Semi Supervised Binary Crop Type Mapping

Stephen Korir

November 2024

1 Introduction

Accurate crop type mapping is a crucial step in agricultural management influencing critical aspects like creation of policies, resource allocation, creation of a country food balance sheet, yield estimation among others. With the increasing population, exacerbated by climate change, there is a global demand for sustainable agricultural practices. Consequently, this requires accurate datasets to help optimize processes and decisions in the agricultural value chain. Contrary to the conventional sampling procedures for crop area estimation, remote sensing has changed and transformed crop type mapping by introducing efficiency, accuracy and providing a wide spatial coverage and temporal resolution enabling researchers to monitor agricultural regions effectively.

Crop classification however faces some challenges like spectral similarity between crops which makes it hard to segregate some crops and lack of extensive ground truth data to help in the modeling process. To address some of these challenges, semi-supervised approaches offer a promising solution . Using unsupervised clustering techniques with supervised classification models, these approaches make use of large amounts of unlabeled data available in satellite imagery while using limited amounts of georeferenced ground truth for precise mapping.

This study investigates the application of semi-supervised learning for crop type mapping in Uasin Gishu County, Kenya, which is a maize basket and a major maize producing county in the country. Using sentinel-2 datasets and positive ground truth labels, this research aims to evaluate the potential of combining clustering algorithms and supervised learning to enhance crop classification accuracy. The results and findings will immensely contribute to improving remote sensing methodologies and agricultural planning.

2 Literature Review

Remote sensing technologies have transformed the agriculture spectrum by providing mechanisms for monitoring, planning and decision making. Sentinel-2 imagery, which is part of the Copernicus program, with a relatively moderate to high spatial resolution of 10m to 20m, and a high revisit frequency of

approximately 5 days, has been particularly critical for crop classification tasks ([Foerster et al., 2012]). These satellites capture spectral bands from the electromagnetic spectrum which are sensitive to vegetation changes. This is a precursor to crop type segregation aided by their unique spectral signatures. Campbell and Wynne [2011] emphasized that remote sensing allows for scalable agricultural monitoring by overcoming the logistical challenges and cost pegged to field based sampling and mapping surveys, with increased efficiency and accuracy.

As much as remote sensing caters for other downstream tasks like yield prediction, water stress management, crop monitoring (Lu and Weng [2007]), accuracy and precision of these applications are heavily tied to the accuracy of the crop type maps mined from satellite imagery.

Gomez et al. [2019], enumerated three primary issues that affect crop type classification. These include, mixed pixels in imagery, spectral similarity between crops and phenological variability brought about by varying climatic factors.Crops from the same scientific family, overlapping growth cycles and similar canopy structures can exhibit the same spectral signatures which would make it hard to segregate them, and this would consequently lead to misclassification. Misclassification can also be contributed by spatial resolution. Medium to low coarse resolution will always have different land cover types or crop types being represented by a single pixel. This will therefore lead to accuracy issues in the resulting model.

According to Xie et al. [2017], ground data collection is time,location and resource intensive, making it unsustainable for large scale studies. This will consequently lead to scarcity of data thereby limiting the classification accuracy.

Semi supervised learning creates an intersection between labeled and unlabeled datasets by leveraging vast amounts of remotely sensed imagery to compensate for the limitations of ground truth data. As demonstrated by Ng and Li [2017], integration of unsupervised and supervised methods significantly improves the classification outcomes. By clustering unlabeled data into groups, and using labeled samples to get the negative and positive samples, semi-supervised learning reduces dependency on labeled datasets.

Li et al. [2020] highlighted the effectiveness of using K-means clustering for extracting natural groupings in spectral data. When used with machine learning algorithms like Random Forest, these clusters act as proxy labels enabling more accurate classification of crop types.

Ensemble learning methods like Random forest have proven to be robust in handling high-dimensional and correlated datasets(Belgiu and Dragut̨ [2016]), and have shown superior performance than traditional algorithms like maximum likelihood. On the other hand, deep learning approaches like Convolutional Neural Networks have also high potential for remote sensing. Gomez et al. [2019], have leveraged on these approached crop crop type identification, but the downside of this is the high dependence on large amounts of labeled data which semi-supervised learning methods aim to address.

Phenological monitoring is essential for crop identification. Creating a temporal signature for different crops enhances classification accuracy(Foerster et al. [2012]). Carrao et al. [2008] noted that multi-temporal profiles from multispec-

tral sensors will help in distinguishing crops with overlapping spectral profiles The integration of multiple data sources aims to compliment the deficiencies of the different sensors. Optical sensors like sentinel-2 datasets are always affected by clouds, but with sentinel-1 SAR, these can be mitigated since they can penetrate through clouds. Fusing sentinel-1 and sentinel-2 has also great promise in improving the classification accuracy by leveraging on the textural and structural properties from the radar sensors and spectral information for the optical sensors(Van Tricht et al. [2018])

As highlighted earlier, several limitations like mixed pixels and spectral overlaps still pose a challenge to the crop type mapping exercise(Xie et al. [2017]). The accuracy and effectiveness of semi-supervised learning still depends on the quality of the initial clustering. Future improvements may included development of earth foundational models and incorporation of hyperspectral data(Thenkabail et al. [2012])

3 Methodology

3.1 Study Area

The study focusses on Moiben Ward, in Uasin Gishu County, Kenya, which is a predominantly maize growing region. The selected area presents a perfect location to explore semi-supervised crop mapping due to its agricultural significance and availability of ground truth data and remote sensing data.

3.2 Data Collection

Field surveys were conducted to collcet maize polygons, yielding 35 georeferenced samples. Multisepectral sentinel-2 imagery was acquired from Google Earth Engine in the period of (June-August) of maize in the area

3.3 Data Preprocessing

Sentinel-2's QA60 band was used to mask clouds. A composite for the region was created by merging all the cloud masked imagery within that period. Vegetation indices like Normalized Difference Vegetation Index and Enhanced Vegetation were calculated and used as extra bands in the satellite imagery to enhance spectral feature discrimination.

3.4 Unsupervised Clustering

K-Means clustering was applied to the preprocessed imagery with 8 clusters, representing potential maize and non-maize classes Ground truth labels were used to identify clusters corresponding to maize and non-maize classes. Clusters with the highest frequencies of maize labels were designated as positive clusters 100 positive samples(Maize) and 1000 negative samples(non-maize) were extracted to be used for supervised classification

3.5 Supervised Classification

A random forest classifier was trained using the datasets generated from the clustering step. The dataset was split into 70 percent training and 30 percent testing sets. Model performance was evaluated using overall accuracy and a confusion matrix.

3.6 Evaluation Metrics

The classification results were evaluated for overall accuracy and misclassification rate, with analysis on contributing factors such as mixed pixels.

References

M. Belgiu and L. Draguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016. doi: 10.1016/j.isprsjprs.2016.01.011.

J. B. Campbell and R. H. Wynne. *Introduction to Remote Sensing*. Guilford Press, 5th edition, 2011. ISBN 9781609181765.

H. Carrao, P. Goncalves, and M. Caetano. Contribution of multispectral and multitemporal information from modis images to land cover classification. *Remote Sensing of Environment*, 112(5):986–997, 2008. doi: 10.1016/j.rse.2007.07.002.

S. Foerster, K. Kaden, M. Foerster, and S. Itzerott. Crop type mapping using spectral-temporal profiles and phenological information. *Remote Sensing*, 4 (11):3178–3205, 2012. doi: 10.3390/rs4113178.

C. Gomez, J. C. White, and M. A. Wulder. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:12–29, 2019. doi: 10.1016/j.isprsjprs.2016.03.008.

W. Li, Q. Guo, M. K. Jakubowski, and M. Kelly. A new method for segmenting individual trees from the lidar point cloud. *Remote Sensing of Environment*, 113(3):715–723, 2020. doi: 10.1016/j.rse.2020.10.004.

D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007. doi: 10.1080/01431160600746456.

W. Ng and P. Li. Semi-supervised learning for remote sensing image classification using multi-feature combination. *Remote Sensing*, 9(8):786, 2017. doi: 10.3390/rs9080786.

P. S. Thenkabail, J. G. Lyon, and A. Huete. *Advances in Hyperspectral Remote Sensing of Vegetation*. CRC Press, 2012. doi: 10.1201/b11219.

K. Van Tricht, A. Gobin, S. Gilliams, and I. Piccard. Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study in belgium. *Remote Sensing*, 10(10):1642, 2018. doi: 10.3390/rs10101642.

Y. Xie, Z. Sha, and M. Yu. Remote sensing imagery in vegetation mapping: A review. *Journal of Plant Ecology*, 11(1):100–114, 2017. doi: 10.1093/jpe/rtx005.