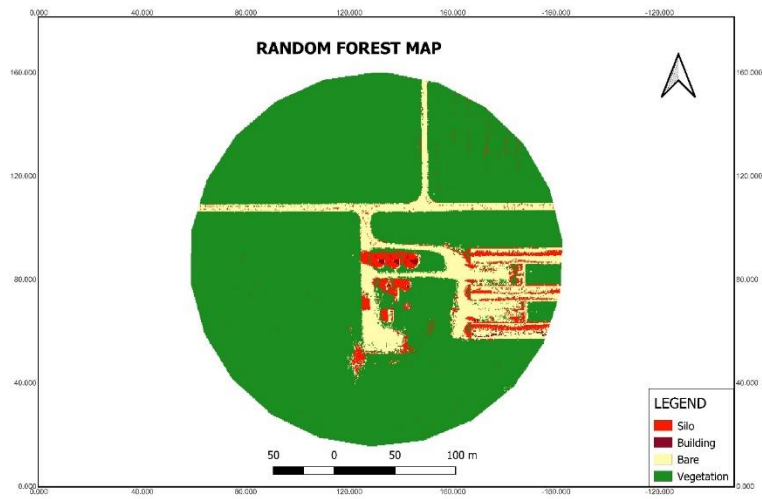


Figure 39: shows the Random Forest classification testing areas.

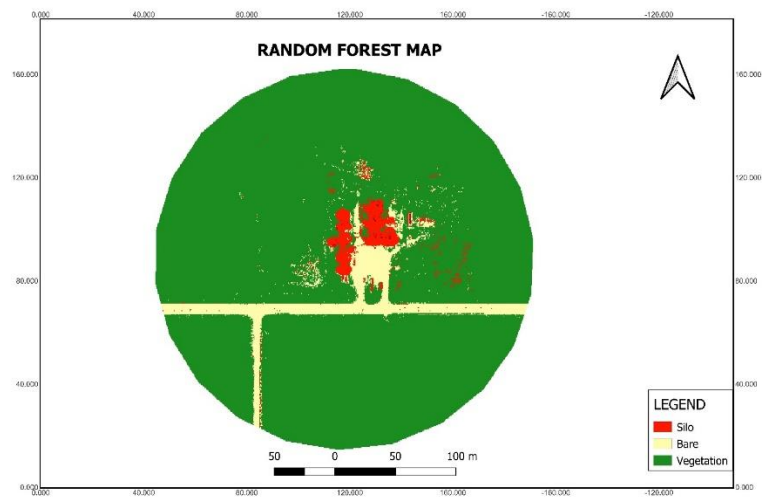


Figure 40: shows the Random Forest classification testing areas.

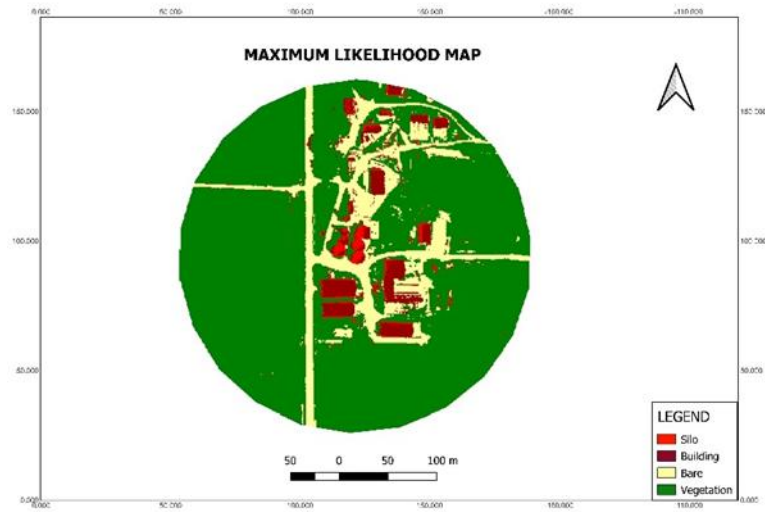


Figure 39: shows the Random Forest classification testing areas.

Table 8: the quantitative results for all four seasons of the 37 selected areas

ACCURACY PER SEASON				
	SPRING	SUMMER	FALL	WINTER
AREA 1	94,00%	50,00%	50,00%	50,00%
AREA 2	74,00%	59,00%	11,00%	52,00%
AREA 3	86,00%	93,00%	43,00%	36,00%
AREA 4	76,00%	76,00%	56,00%	64,00%
AREA 5	73,00%	41,00%	27,00%	55,00%
AREA 6	73,00%	73,00%	53,00%	47,00%
AREA 7	80,00%	92,00%	60,00%	80,00%
AREA 8	72,00%	39,00%	11,00%	87,00%
AREA 9	67,00%	79,00%	43,00%	33,00%
AREA 10	70,00%	50,00%	61,00%	48,00%
AREA 11	100,00%	50,00%	56,00%	0,00%
AREA 12	37,00%	58,00%	89,00%	16,00%
AREA 13	100,00%	100,00%	0,00%	0,00%
AREA 14	89,00%	89,00%	22,00%	0,00%
AREA 15	78,00%	98,00%	56,00%	38,00%
AREA 16	79,00%	89,00%	72,00%	68,00%
AREA 17	100,00%	90,00%	75,00%	30,00%
AREA 18	65,00%	4,00%	74,00%	65,00%
AREA 19	100,00%	50,00%	0,00%	50,00%
AREA 20	100,00%	90,00%	90,00%	23,00%
AREA 21	50,00%	87,00%	50,00%	50,00%
AREA 22	88,00%	87,00%	33,33%	83,00%
AREA 23	83,33%	83,33%	83,33%	66,67%
AREA 24	90,00%	70,00%	10,00%	40,00%

AREA 25	68,57%	42,86%	0,00%	28,57%
AREA 26	16,67%	88,89%	33,33%	44,44%
AREA 27	57,14%	100,00%	0,00%	0,00%
AREA 28	50,00%	80,00%	20,00%	10,00%
AREA 29	42,11%	94,74%	52,63%	42,11%
AREA 30	100,00%	78,00%	100,00%	75,00%
AREA 31	84,62%	92,00%	100,00%	84,62%
AREA 32	50,00%	75,00%	25,00%	0,00%
AREA 33	100,00%	50,00%	100,00%	100,00%
AREA 34	53,85%	53,85%	53,85%	61,54%
AREA 35	40,00%	80,00%	60,00%	60,00%
AREA 36	90,00%	67,00%	60,00%	80,00%
AREA 37	98,00%	50,00%	50,00%	0,00%
AVERAGE	75,01	0,7161	0,4812	0,4508

STANDARD DEVIATION

	0,2138	0,2203	0,298	0,2846
--	--------	--------	-------	--------
