Sumatra Journal of Disaster, Geography and Geography EducationJune, 2022. Vol.7. No.1. pp. 45-50
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN : **2580 - 4030** (Print) **2580 - 1775** (Online), Indonesia

LAND-COVER/LAND-USE RANDOM FOREST CLASSIFICATION USING SENTINEL 1A RADAR IMAGERY

Grefie Dwinita[1], Projo Danoedoro[2], Prima Widayani[2]
[1]Master Student of Remote Sensing, Gadjah Mada University, Yogyakarta
[2]Lecturer of Remote Sensing, Gadjah Mada University, Yogyakarta
Email :grefiedwinita@mail.ugm.ac.id[1]

*Corresponding Author, Received: April 20, 2023. Revised: Mei 21, 2023. Accepted: June 26, 2023

ABSTRACT : Extraction of existing land use/land cover information can be obtained using remote sensing data. One of them is by using Sentinel 1A Imageas a Synthetic Aperature Radar (SAR) active system image with 20 meters resolution and 12 daystemporal resolution. This study aims to map land use/land cover in the city of Padang using the random forest classification (RFC) method through active system remote sensing data extraction. Accuracy test of the classification results using the confusion matrix method with overall accuracy of 85% and kappa value of 0.78. The most dominant land use in the study area was forest land use with a total area of 30,285 Ha reaching 43.83% of the City of Padang.

Keywords: Sentinel 1A, Synthetic Aperature Radar (SAR), Land Use/Land Cover, Random Forest Classification

1. INTRODUCTION

Remote sensing imagery is one of the main data sources in extracting geospatial information. Remote sensing technology has so far been developed and carried out by several researchers using both optical and radar data. Advances in Synthetic Aperture Radar (SAR) technology have the advantage of being able to penetrate clouds and canopies and could be operated day and night, regardless of sun conditions and rainfall, as a result SAR data can be utilized in permanently cloud areas [7]. Sentinel 1A imagery is categorized in the weather radar system with two C-Band polarization modes VV and VH. Sentinel 1A has a 5.405 GHz sensor with an angle of incidence between 20° and 45◦. The platform follows a circular, near-polar, sun-synchronous orbit at an altitude of 693 Km. There are 4 modes in Sentinel 1 IncludingStripmap (SM), Interferometric Wide Swath (IW), Extra Wide (EW) and Wave with various area, incident angle and polarization [8].

The city of Padang is an area that is directly liberated by the Indian Ocean. In this area, the rainfall is quite high so it can affect the condition of the cloud-covered image. In mapping land use/land cover, an image is required without any cloud disturbance so that the object can be mapped as a whole using an active image system.

Classification to map land use/land cover using remote sensing data is carried out in several ways, including by using parametric and non-parametric methods. Non-parametric assessment is increasingly used as an approach for classifying multi-source data [5]. The Random Forest Classification (RFC) algorithm is an ensemble classification proposed by Breiman combining several decision trees [10]. Determination of a decision tree when a single decision tree is installed excessively would result in a lower bias but a higher prediction variation. The variance of the RFC algorithm is reduced by averaging the output of several decision trees. The algorithm for each decision tree is constructed using a randomly selected subset of the polarimetric input descriptors to reduce the correlation among the different decision trees [9]. The parameters needed in determining the number of decision trees to create (k) variables and the number of randomly selected variables (m) are considered to divide each node in the decision tree [4]. The polarization parameter in SAR data has a correlation which leads to redundancy of information when using a single data. Single data is one observed radar data, which may not be able to characterize plant phenology due to the presence of non-linear signals between the radar and plant descriptors [9]. Thus, random forest algorithm combiningsome decision trees in the form of machine learning method can be used to resolve non-linearity and complexity in identifying plant phenology.

This study used Sentinel 1A imagery to obtain information on land landuse/land cover using the random forest classification method in the city of Padang. The use of Sentinel 1A imagery used VV and VH polarization to detect land landuse/land cover. The purpose of this study was to classify land use/land cover using the random forest method based on Sentinel 1A imagery and calculate the overall accuracy of the existing land use/land cover map.

2. METHODS

Research site

The re location was conducted in the City of Padang, West Sumatra Province, which is geographically located at 00º44'00” and 1°08'35” South Latitude and between 100°05'05” and 100°34'09” East Longitude. The research location is presented in Figure 1.

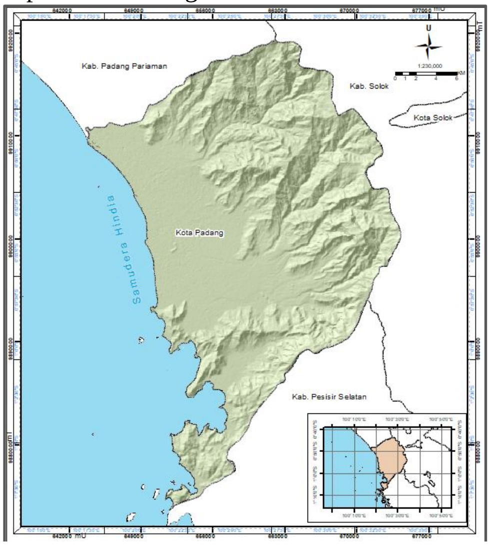

Figure 1. Research Location

Data

The data in this study used Sentinel 1AImagerywith Polarization VV and VH recording in October 21st, 2022: (S1A_IW_GRDH_1SDV_20221021T114126_20221021T114151_045539_0571B5_2D40) downloaded at https://search.asf.alaska.edu/#/.

Data Processing Stages

A. Sentinel 1A ImageProcessing

Sentinel 1A imagery radar data processingused SNAP Softwareversion 8.0 toproduceSentinel 1A image in Sigma NoughtDesiblevalues in the form of GeoTif format. The processofrecording radar imagewascarriedout in a side-lookingmannerconsequently,itcauseddisturbances in theimagerecordingresultssuch as specklenoise, foreshortening, layoverandshadow. An imagecalibrationprocesswasneededtoproducedetailedandclearinformationonimageappearance. Therewereseveralsteps in processing radar data, includingsubsetimage, apply orbit-file, thermalnoiseremoval, remove-GRD-Border Noise, Calibration, Speckle Filter, TerrainCorrection, Linear T FromdB. Accordingto ESA (2015)that in processing radar imagesincludingSentinel 1A Image usingseveral main steps[3].In this study, the image process was presented in graph builder process by processing image data automatically based on the built graph (Figure 2).

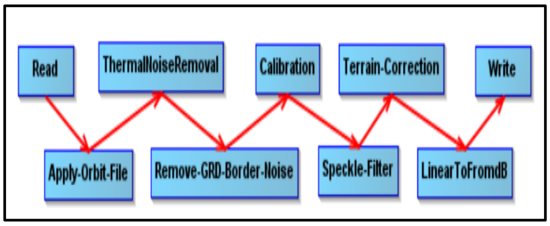

Figure 2. Sentinel 1A ImageProcessing

B. SampleTraining Area

Before starting the machine learning random forest classification process in SNAP software version 8.0, Sentinel 1A image whichhas already been calibrated and converted to Sigma Nought Desible (dB) polarization, in both VV and VH polarizations, then continued on process of training area sample or region of interest (ROI) to obtain pixel information on land cover classes in the study area. The sampling was carried out based on the RGB composite results such as VV Red (R), VH Blue (B) and VV-VH Green (G) Sigman Nought (Sigma0_dB) with band B (Blue) which was synthetic band. RGB composite arrangement aimed to make it easier in identifying objects visually.

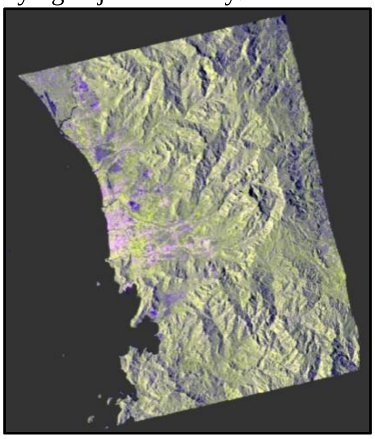

Figure 3. RGB Composite (VV,VH,VV-VH)

C. Random Forest Classification (RFC)

Random Forest Classification is an algorithm consisting of tree-based classifications collection based on equation 1.1:

h(x, θk), k=1,...}.............. Equation 1.1

in which, (x) is the input vector while (k) is the independent variable which is randomly distributed identically. Each decision tree is

Sumatra Journal of Disaster, Geography and Geography EducationJune, 2022. Vol.7. No.1. pp. 45-50
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN : **2580 - 4030** (Print) **2580 - 1775** (Online), Indonesia

built using a deterministic algorithm with bootstrap which leaves the remaining data points for validation and in a unit would choose the most popular class in the input (x) [5]. Therefore, the bootstrap procedure as a substitute for increasing the diversity of tree classifications allocating each pixel to the next class with the maximum numbers of vote/input in the decision tree collection. In carrying out the classification using the random forest method, the classification uses two specified parameters, which are the number of decision trees and the number of predictors to maximize model accuracy [7]. The more number of decision trees used, the higher the computational cost. Based on the research conducted, the number of 100 trees used in the classification was sufficient to ensure the average out-of-bag squared error (OOB-error) relatively stable to construct classification using the RFC method.

The advantages of the random forest algorithm consisting of some trees in which each tree was trained on a sample data. The class was determined based on the number of votes from each tree. The random forest and ensemble learning method have a higher level of accuracy compared to other machine learning algorithms because a collection of several classifiers provides more accurate results than a single classifier. Hence, the classification result was considered quite efficient by using the random forest classification method. The random forest algorithm was a non-parametric method as the random forest algorithm did not depend on data distribution assumptions. In which there was a relationship between predictor and response data so decision rules were made from input data to get the best model.The lack of a pixel-based random forest algorithm when applied to radar data whose speckle noise would result in a classification result in the form of salt and pepper even in a homogeneous area [6].

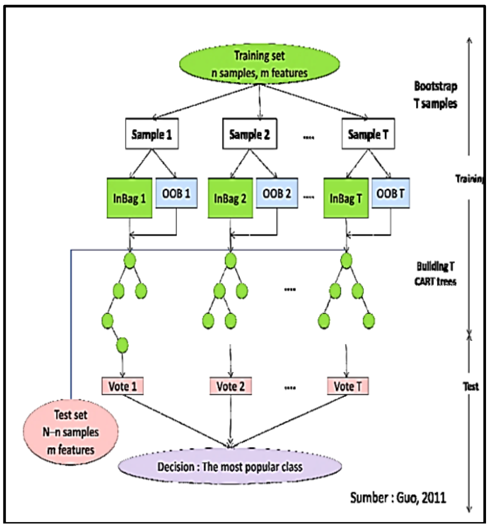

Figure 4.
RandomForestClassificationProcess

The Methodused wasmachinelearningwith therandomforestclassification (RFC) methodtodeterminetheland use/land cover classificationbasedonland cover classification SNI-7645-1-2014 andadditionalmodificationsweremadebasedo ntheconditionsof thefield in the study area. The randomforestmethod hadtheadvantageofdeterminingthesample as a training area basedon a collectionofsomeclassifiers, givingmoreaccurateresultsthan a singleclassifierin order to produce a moreaccurateclassification. The resultsofthelandcover wouldbefollowedby data extractionprocesstoobtaintheexistinglanduse in thefield.

D. Land Use Matching

Analysisofsensingimage data couldproducelandcovercategoriesrelatedtosp eciesdifferentiation, canopydensity, andvegetationphenology. In which ifecologicalcontextandfield data were involved couldobtaininformationrelatedtoecosystemsi n terms ofsocio-economicclass function [1].The processofextractinglanduseinformationwasc arriedoutusingthematchingmethod, whichwas a methodcombiningtheresultsofremotesensing data extractionwiththelandform in the study area. The goalwastoreducetheresultinglanduseinforma tionhavinghighaccuracy. Land usewasquitedifficulttodistinguishbetweenon elanduseclassandtheotheras a result,thecombinationofthetwo data informationcouldaddtotheexistingaccuracyv alueoflanduse. The matchingmethodwasused in a dimensiontablebycombiningtherowpartsofthelandcoverclassobjectswiththecolumnpartsofthelandformclassscattering in the study area.

E. SampleandFieldSurvey
Field sample was taken based on the distribution of land use/land cover as a result of Sentinel 1A imageryradarextraction using the random forest method. Sample method of *stratified random sampling*based on land cover/use area. There were 101 samples used as a test for the accuracy of land use maps in the study area. Regarding the number of samples using the confusion matrix accuracy test method, according to Jensen (2005) and Tso and Maher (2009) does not refer to any recommendations regarding the number of samples and the process [2].Confusion matrix accuracy test of kappa coefficient and overall accuracy (OA) were used to calculate the overall accuracy of different classifying measures. For the confusion matrix, Producer Accuracy (PA) and User Accuracy (UA) were calculated using the highest Overall Accuracy (OA) point of each classifier for the overall assessment [10].

3. RESULTS AND DISCUSSION

3.1 RESULTS

The resultsofdeterminingthetraining area atthe study siteobtained 10 landcoverclassesin whichthelandcoverextractionstagewascarriedoutusing a multispectralalgorithmwith randomforestclassification (RFC) method. This methodwasbasedonpreviousresearchwith therandomforest (RF) methodwhich wasconsideredobtainingbetterandmorepreciseresultscomparedtoothermethods. The randomforestmethodwas in theprocessoffacingoverfittingastreesorclassifiedproducedbytherandomforestalgorithmweremappedrandomly. The compilationofSentinel1A imagelandcover was toextractland use/land cover. The compiled map usingSentinel 1A imagerecordedinOctober 21st, 2022 obtained 10 landcoverclasses, includingwaterbodies, built-upland, open landwetlands, open landmoistland, open landdryland, lowlandvegetation in moistsoil, lowlandvegetation in dryland, highlandvegetationhighdensity, medium highlandvegetation, andlowdensityhighlandvegetation. The resultwasused as the basis forthepreparationoflandusewiththematchingmethodusinganapproachfromlandcoverandlandformstudies.

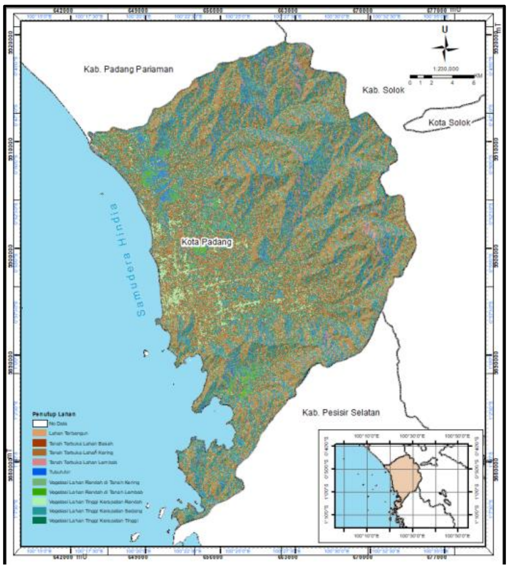

Figure 5. Land Cover Mapping

The landform map wasextractedfromSentinel 1A Digital Elevation Model (DEM) data. Processingwasconductedusing visual interpretationbydigitizingonscreen. Sentinel 1A imagewith RGB VV, VH and VV/VH compositeswas used toobtainlandcover. Hillshadeprocess was used on Sentinel 1A aimingto highlight visuallyappearanceofroughnessonthesurfaceofthe object. The classificationforthepreparationofthelandformmap resulted in 4 main landformsand 8 sub-landunits in the study area. The preparationofthelandform map obtainedthemostdominantlandform area in the study field such as structurallandformswithsynclinalmountain sub-units (S6) withan area of 33,800 Ha (48.9%) andsynclinalhills (S7) withan area of 12,937 Ha (18.7%)).The distribution of landforms in the study area started from the marine landforms in the west and hills to the mountains in the north and the central part of the fluvial landforms with sub-units of alluvial plains (F1), back swamps (F5), floodplains (F7) andalluvial fan (F16).

Sumatra Journal of Disaster, Geography and Geography EducationJune, 2022. Vol.7. No.1. pp. 45-50
Disaster, Geography, Geography Education http://sjdgge.ppj.unp.ac.id/index.php/Sjdgge
ISSN : **2580 - 4030** (Print) **2580 - 1775** (Online), Indonesia

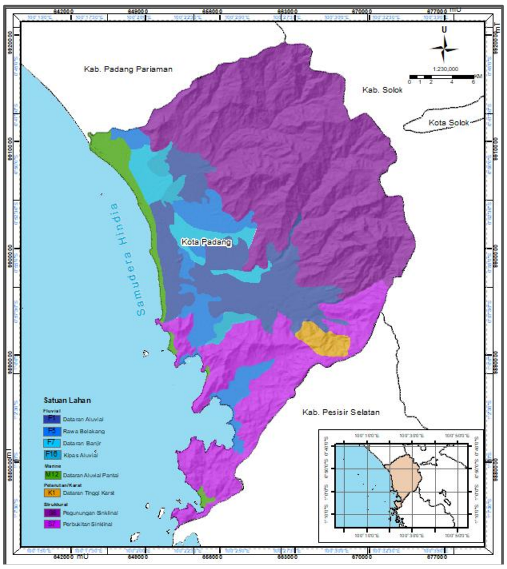

Figure 6. Landform Map

The spatialecologicalapproachassumedthatthephysical characteristicsofthelandaffectedlanduse. The stagesofintroducinglandusebasedon a two-dimensionalmatchingtablebetweenlandformsdisplayed in rowsandlandcoverdisplayed in columns. Land usewasidentifiedbasedonthelanduserelationshipthatemergedfromthose twoapproaches. In thelandcoverstudy arealocatedon open soil ofmoistlandwithfloodplainlandform couldbeassumedtobe a paddyfieldlanduseclass. Therewere 10 randomforestlandcoverclassesand 8 landformclasses,thus 8 landuseclasseswereobtained including settlements, paddyfield, field/moor, open land, mixedgarden, shrub, forestandriver.Thedistributionoflandusecanbeseen in Figure 7.

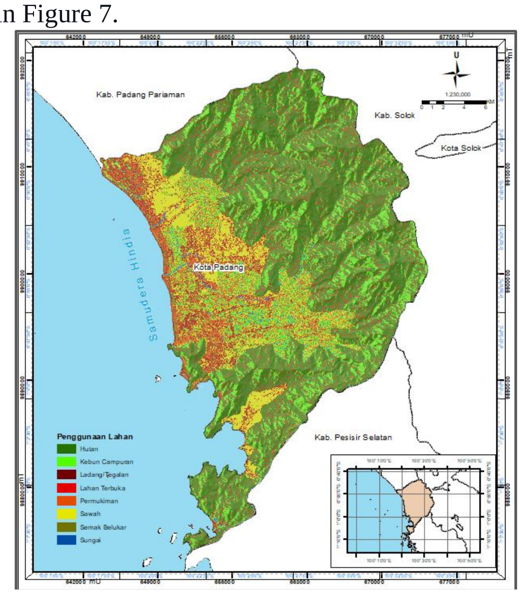

Figure 7. Land Use Map

The results of the accuracy value using the confusion matrix method, the overall accuracy (OA) value on the land use map was 85% with a kappa index of 0.78. When viewed from the kappa index value, the land use map was still acceptable, which was in the range of 85% -89%. The overall accuracy (OA) value showed the overall value of the resulting land use map, but it had not shown the accuracy of each land use class category yet. In addition to knowing the overall accuracy value, it was also necessary to calculate producer accuracy (PA) and user accuracy (UA) which aimed to determine the value in each land use class. Producer Accuracy (PA) had a relationship with omission errors in the form of omissions,in which the existing objects did not match or undescribed on map. Meanwhile, User Accuracy (UA) was related to commission error in the form of additions as objects depicted on the map did not match conditions in the field or undescribed. Out of 9 land use classes, open land objects and shrubs had the lowest accuracy for both Producer Accuracy and User Accuracy. The land use class of open land and shrubs showed 66.67% (producer accuracy) and 50% (user accuracy). Meanwhile, the highest accuracy was on the object of using paddy fields with a value of 96.36% (producer accuracy) and 94.6% (user accuracy). The low accuracy of open land objects and shrubs both on producer accuracy and user accuracy was due to the presence of mixed garden objects mostly associated with yards in settlements and paddy fields. Furthermore, there had been influence from an image used as the main data in extracting land use, by using Sentinel 1A imagery radar data classified as medium resolution, 20 meters resolution allowingmixpixels on the object. Moreover, there were factors on ROI determination to decide land cover classes and land use matching processes between land cover classification results and landform. Errors in determining ROI affected the pixels taken, in which the pixels taken were supposed to be homogeneous and unmixed with pixels of other objects, causing mixed pixels in the ROI samples used in extracting the land cover.Completion of land use with the matching method between land cover and landform wouldaffect on the resulting land use map.

The results of the land use classification obtained the area of each class. Out of 8 land use classes in the research area, the forest land use class dominated in area of 30,285ha (43.82%), which was almost half of the research area was forest area. The reason of this was the northern to eastern parts of Padang City are structural landforms, such as mountains to synclinal hills which are part of row onBarisan hills. Additionally, the use of mixed garden land covering an area of 11,570 ha (16.4%), shrubs 7,457 ha (10.79%), rice fields 7,251 (10.49%), settlements 4,296 ha (6.22%), open land 4,236 ha

(6.13%), fields/moor fields 3,495 ha (5.06%) and rivers 523 ha (0.76%).

4. CONCLUSIONS

The resultsofcurrent studyto map landuse in thecityof Padang hadtheaccuracyoflanduseusingtherandomforestmethodwhichwasquitefeasiblewithanoverallaccuracy valueof 85% with kappa valueof 0.78. This study still hadweaknessesduetotheinfluenceofpixel mix on ROI sampling in extractinglanduse. Hence,accuracywasneeded in determiningland use/land coversamplesathomogeneouspixelstominimizeerrors. In addition to this, it was found theeffectofnoise in theformofsaltandpepperonSentinel 1A ImageryRadar impactingtheclassificationresults. The useof radar imagery in the studywasquitedifficulttodistinguishthelanduseofwaterbodiesandwetlandssincetheyhadthe similarbackscattervalue.

5. ACKNOWLEDGMENTS

The authorwouldliketoextend her deepest gratitude to lecturers of Remote Sensing Program, Gadjah Mada University Yogyakarta, BappedaPadang, https://search.asf.alaska.edu/#/, and Ina-Geoportal Geospatial Information Agency as data provider for this study.

6. REFERENCES

[1] Danoedoro, P., Hidayati, I. N., &Widayani, P. (2010). The effect of scales in the analysis of agricultural land- use fragmentation based on satellite images of Semarang area , Indonesia. *Seaga*, 1–17.

[2] Danoedoro, P. (2015). PengaruhJumlah dan MetodePengambilanTitikSampelPengujiterhadap Tingkat AkurasiKlasifikasi Citra Digital PenginderaanJauh. *Simposium Nasional Sains Geoinformasi Ke-4, November 2015*.

[3] ESA. 2015. Sentinel-2 User Handbook. *ESA Standard Document User Handbook*. European Space Agency.

[4] Onojeghuo, A. O., Blackburn, G. A., Wang, Q., Atkinson, P. M., Kindred, D., &Miao, Y. (2018). Mappingpaddyricefieldsbyapplyingmachinelearningalgorithmstomulti-temporal sentinel-1A andlandsat data. *International JournalofRemoteSensing*, *39*(4), 1042–1067. https://doi.org/10.1080/01431161.2017.1395969

[5] Rahayu, M. I., &Sambodo, K. A. (2013). Random forest classification of jambi and south sumatera using alospalsar data. *34th Asian Conference on Remote Sensing 2013, ACRS 2013*, *4*(2), 3255–3262.

[6] Sambodo, K. A., Rahayu, M. I., Indriasari, N., &Natsir, M. (2014). KlasifikasiHutan-Non Hutan Data AlosPalsarMenggunakanMetode Random Forest. *Prosiding Seminar Nasional PenginderaanJauh 2014*, 120–127.

[7] Son, N. T., Chen, C. F., Chen, C. R., & Minh, V. Q. (2018). Assessment of Sentinel-1A data for rice crop classification using random forests and support vector machines. *Geocarto International*, *33*(6), 587–601. https://doi.org/10.1080/10106049.2017.1289555

[8] Torbick, N., Chowdhury, D., Salas, W., & Qi, J. (2017). Monitoring Rice Agriculture across Myanmar Using Time Series Sentinel-1 Assisted by Landsat-8 and PALSAR-2. *Remote Sensing*, *9*(2), 119. https://doi.org/10.3390/rs9020119

[9] Wang, H., Magagi, R., Goïta, K., Trudel, M., McNairn, H., &Powers, J. (2019). Cropphenologyretrieval via polarimetric SAR decompositionandRandomForestalgorithm. *RemoteSensingofEnvironment*, *231*(May). https://doi.org/10.1016/j.rse.2019.111234

[10] Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., &Feng, M. (2019). Evaluationofthreedeeplearningmodelsforearlycropclassificationusing Sentinel-1A imagerytimeseries-a case study in Zhanjiang, China. *RemoteSensing*, *11*(22). https://doi.org/10.3390/rs11222673