Evaluating Machine Learning Algorithms for Onion Mapping in Nueva Ecija, Philippines Using Sentinel-2 Imagery

Reymar R. Diwa[1], Ariel C. Blanco[1,2]

[1] Space Information and Infrastructure Bureau, Philippine Space Agency, Quezon City 1109, Philippines – reymar.diwa@philsa.gov.ph

[2] Department of Geodetic Engineering, College of Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines – acblanco@up.edu.ph

Keywords: machine learning, onion mapping, CatBoost, LightGBM, XGBoost

Abstract

High-value crops like onion are vulnerable to price fluctuations for several reasons, including production shortage, infestation, inflation, importation-related issues, and climate impacts, resulting in high risk for local farmers. Accurate mapping and monitoring can be invaluable in managing these price fluctuations and ensuring long-term stable supply chains, as they enable detailed crop monitoring and yield estimation for onions. In this work, we utilized Sentinel-2 multispectral imagery for onion mapping, applying machine learning algorithms (MLAs) such as Extreme Gradient Boosting Machine (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost Classifier. The input data for the analysis included the 10 RGB, VRE, NIR, and SWIR bands of Sentinel-2 as well as 25 biophysical indices and terrain variables. These indices encompass key indicators for monitoring crop health and suitability like overall vegetation health, chlorophyll content, nitrogen content, soil moisture, soil salinity, soil clay content, Leaf Area Index (LAI), etc. The results showed that among the MLAs tested, CatBoost achieved the highest accuracy (90.0 %), followed by LightGBM (86.7 %), and XGBoost (84.7 %). Among the bands and indices used, the Clay Minerals Ratio (CMR) and Modified Photochemical Reflectance Index (PRI) were consistently identified as the most important features, strongly suggesting that onions are distinguished based on a combination of soil properties and canopy pigment traits.

1. Introduction

Onion (*Allium cepa* L.) is among the most cultivated crops in the world (Opara, 2003), with global production reaching more than 70 million tons, mostly produced by China and India (Pareek et al., 2017). In the Philippines, onion production grew more than twofold to 283,000 metric tons between 2011 and 2022. However, the crop remains highly vulnerable to price volatility, mainly because of several reasons including production shortfalls, extreme weather events, pest infestations, inflation, importation issues, and climate change impacts (Hutchinson, 2023). Monitoring onion cultivation areas is therefore critical to support government interventions aimed at stabilizing the market and ensuring a reliable supply which is a persistent challenge in the country.

Conventional field surveys remain the primary method for monitoring of onion farming areas and production in the Philippines (Philippine Statistics Authority, n.d.). However, these surveys are time-consuming, labor-intensive, prone to sampling bias, and are often subject to logistical constraints (Liu et al., 2018). Remote sensing (RS) provides a scalable approach to crop identification and monitoring, minimizing dependence on field-based surveys (Doraiswamy et al., 2005). In fact, RS serves as an efficient tool for improving soil and crop productivity (Kingra et al., 2016).

RS is an effective tool for monitoring cropping practices (Bégué et al., 2018). Numerous works used RS for crop-type mapping and yield prediction (Joshi et al., 2023; Karthikeyan et al., 2020; Moran et al., 1997; Navid & Vieira, 2019). Although garlic, a similar-looking crop, has been accurately identified and mapped using combined Sentinel-1 and Sentinel-2 in a previous work (Y. Chen et al., 2022), onion mapping using RS remains limited despite its very huge potential. Nevertheless, in a previous work by Diwa et al. (2024), hyperspectral imagery from the PRISMA satellite was used to map onion fields in Bongabon, Nueva Ecija - the onion capital of the Philippines - achieving a 79 % accuracy using linear spectral unmixing (LSU). Hyperspectral imagery are generally more advantageous for differentiating crop types than multispectral images because of its high spectral resolution (Mariotto et al., 2013), enabling the differentiation of onion from other landcover types. However, the study also revealed two major limitations: the difficulty to obtain cloud-free hyperspectral images during the crucial onion growth period and the relatively coarse spatial resolution (30 m), which is insufficient for mapping of smallholder and fragmented farm parcels, a significant concern given that 38 % of farms in the Philippines are smaller than 0.5 hectares (ha) (SEARCA, n.d.).

This study addresses these limitations by leveraging open-source Sentinel-2 imagery (S2), which offers more frequent revisit time (i.e., 5 days at the equator) and higher spatial resolution (10–20 m). Like the PRISMA, S2 also captures data in the visible, near-infrared, and shortwave infrared ranges of the electromagnetic spectrum, albeit in broader spectral bands (Diwa et al., 2024; Kaplan, 2018). The S2 bands combined with twenty-two (22) biophysical and three (3) terrain variables based on critical conditions for onion growth provide a robust dataset for applying classical machine learning algorithms (MLAs) for onion mapping. These indices encompass key indicators for monitoring crop health and suitability such as overall vegetation health, chlorophyll content, nitrogen content, soil moisture, soil salinity, soil clay content, Leaf Area Index (LAI), etc.

Premised upon the main goal of developing an operational workflow for onion mapping, this work evaluates the performance of various gradient boosting MLAs (i.e. CatBoost, Light Gradient Boosting Machine, and Extreme Gradient Boosting Machine) in accurately mapping onion fields in Bongabon, Nueva Ecija, using S2 imagery. This study underscores the potential of blending S2 multispectral data with MLAs and auxiliary growth variables for onion mapping. The findings contribute to the goal of

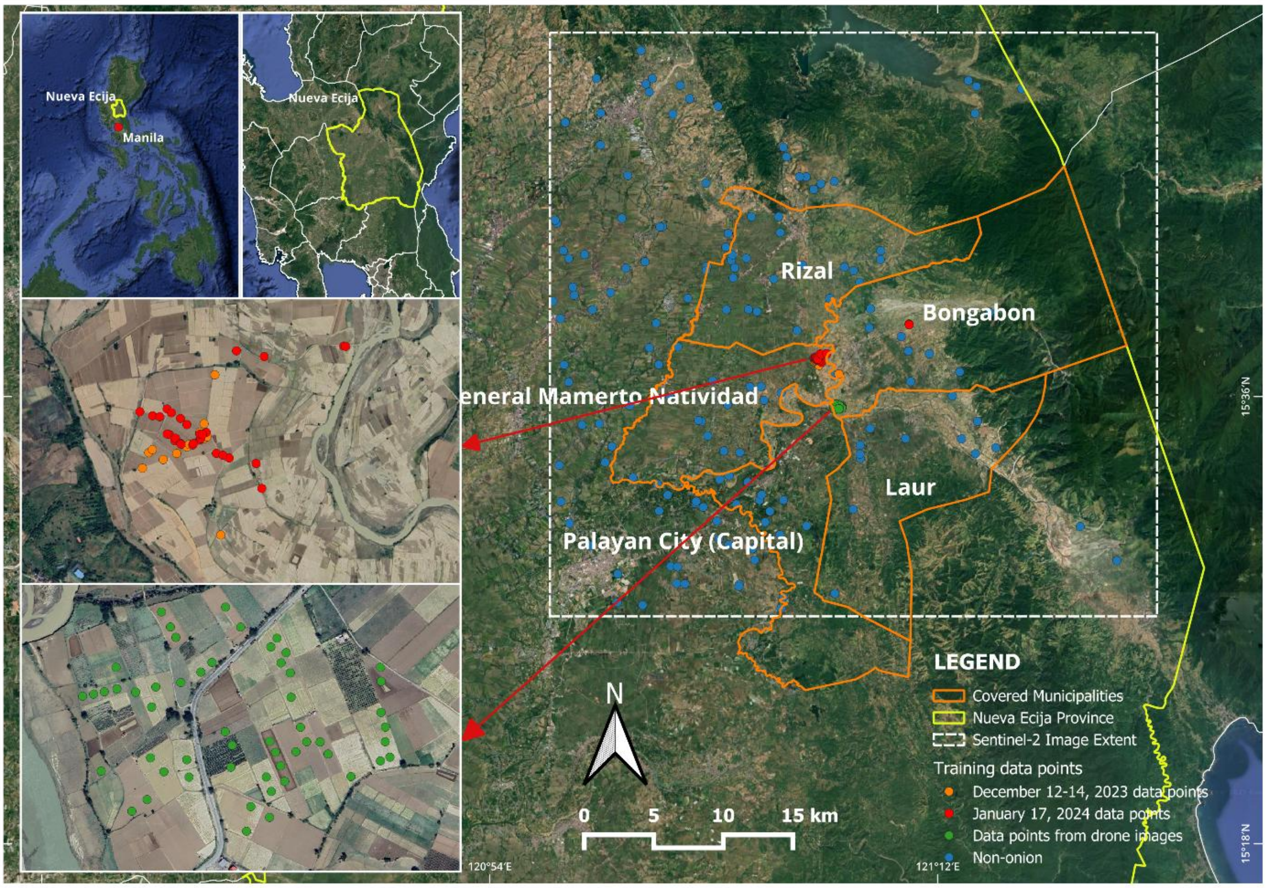

Figure 1. Location map of the study area and ground data points of onion and non-onion collected in Nueva Ecija, Philippines

achieving more accurate onion mapping, providing a scalable approach to monitor onion production and mitigate market volatility.

Figure 2. RGB composite (Bands 4-3-2) of Sentinel-2 image for the study area captured on February 6, 2024

2. Materials And Methods

2.1 Location of Study Area and Ground Data Collection

The study was conducted in neighbouring municipalities of Bongabon, Rizal, General Mamerto Natividad, Laur, and Palayan City in the province of Nueva Ecija, approximately 144 km north of Manila, the capital of the Philippines (Figure 1). Often referred to as the “Onion capital of the Philippines”, Bongabon is the leading producer of onion in the country and in Southeast Asia with approximately 27 km^2 of arable land for onion production (Gavino & Tiw-an, 2020). The onion cultivated in the study area has a growing period of 90 – 110 days (3 – 4 months) and is normally planted from September to January and harvested from December to April. Aside from onion, rice and other seasonal crops (e.g., corn) are also cultivated in Bongabon.

Three field campaigns were conducted to gather ground truth data of onion cultivation sites using the ODK collect app (Bokonda et al., 2019). The first two field visits which took place on December 12-14, 2023 and January 17, 2024 were carried out to collect onion ground locations (including drone images in January 2024). To note, the onions were at vegetative and bulbing states during the December 2023 and January 2024 field visits, respectively. During the third visit on April 15–17, 2024, local farmers were shown a February 2024 Sentinel-2 mosaic of the study area. Using reference points provided to guide them, the farmers identified the locations of onion and rice planting areas directly on the map. The onion and non-onion data points from the field visits are also shown in Figure 1.

2.2 Satellite Image and Pre-processing

To perform onion mapping, we utilized S2 Level 2A image acquired on 6 February 2024 covering the study area (Figure 2), selected based on the phenological stage of onion growth (i.e., bulbing to maturity) when spectral contrast is maximized. Since S2 Level-2A data are pre-corrected for atmospheric and geometric effects, the image downloaded from the Copernicus Browser required no additional preprocessing before analysis. Ten (10) spectral bands were used as follows: RGB (Bands 2, 3, and 4), Vegetation Red Edge (VRE; Bands 5, 6, 7, and 8A), Near-Infrared (NIR; Band 8), and Shortwave Infrared (SWIR; Bands 11 and 12).

In addition to the spectral bands, twenty-two (22) derived indices were calculated using both GEE and Sentinel Application Platform (SNAP). These indices were selected based on agronomic studies highlighting optimal onion growth conditions (Boyhan & Kelley, 2007), particularly: (1) well-drained soils, (2) low sulfur content, (3) adequate nitrogen and phosphorus levels, and (4) moisture content. Table 1 summarizes the biophysical indices used in this study as well as their formulas and relevance for onion mapping.

The S2 downloaded from the Copernicus Browser was used to calculate the Leaf Area Index (LAI), Fraction of Vegetation Cover (fCover), Fraction of Absorbed PAR (fAPAR), Canopy Water Content (CW), and Canopy Chlorophyll Content (CCC) in SNAP. All the S2 bands were first resampled to 10 meters prior to calculation of these indices. The seventeen (17) other biophysical indices were calculated in GEE.

Terrain variables were also incorporated in the analysis, as onions typically grow only in flat areas. This helps minimize potential misclassification of onions with grasses or low-lying shrubs in steeper terrains. Terrain variables such as the slope, roughness, and Topographic Position Index (TPI) were calculated in QGIS v3.34.2 using Interferometric Synthetic Aperture Radar (IfSAR) Digital Elevation Model (DEM) with a spatial resolution of 5 m. The spectral bands, biophysical indices, and terrain variables were stacked into one image file and then crop masked using the annual and perennial crop layers of the 2020 land cover map of National Mapping and Resource Information Authority (NAMRIA) (*Geoportal PH*, n.d.). A flow chart showing the steps for preprocessing and succeeding classification techniques used in this study is shown in Figure 3.

2.3 Machine Learning Algorithms

Three gradient boosting machine learning algorithms (MLAs) such as Light Gradient Boosting Machine (LigthGBM), Categorical Boosting Classifier (CatBoost), and Extreme Gradient Boosting Machine (XGBoost) were evaluated to classify onion and non-onion areas. Briefly, these three MLAs share key characteristics that make them particularly advantageous for onion mapping using S2, such as (1) high predictive performance even with complex, nonlinear relationships, (2) less sensitive to multicollinearity, and (3) strong performance on small to moderate-sized datasets, among others.

To improve the weak learner classifiers, gradient boosting algorithms iteratively train a series of individual classifiers, with each stage aimed at correcting the errors of the previous ones (Kavzoglu & Colkesen, 2013; Mahesh & Soundrapandiyan, 2024; Waske & Braun, 2009). Among the widely used implementations, LightGBM, CatBoost, and XGBoost are gradient boosting frameworks built on decision trees but differ in how the trees are constructed. LightGBM adopts the leaf-wise strategy to grow the trees by splitting the leaf with the largest loss reduction first which often results in deeper and more complex trees (Ustuner & Balik Sanli, 2019). CatBoost uses the symmetric tree growth strategy by ensuring balanced trees at each level (Hancock & Khoshgoftaar, 2020). In contrast, XGBoost uses the level-wise tree growth strategy by expanding all leaves at the same depth

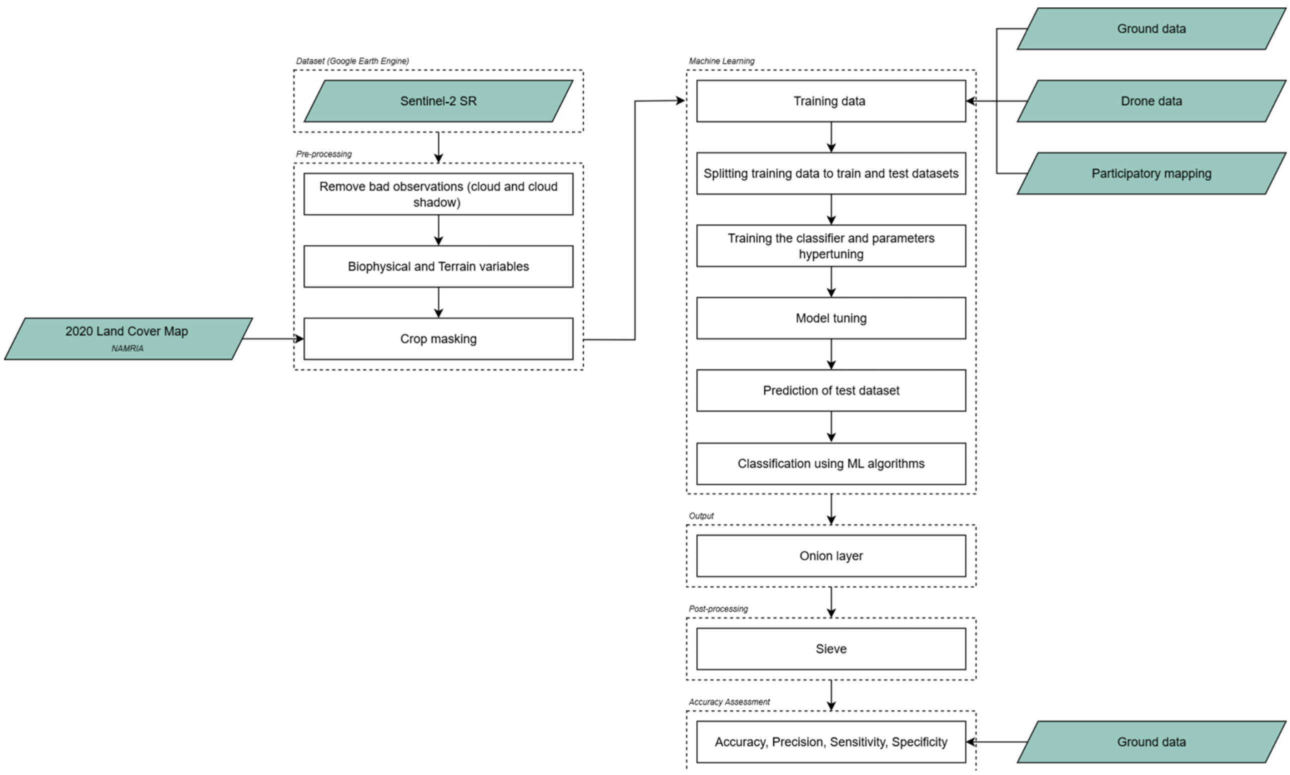

Figure 3. Methodological flow diagram for mapping onion fields using machine learning.

Index	Formula	Relevance for Onion Mapping
Vegetation Health and Biomass		
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$	Indicates general plant vigor and greenness.
Enhanced Vegetation Index (EVI)	$EVI = 2.5 \times \frac{(NIR - Red)}{(NIR + 6Red - 7.5Blue + 1)}$	Improves vegetation monitoring by reducing soil background noise.
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{(NIR - Green)}{(NIR + Green)}$	Sensitive to chlorophyll; useful for assessing nitrogen-related growth.
Leaf Area Index (LAI)	Empirically derived from reflectance or radiative transfer models	Measures canopy density, useful for estimating growth stage and biomass.
Fraction of Vegetation Cover (fCover)	$fCover = \frac{LAI}{LAI_{max}}$	Shows proportion of ground covered by vegetation; assesses uniformity of growth.
Fraction of Absorbed PAR (fAPAR)	Derived from canopy reflectance	Evaluates how efficiently plants use light for photosynthesis.
Moisture and Water Stress		
Normalized Difference Water Index (NDWI) [a]	$NDWI = \frac{(Green - NIR)}{(Green + NIR)}$	Detects water content in vegetation.
Moisture Stress Index (MSI) [b]	$MSI = \frac{SWIR1}{NIR}$	Identifies water stress in plants.
Normalized Difference Infrared Index (NDII) [c]	$NDII = \frac{(NIR - SWIR1)}{(NIR + SWIR1)}$	Sensitive to water content and hydration levels.
Canopy Water Content (CW)	Derived from radiative transfer model (i.e., PROSAIL) and incorporated by Artificial Neural Network (ANN)	Detects water content and stress
Nutrient and Chlorophyll Status		
Chlorophyll Index (CI_{green})	$CI_{green} = \frac{NIR}{Green} - 1$	Estimates chlorophyll concentration, a proxy for nitrogen levels.
Normalized Difference Red Edge (NDRE)	$NDRE = \frac{(NIR - VRE1)}{(NIR + VRE1)}$	Tracks nitrogen content and stress.
Red Edge Chlorophyll Index ($CI_{red\text{-}edge}$)	$CI_{red-edge} = \frac{NIR}{VRE1} - 1$	Highlights chlorophyll status using red edge sensitivity.
Sentinel-2 Red Edge Position Index (S2REP) [d]	$S2REP = 705 + 35 \times \left[\frac{(NIR + Red)/(2 - VRE1)}{VRE2 - VRE1}\right]$	Tracks shifts in chlorophyll due to nutrient changes.
Modified Chlorophyll Absorption Ratio Index (MCARI)	$MCARI = [(VRE1 - Red) - 0.2(VRE1 - Blue)] \times \frac{VRE1}{Red}$	Enhances chlorophyll estimation by minimizing background noise.
Canopy Chlorophyll Content (CCC)	Derived from radiative transfer models using Red Edge and NIR	Measures total chlorophyll in the canopy.
Soil and Substrate Characteristics		
Soil-Adjusted Vegetation Index (SAVI)	$SAVI = \frac{(NIR - Red)}{(NIR + Red + L)} \times (1 + L); L = 0.5$	Adjusts for soil background, useful in low vegetation areas.
Clay Minerals Ratio (CMR)	$CMR = \frac{SWIR1}{SWIR2}$	Identifies soil type (sandy vs. clayey), critical for sulfur content and bulb pungency.
Normalized Difference Salinity Index (NDSI) [e]	$NDSI = \frac{(Red - NIR)}{(Red + NIR)}$	Detects saline soils, which may hinder onion growth.
Brightness Index (BI) [f]	$BI = \sqrt{Red^2 + NIR^2}$	Lighter soils (often sandy, well-drained) tend to have higher brightness.
Soil Composition Index (SCI)	$SCI = \frac{(SWIR1 - NIR)}{(SWIR1 + NIR)}$	Helps distinguish different soil types.
Photosynthetic Activity and Stress		
Modified Photochemical Reflectance Index (PRI) [g]	$PRI = \frac{(Green - Red)}{(Green + Red)}$	Measures changes in photosynthetic efficiency, indicating stress.

Table 1. Biophysical indices used in this study***

*** [a]McFeeters (1996), [b]Hunt & Rock (1989), [c]Klemas & Smart (1983), [d]Frampton et al. (2013), [e]Khan et al. (2005), [f]Khan et al. (2001), [g]Peñuelas et al. (1995)

before proceeding deeper (T. Chen & Guestrin, 2016; H. Zhang et al., 2019). Several works used these gradient boosting algorithms for estimation of crop yield (Mahesh & Soundrapandiyan, 2024) and crop mapping (Ustuner et al., 2019). In fact, Ustuner et al. (2019) noted that gradient boosting algorithms outperform linear boosters in terms of crop mapping accuracy. The MLA classifications were performed using PyCaret 3.0 python package, an open-source machine learning library (*Pycaret*, n.d.).

To train the MLA models, a total of 250 datapoints were generated from the field surveys, consisting of 100 onions (17 from December 2023 ODK, 27 from January 2024 ODK, and 56 from January 2024 drone data) and 150 non-onion samples (collected from the participatory mapping). The dataset was then randomly split into 70 % for training, 15 % for tuning, and 15 % for testing to evaluate the model performance of the MLAs.

3. Results And Discussion

3.1 Onion Maps and Classification Accuracy

Gradient boosting algorithms such as LightGBM, CatBoost, and XGBoost were applied to the pre-processed S2 containing the spectral bands, biophysical indices emphasizing the ideal conditions for onion growth, and pertinent terrain variables to delineate the onion fields in Bongabon and nearby municipalities in Nueva Ecija, Philippines. Unlike linear algorithms, gradient boosting algorithms are advantageous for complex, nonlinear relationships as they work by combining many simple decision trees, with each tree improving the errors of the previous ones to achieve more accurate mapping of onion fields (Waske & Braun, 2009). The models were trained using 250 ground data points consisting of 100 onions and 150 non-onions, split into 70 % training, 15 % tuning, and 15 % test. The accuracy metrics of the training and tuned models as well as the test dataset are shown in Figure 4. To note, the accuracies of the training models were 0.9430 (LightGBM), 0.9267 (CatBoost), and 0.9183 (XGBoost).

The post-processed onion classification maps generated using the trained MLA models are shown in Figure 5. To assess the agreement between the three MLA models in classifying onion, an onion consistency map was created, counting how often a pixel is classified as onion, also shown in Figure 5. To validate and assess the accuracy of the onion maps, ground truthing was carried out on 15-17 April 2024 using the Open Data Kit (ODK) collect app (Bokonda et al., 2019) and participatory mapping with local farmers. The points visited for field validation were chosen by random sampling of 150 points, comprising of 60 onions and 90 non-onion ground data points. To note, onion locations were verified through direct field visits, while non-onion points (e.g., built-up areas, rice fields, bare soil, etc.) were identified based on participatory mapping results and satellite images. The ground truth data were then used to calculate the accuracy, precision, sensitivity, and specificity of the onion classification (Table 2).

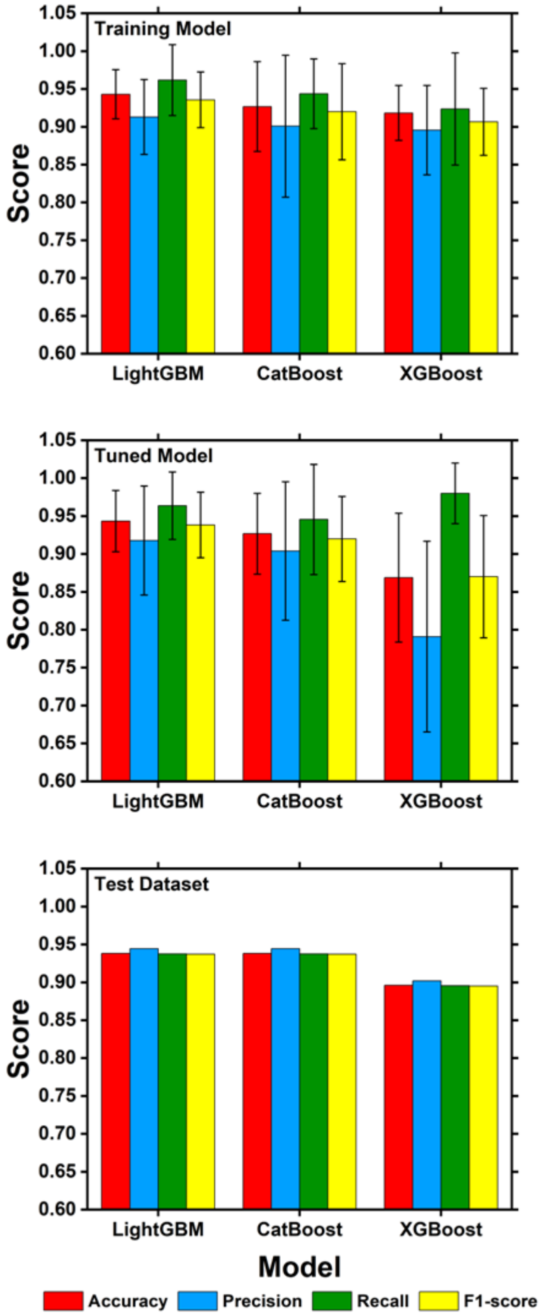

Figure 4. Accuracy metrics of the training and tuned models and the overall test dataset

The accuracies of the MLAs are as follows: CatBoost (90.0 %), LightGBM (86.7 %), and XGBoost (84.7 %). Notably, the classified maps achieved high precision and specificity levels, indicating their strong ability to minimize overestimation of onion classification and their effectiveness in correctly identifying non-onion pixels, respectively. However, the maps exhibited relatively modest sensitivity, indicating challenges in detecting actual onion pixels, which may be attributed to the overlapping spectral features between onions, other crops, and bare soil which was also mentioned in Diwa et al. (2024). Moreover, the “3” classification in the consistency map only slightly increased the accuracy to 90.5 %, though yielding insignificant improvement to other accuracy metrics.

Accuracy Metrics		LightGBM	CatBoost	XGBoost	Consistency Map
True Positive	TP	43	48	39	37
True Negative	TN	87	87	88	87
False Positive	FP	3	3	2	2
False Negative	FN	17	12	21	11
Accuracy		86.7%	90.0%	84.7%	90.5%
Precision		93.5%	94.1%	95.1%	94.9%
Sensitivity		71.7%	80.0%	65.0%	77.1%
Specificity		96.7%	96.7%	97.8%	97.8%

Table 2. Accuracy metrics of the onion and non-onion classification

3.2 Comparative Analysis of the MLA results

Detailed examination of onion classification in Figure 5 shows very little difference in onions classified by the three MLA models. Generally, the models provided a smooth transition between onions and other land cover types in the study area (i.e., rice and bare soil). Relative to the earlier study by Diwa et al. (2024) that utilized PRISMA hyperspectral image, the present approach provided a more accurate depiction of onion distribution, particularly in mixed vegetation zones (Figures 5B&D) and along transitional boundaries between crop types (Figures 5A–C). Minor misclassifications remained in certain parts of the area (Figure 5A) but occurred less frequently than in the previous results. Notably, this study’s models more effectively distinguished onions from rice, fallow land, and burned areas.

A notable observation is that the CatBoost model occasionally classified early-stage rice fields as onions (Figures 5A and 5D). This misclassification likely stems from the spectral resemblance between onions in their bulbing or mature phases and rice during its initial growth stage. Because young rice crops only partially cover the soil surface, their reflectance often represents a mixture of soil and vegetation signals, making them spectrally similar to onion plots. Although CatBoost achieved the highest classification accuracy, it still exhibited confusion when both crops displayed limited canopy cover and comparable greenness levels (Cheng et al., 2023).

3.3 Variable Importance

Among the bands and indices used, the CMR and PRI were consistently identified as the most important variables in classifying onion. This finding underscores the critical role of soil properties and moisture in shaping onion cultivation patterns. As shallow-rooted crops, onions are highly sensitive to moisture stress during the bulbing stage (Rao, 2016). The importance of CMR possibly indicates that soil mineral composition (i.e., clay) that affects water retention in soils plays a significant role in onion growth (Gaiser et al., 2000). This is supported by earlier discussion on optimal onion growth conditions, particularly, well-drained soils (Boyhan & Kelley, 2007).

Meanwhile, PRI is also recognized as a water-sensitive index (Thenot et al., 2002). Beyond this, its sensitivity to photosynthetic efficiency possibly enhances its effectiveness in distinguishing onion canopies, which exhibit distinct physiological and spectral responses compared to surrounding crops (Zhang et al., 2016). Overall, the importance of PRI in mapping onions is likely due to its responsiveness to key phenological stage/s of onion growth, particularly the bulbing stage, when onions exhibit unique physiological characteristics and heightened sensitivity to moisture stress.

All in all, the results of this work feature the importance of combined static (soil) and dynamic (physiological growth of onion)

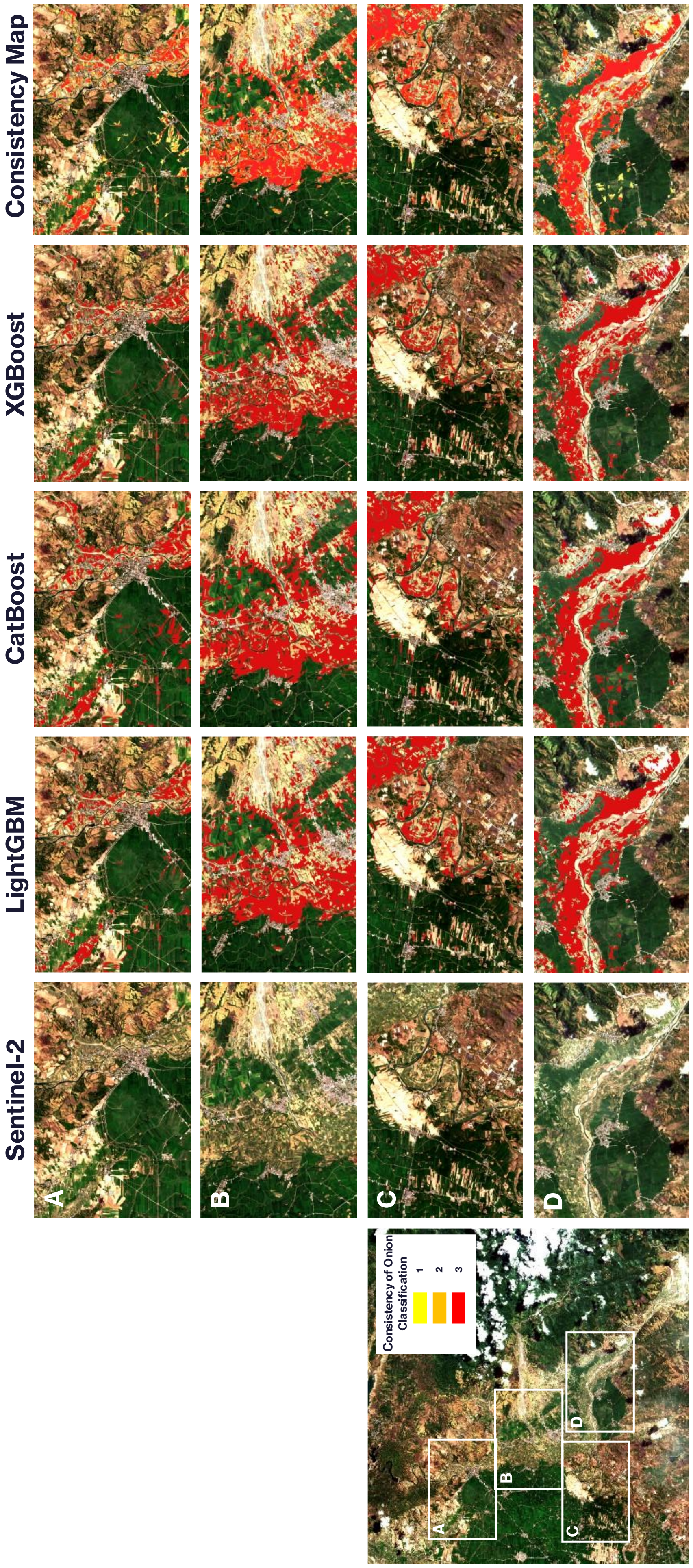

Figure 5. Onion classification maps generated using LightGBM, CatBoost, and XGBoost, with onion areas highlighted in red. The map also illustrates the spatial consistency of onion classification across all machine learning algorithms, as indicated by the legend in the inset map.

factors for accurate onion mapping. Moreover, this work highlights opportunities for targeted soil management and water-use optimization in onion cultivation.

3.4 Conclusion and Future Work

This study showed the comparative analysis for onion mapping using gradient boosting algorithms such as LightGBM, CatBoost, and XGBoost. The robustness of CatBoost in classifying onion fields was demonstrated in this study. Future works include development of ways to address the misclassification of early-stage rice as onions by CatBoost.

Acknowledgements

The authors would like to thank the local government of Bongabon, Nueva Ecija, for their invaluable support for the implementation of the field visits. We would also like to thank our colleagues in the Philippine Space Agency for their support in this work. Lastly, we would like to thank the Department of Agriculture – Bureau of Agricultural and Fisheries Engineering (DA-BAFE) for the DigitalAgri Project.

References

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., R. Verón, S., 2018. Remote Sensing and Cropping Practices: A Review. Remote Sensing 10, 99. https://doi.org/10.3390/rs10010099

Bokonda, L., Khadija, O.T., Souissi, N., 2019. Mobile Data Collection Using Open Data Kit. pp. 543–550. https://doi.org/10.1007/978-3-030-36778-7_60

Boyhan, G.E., Kelley, W.T., 2007. 2007 Onion Production Guide.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. Presented at the Proceedings of the 22nd international conference on knowledge discovery and data mining, pp. 785–794.

Chen, Y., Guo, Y., Qiao, L., Xia, H., 2022. Coupling optical and SAR imagery for automatic garlic mapping. Front. Sustain. Food Syst. 6. https://doi.org/10.3389/fsufs.2022.1007568

Cheng, G., Ding, H., Yang, J., Cheng, Y., 2023. Crop type classification with combined spectral, texture, and radar features of time-series Sentinel-1 and Sentinel-2 data. International Journal of Remote Sensing 44, 1215–1237.

Diwa, R.R., Borlongan, N.J.B., Tabardillo, J.I.A., Sabuito, A.J.C., Candido, C.G., de la Cruz, R.M., Blanco, A.C., 2024a. DEVELOPMENT OF METHODOLOGY FOR MONITORING AND ASSESSING THE NATIONAL GREENING PROGRAM USING OPTICAL AND SAR DATA. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4-W8-2023, 175–181. https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-175-2024

Diwa, R.R., Cabello, K.E., Blanco, A.C., 2024b. Onion mapping using PRISMA hyperspectral image: comparative analysis of K-means classification and linear spectral unmixing [WWW Document]. URL https://www.spiedigitallibrary.org/conference-proceedings-of-spie/13266/132660P/Onion-mapping-using-PRISMA-hyperspectral-image--comparative-analysis-of/10.1117/12.3041512.short (accessed 6.11.25).

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. Remote sensing of environment 97, 192–202.

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS journal of photogrammetry and remote sensing* 82, 83-92.

Gaiser, T., Graef, F., Cordeiro, J.C., 2000. Water retention characteristics of soils with contrasting clay mineral composition in semi-arid tropical regions. Soil Res. 38, 523–536. https://doi.org/10.1071/sr99001

Gavino, R.B., Tiw-an, C.C., 2020. Assessment of mechanization level of onion production in Nueva Ecija. CLSU International Journal of Science & Technology 4, 81–98.

Geoportal PH [WWW Document], n.d. URL https://www.geoportal.gov.ph/ (accessed 10.1.24).

Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. Journal of big data 7, 94.

Hunt Jr, E. R., & Rock, B. N., 1989. Detection of changes in leaf water content using near-and middle-infrared reflectances. *Remote sensing of environment* 30(1), 43-54.

Hutchinson, D., 2023. The continuing struggle for onion farmers in the Philippines. PlantwisePlus Blog. URL https://blog.plantwise.org/2023/02/14/the-continuing-struggle-for-onion-farmers-in-the-philippines/ (accessed 9.26.24).

Joshi, A., Pradhan, B., Gite, S., Chakraborty, S., 2023. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sensing* 15, 2014. https://doi.org/10.3390/rs15082014

Kaplan, G., 2018. Sentinel-2 Pan Sharpening—Comparative Analysis. Proceedings 2, 345. https://doi.org/10.3390/ecrs-2-05158

Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. Journal of Hydrology 586, 124905. https://doi.org/10.1016/j.jhydrol.2020.124905

Kavzoglu, T., Colkesen, I., 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. International journal of remote sensing 34, 4224–4241.

Khan, N. M., Rastoskuev, V. V., Shalina, E. V., & Sato, Y., 2001. Mapping salt-affected soils using remote sensing indicators-a simple approach with the use of GIS IDRISI.

Khan, N. M., Rastoskuev, V. V., Sato, Y., & Shiozawa, S., 2005. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management* 77(1-3), 96-109.

Kingra, P.K., Majumder, D., Singh, S.P., 2016. Application of remote sensing and GIS in agriculture and natural resource management under changing climatic conditions. Agricultural Research Journal 53.

Klemas, V., & Smart, R., 1983. The influence of soil salinity, growth form, and leaf moisture on-the spectral radiance of. *Photogramm. Eng. Remote Sens* 49, 77-83.

Liu, J., Feng, Q., Gong, J., Zhou, J., Liang, J., Li, Y., 2018. Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data. International journal of digital earth 11, 783–802.

Mahesh, P., Soundrapandiyan, R., 2024. Yield prediction for crops by gradient-based algorithms. Plos one 19, e0291928.

Mariotto, I., Thenkabail, P.S., Huete, A., Slonecker, E.T., Platonov, A., 2013. Hyperspectral *versus* multispectral crop-productivity modeling and type discrimination for the HyspIRI mission. *Remote Sensing of Environment* 139, 291–305. https://doi.org/10.1016/j.rse.2013.08.002

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* 17(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Moran, M.S., Inoue, Y., Barnes, E.M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sensing of Environment 61, 319–346. https://doi.org/10.1016/S0034-4257(97)00045-X

Navid, R., Vieira, E., Vania, 2019. Applications of Image Processing and Soft Computing Systems in Agriculture. IGI Global.

Opara, L.U. ;Mejía, 2003. ONIONS: Post-Harvest Operation [WWW Document].

Pareek, S., Sagar, N., Sharma, S., Kumar, V., 2017. Onion (Allium cepa L.): Chemistry and Human Health, 2nd Edition. pp. 1145–1162. https://doi.org/10.1002/9781119158042.ch58

Peñuelas, J., Filella, I., & Gamon, J. A., 1995. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytologist* 131(3), 291-296.

Philippine Statistics Authority, n.d. Costs and Returns Survey of Onion Production 2006 [WWW Document]. URL https://psada.psa.gov.ph/catalog/37 (accessed 9.27.24).

Pycaret [WWW Document], n.d. . PyCaret. URL https://pycaret.org/ (accessed 8.17.25).

Rao, N.K.S., 2016. Onion, in: Rao, N.K.S., Shivashankara, K.S., Laxman, R.H. (Eds.), Abiotic Stress Physiology of Horticultural Crops. Springer India, New Delhi, pp. 133–149. https://doi.org/10.1007/978-81-322-2725-0_8

SEARCA, n.d. 2nd Small and Family Farmers | New and Beginning Farmers National Conference - SEARCA [WWW Document]. URL https://www.searca.org/events/conferences/2nd-small-and-family-farmers-new-and-beginning-farmers-national-conference (accessed 8.15.25).

Thenot, F., Méthy, M., Winkel, T., 2002. The Photochemical Reflectance Index (PRI) as a water-stress index. International Journal of Remote Sensing 23, 5135–5139. https://doi.org/10.1080/01431160210163100

Ustuner, M., Balik Sanli, F., 2019. Polarimetric Target Decompositions and Light Gradient Boosting Machine for Crop Classification: A Comparative Evaluation. ISPRS International Journal of Geo-Information 8, 97. https://doi.org/10.3390/ijgi8020097

Ustuner, M., Balik Sanli, F., Abdikan, S., Bilgin, G., Göksel, Ç., 2019. A Booster Analysis of Extreme Gradient Boosting for Crop Classification using PolSAR Imagery. https://doi.org/10.1109/Agro-Geoinformatics.2019.8820698

Waske, B., Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS journal of photogrammetry and remote sensing 64, 450–457.

Zhang, C., Filella, I., Garbulsky, M.F., Peñuelas, J., 2016. Affecting Factors and Recent Improvements of the Photochemical Reflectance Index (PRI) for Remotely Sensing Foliar, Canopy and Ecosystemic Radiation-Use Efficiencies. Remote Sensing 8, 677. https://doi.org/10.3390/rs8090677

Zhang, H., Eziz, A., Xiao, J., Tao, S., Wang, S., Tang, Z., Zhu, J., Fang, J., 2019. High-Resolution Vegetation Mapping Using eXtreme Gradient Boosting Based on Extensive Features. Remote Sensing 11, 1505. https://doi.org/10.3390/rs11121505