JGISE Vol. 7 No. 2 (2024), pp. 127 - 132 | https://doi.org/10.22146/jgise.94152

JGISE
Journal of Geospatial Information Science and Engineering
ISSN: 2623-1182 | https://jurnal.ugm.ac.id/jgise

Sentinel-2 satellite image processing using machine learning algorithms of the Manombo nature reserve

Valérien Eugène Tsaramanana [1,2,3], Arisetra Razafinimaro[1,2,3], Aimé Richard Hajalalaina [1,2,3],

[1] School of Management and Technological Innovation, University of Fianarantsoa, Madagascar
[2] Laboratory of Computer Science and Mathematics Applied to Development, University of Fianarantsoa, Madagascar
[3] Computer Science, Geomatics, Mathematics and Applications, Host Team, Fianarantsoa, Madagascar

Corresponding : Valérien Eugène Tsaramanana **| Email :** zoherval@gmail.com

Diterima (*Received*): 18/Feb/2024 Direvisi (*Revised*): 15/Dec/2024 Diterima untuk Publikasi (*Accepted*): 16/Dec/2024

ABSTRACT

This paper is based on the fields of satellite image processing and analysis using Sentinel-2 satellite images with machine learning algorithms under Google Earth Engine for the study of land cover evolution in the Manombo Madagascar, nature reserve. The objectives of the study are to identify the elements that occupy the land in the reserve. During our experiments, we compared the best machine learning algorithm using CART, Random Forest, Naive Bayes, SVM to determine the best machine learning algorithm for our Sentinel-2 data. So, we have proposed a methodology to do the treatment and, in the end, we have treatment results. From our treatments, we can conclude that the use of Random Forest classifier gave the most accuracy on the correct classification.

Keywords: Land use, Sentinel2, Supervised Classification, machine learning, satellite images.

1. Introduction

This work is part of the promotion of the use of Sentinel-2 satellite images for the study and monitoring of the evolution of land use. Researchers are interested in this type of image for the study of land cover in their respective countries, like (Inglada, 2016), which proposes the land cover map of France in 2016, (Akodéwou et al, 2019) which carries out the study of land cover in and around the Togodo protected area in Togo, (Delalay et al., 2019) who map the land use and land cover of a mountainous region of Nepal, (Ayoubi, 2017) who maps the land cover of Reunion Island, and finally (CIRAD, 2018) who maps the land cover of the Antananarivo Madagascar agglomeration using Sentinnel-2 with Landsat8. This leads us to reflect on the valorization of the potential of Sentinel-2 images for the monitoring study of the land cover of the Manombo nature reserve in Madagascar. In our processing approach, we have focused on the use of different classification algorithms based on supervised machine learning, which are still little exploited in Madagascar. Algorithms based on this technique use a variety of sources of inspiration, ranging from probability theory to geometric intuitions and heuristic approaches (Djaloul, 2017). The machine learning technique makes it possible to learn automatically from past data and experiences, and it seeks to best solve a given problem (Ah-Pine, 2019). These algorithms compromise supervised classifiers: CART (Classification And Regression Trees) (Breiman et al., 1984), Random Forest (Breiman, 2001), Naive Bayes (Rish, 2001), SVM (Weston and Watkins, 1998).

2. Data, methods and tools

2.1. Presentation of methods and techniques by machine learning

Machine learning refers to the development, analysis, and implementation of methods that allow a machine to evolve through a learning process, and thus to perform tasks that are difficult or impossible to perform by more traditional algorithmic means (Bessette, 2016). It then presents two main categories of learning: supervised and unsupervised (Soofi et al., 2017). Technically speaking, the terms supervised and unsupervised learning refer to

whether the raw data used to train the algorithms has been pre-labeled or unpre-labeled (Lawton G, 2020).

In this case, we used a supervised approach throughout this study. We hear more about supervised learning, as it's usually the last step in building an algorithmic model. This term includes techniques to facilitate image recognition and obtain better predictions. Using supervised learning, we manually create the necessary samples, called areas of interest. Thus, this phase produces a prediction model necessary for the classification phase, which consists of using the acquired prediction model. In this phase, new data are applied, which will be the optical satellite images to be classified to produce the classification. Then, validation confirms whether the classification results are acceptable or not. Figure 1 illustrates this method.

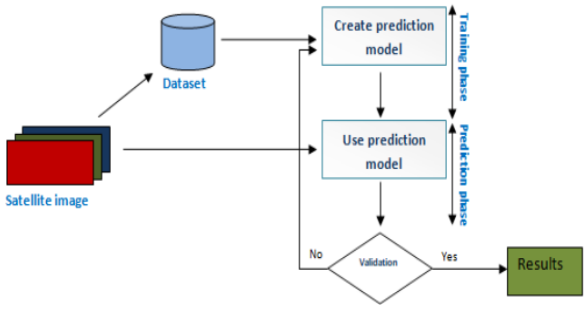

Figure 1 Satellite image classification methods using machine learning (Razafinimaro et al., 2022)

2.2. Study area

Manombo is located in the southeastern part of Madagascar, lying between 22°58' - 23°08' South and 47°38 - 47°48' East, Ankarana Miraihina Commune, Farafangana District and Southeast Region of Madagascar. Manombo is a special reserve located 26 km from Farafangana following the RN12 linking Farafangana to Vangaindrano. It has 04 contiguous rural communes: Ankarana Miraihina, Iabohazo in the west, Mahabo Mananivo in the south, Manambotra Atsimo in the north. The reserve covers 5,320 ha (divided into two plots). The climate is humid and hot, with an annual rainfall of about 2,500 mm and an average temperature of around 23°C.

2.3. Presentation of data used

The Sentinel-2 satellites provide views in 13 spectral bands in visible and near-infrared light with a resolution between 10 and 60 meters, but in this study, we used 10m resolution Sentinel-2 data with Level 1C processing provided by GEE, bands 2, 3, 4 and 8 were used. These data were orthorectified and radio-corrected to provide reflectance values at the top of the atmosphere. Table 1 shows the Sentinel-2 characteristics with a resolution of 10m.

Table 1: Sentinel-2 10m resolution characteristics

Sentinel-2 Bands	***Spatial resolution (m)***	***Wavelength (nm)***
B 2 – Bleu	*10*	*492.4*
B 3 – Vert	*10*	*559.8*
B 4 – Rouge	*10*	*664.6*
B 8 – NIR	*10*	*832.8*

2.4. Tools

Google Earth Engine[1] (GEE) is a cloud-based platform for analyzing environmental data on a global scale. This platform is a well-known cloud computing solution in the geospatial field. This software provides registered users with access to most of the free and open access data catalogs, including full-resolution data from MODIS, Landsat, Sentinel-1 and Sentinel-2, stored on Google's cloud storage infrastructure.

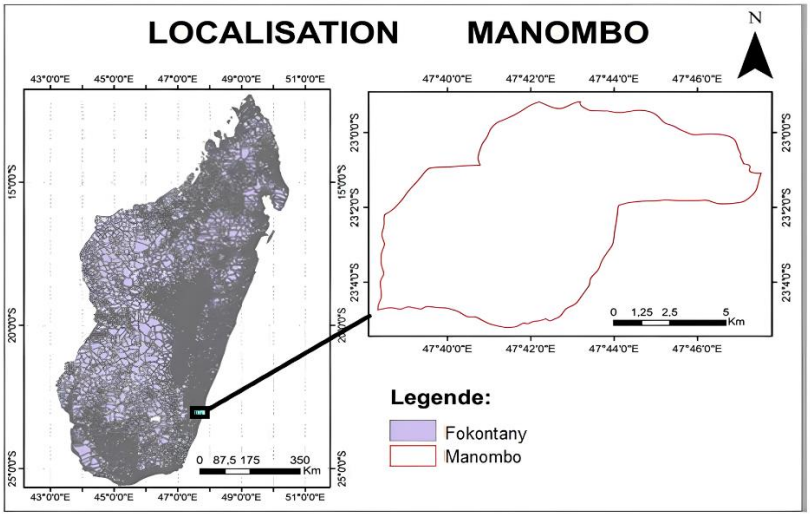

Figure 2 Manombo Reserve Location.

[1] https://earthengine.google.com/

2.5. Classification algorithms

Throughout the processing, we used the Classifier package that handles classification supervised by machine learning algorithms running in Google Earth Engine. These classifiers include CART, Random Forest, Naive Bayes, and SVM.

2.5.1. CART

CART (Classification And Regression Trees) refers to a statistical method, which constructs predictors per tree in both regression and classification. It corresponds to two very distinct situations depending on whether the variable to be explained, modelled or predicted is qualitative (discrimination) or quantitative (regression). The first decision tree classification algorithms are old. The two most significant works were the creation of CART (Breiman, 1984).

2.5.2. Random Forest

Random Forest is a supervised learning technique that combines an aggregation technique, BAGGING, and a particular technique for inducing decision trees. It is an ensemble learning algorithm that combines multiple classifications of the same data to produce a more accurate classification than other forms of decision trees (Cutler, 2007).

2.5.3. Naive Bayes

The naïve Bayesian classifier is a supervised learning method that is based on a strong simplifying assumption: the descriptors (Xj) are two by two independent conditionally to the values of the variable to be predicted (Y) 1. Yet, despite this, it proves to be robust and effective. Its performance is comparable to other learning techniques. The basic idea of classification comes from Bayes' formula:

$$P\,(CjX) = P\,(CP)P\,(X\,(X)\,jC). \qquad (1)$$

Since the attached conditional probability P *(XjC)* is difficult to estimate, the naïve version (later called NB) of classification is used. The probability of the class becomes in this case (Salperwyck et al., 2014). The naïve Bayesian classification is a type of simple probabilistic Bayesian classification based on Bayes' theorem with a high (so-called naïve) independence of assumptions. It implements a naïve Bayesian classifier, or naïve Bayes classifier, belonging to the family of linear classifiers. A Bayesian classifier is based on a probabilistic approach employing Bayes' rule. Let P *(Ci)* be the a priori probability of a class Ci, P *(x)* the probability of observing a characteristic vector x, and P *(xjCi)* the probability of observing the vector x knowing that the class is *Ci*. Bayes' rule can then be used to calculate the posterior probability of the class Ci when x is observed (Chouaib, 2011):

$$P(C_i/x) = \frac{P(x/C_i)P(C_i)}{\sum_j P(x/C_i)P(C_i)} \qquad (2)$$

In practice, since the denominator of Bayes' formula does not depend on *Ci*, we are only interested in the numerator. The probabilities P *(Ci)* of each class as well as the distributions P *(xjCi)* must be estimated beforehand from a training sample.

2.5.4. SVM

The Support Vector Machine (SVM) is a learning algorithm that produces a linear classifier (Pascal, 2009). The SVM belongs to the category of linear classifiers (which use a linear separation of data), and which has its own method of finding the boundary between categories. SVM is a two-class classification method that attempts to separate positive examples from negative examples in the set of examples. The method then looks for the hyperplane that separates the positive examples from the negative examples, ensuring that the margin between the closest to the positives and the negatives is maximized. This ensures that the principle is generalized, as new examples may not be too similar to those used to find the hyperplane, but may be located on either side of the boundary. The interest of this method is the selection of support vectors that represent the discriminant vectors by which the hyperplane is determined. The examples used when searching for the hyperplane are then no longer useful and only these carrier vectors are used to classify a new case, which can be considered an advantage for this method.

2.6. Treatment methodology

For the processing of satellite images under GEE (Shelestov et al., 2017; Mutanga et al., 2019), we used a method shown in Figure 3. First, we used input data as Sentinel-2 images; the study area and the field data, then, at the processing, we made corrections to the images, so, we filtered the images by dates, and masked the clouds using the cirrus band, then we mosaiced and cut the image with the study area, selected the 10m resolution bands, collected samples, chose the algorithm, and at the output, we have classified images with statistics and classification accuracy.

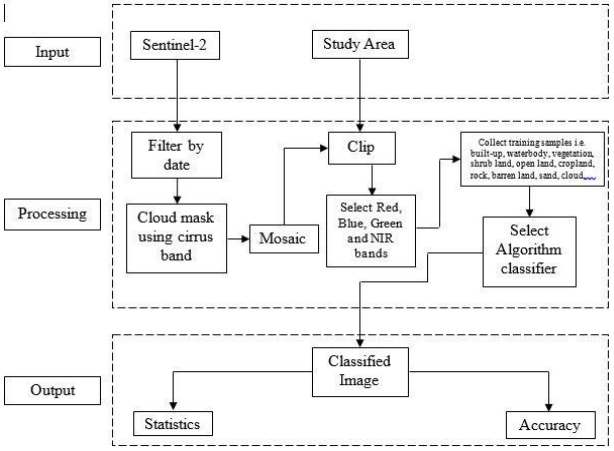

Figure 3 Proposed image processing methodology under GEE

3. Results and Discussions

From all our experiments, we have classified images and compared all the classifiers of the machine learning against its Kappa index and classification accuracy.

The land cover mapping of our study area is shown in Figure 4, 5, 6 and 7. Qualitative assessment of interpretation accuracy was performed using the Kappa index determined from a confounding matrix.

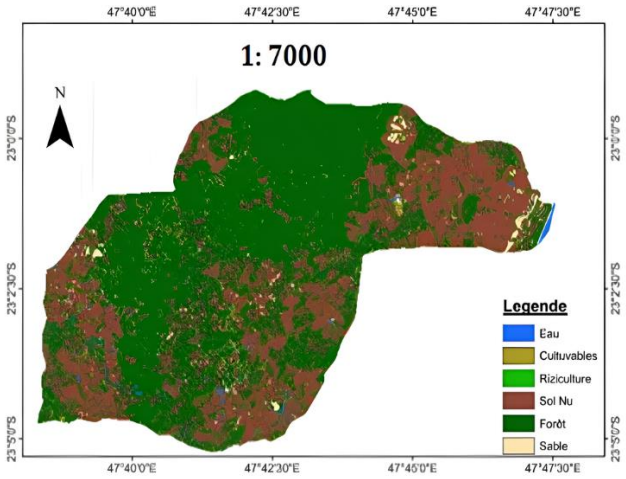

Figure 4 Classification result using Classification And Regression Trees (CART) classifiers

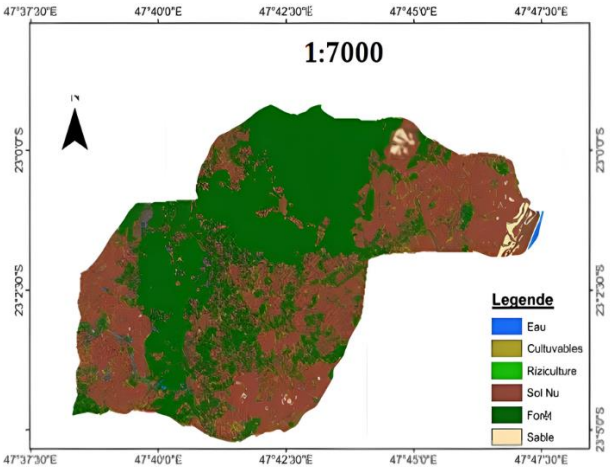

Figure 6 Classification result using Support Vector Machine (SVM) classifiers

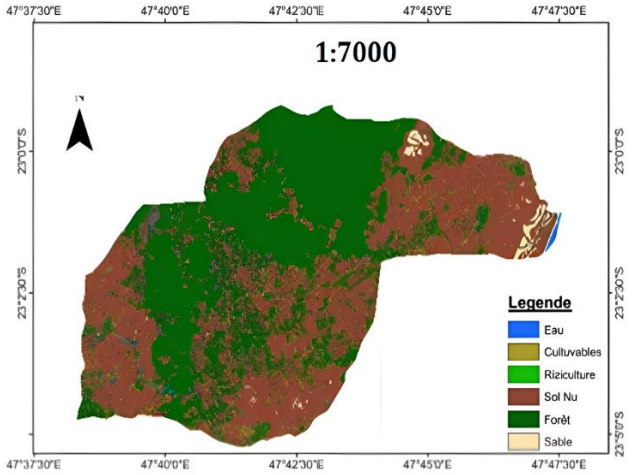

Figure 5 Classification result using Random Forest (RF) classifiers

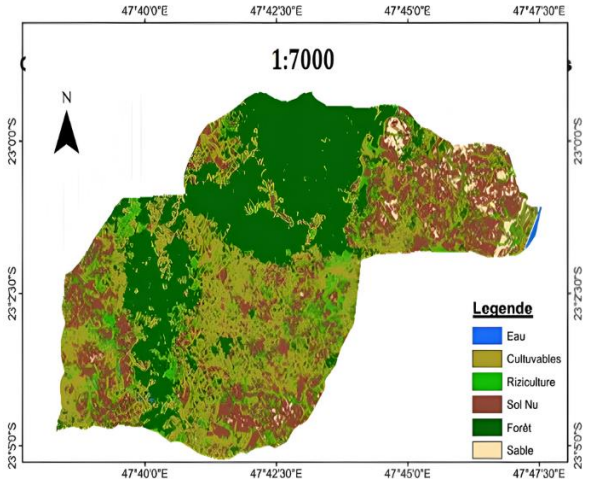

Figure 7 Classification result using NaiveBayes (NB) classifiers

After obtaining the results of the classifications, we obtained confounding matrices for each treatment. That is, it's an error matrix for qualitative data. It is a tool used to measure the quality of a classification system. Thus, each column in the matrix represents the number of occurrences of an estimated class. The confounding matrix is a tool used to evaluate the performance of a classification model, while the kappa coefficient is a measure of the agreement between observed classifications and those predicted by a model. The confusion matrix is a comparison of the classification with reality. They are presented in tables 2, 3, 4 and 5.

Table 2 confusion matrix for Classification And Regression Trees (CART) classifier

	Water	Cultivable land	Rice growing	**B**arren land	Forest	Sand
Water	187	1	0	3	2	0
Cultivable land	0	262	0	13	16	1
Rice growing	0	0	44	18	3	0
barren land	0	0	0	4006	29	1
Forest	0	9	0	103	12665	0
Sand	0	0	0	2	1	323

Table 3 confusion matrix for Random Forest (RF) classifier

	Water	Cultivable land	Rice growing	barren land	Forest	Sand
Water	239	0	3	4	7	0
Cultivable land	0	140	2	61	3	0
Rice growing	2	1	27	8	0	0
barren land	2	74	31	8657	65	0
Forest	4	1	0	43	22304	0
Sand	0	0	0	1	0	733

Table 4 confusion matrix for Support Vector Machine (SVM) classifier

	Water	Cultivable land	Rice growing	barren land	Forest	Sand
Water	143	0	2	0	1	0
Cultivable land	0	67	0	67	33	0
Rice growing	2	0	36	1	71	0
barren land	0	0	3	1146	53	0
Forest	0	0	10	44	19602	0
Sand	0	0	0	2	0	437

Table 5 confusion matrix for Naive Bayes classifier

	Water	Cultivable land	Rice growing	barren land	Forest	Sand
Water	238	6	15	1	1	1
Cultivable land	0	9	0	0	0	0
Rice growing	0	29	79	51	0	9
barren land	0	498	690	7691	0	1091
Forest	0	142	1	0	19021	0
Sand	0	0	0	3	0	337

After having confusion matrices, we have the Kappa indices; that is, an index of relative accuracy. It is a quality estimator that accounts for row and column matrix errors. It ranges from 0 to 1. Kappa is the most well-known index for evaluating a supervised classification. The Kappa coefficient is a quality estimator that accounts for errors in rows and columns.

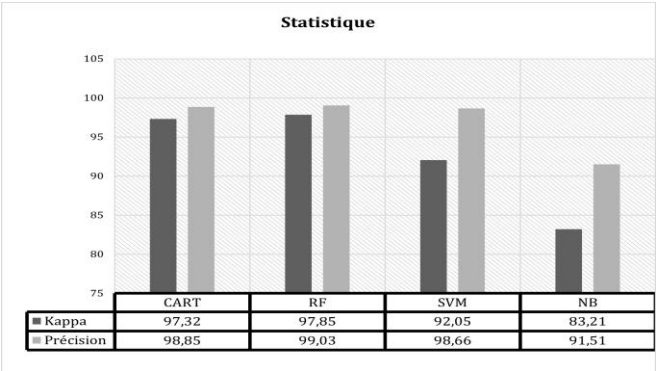

Figure 8 Kappa indices and Machine Learning classifier classification accuracies

Our work is based on the processing of satellite images for the land use of the Manombo Nature Reserve. In our case, we used high-resolution optical images, Sentinel-2 with the resolution of 10m, in order to obtain classification results with validations and a land cover map. The results of satellite image processing are quantitatively analyzed based on the confounding matrix, the Kappa index, and the overall accuracy; and qualitatively with a visual evaluation of classifications to estimate the quality of the results and to highlight errors that cannot be detected in the overall statistical evaluation. The estimation of accuracy depends on the validation set, which does not always indicate a ground truth. For this, we have Kappa classifiers of CART, RF, NAÏVE BAYES and SVM respectively 97,32; 97,85; 92.05 and 83.21. With clarifications of classifications 97.58, 98.33, 83.21 and 97.03.

By comparing all the satellite image classification results we studied, we observed that the use of the Random Forest algorithm was the best algorithm in our classification with our Sentinel-2 data.

4. Conclusion

In conclusion, our study aims to promote the use of Sentinnel-2 satellite images for the study of land cover in the Manombo nature reserve in Madagascar using machine

learning algorithms under Google Earth Engine. Throughout our processing, we determined the best classifier in the Sentinel-2 satellite data the one we used for land cover. By comparing all the satellite image classification results from our experiments, we observed that using the Random Forest algorithm is the best algorithm for our satellite image processing classification with our Sentinel-2 data (Rakotoarison et al, 2021) (Dupuy et al. 2018)

5. Conflict of Interest Statement

The authors hereby declare that they have no conflict of interest in the research, analysis or publication of this article.

6. References

Ah-Pine J. (*2019). Machine Learning,* University of Lyon 2

Akodéwou, A., Oszwald, J., Akpavi, S., Gazull, L., Akpagana, K., & Gond, V. (2019). *Problem of invasive plants in southern Togo (West Africa): contribution of landscape systemic analysis and remote sensing*. Biotechnologie, Agronomie, Société et Environnement/Biotechnology, Agronomy, Society and Environment, 23.

Ayoubi S. (2017). *« Reunion | Mapping the island's land cover | GeoDev".* Available on: http://www.theia-land.art-geodev.fr/la-reunion-cartographie-de-loccupation-des-sols-de-lile/

Breiman, L. (2001). *Random forests*. *Machine learning*, 45, 5-32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Cart*. *Classification and regression trees*.

CIRAD. (2018). *"Land Use Mapping 2017 Antananarivo",* July 12, 2018 9:52 a.m., License: CC BY: Attribution (CC BY, "LEGENDE Project | GeoDev". https://www.theia land.art-geodev.fr/projets/legend/

Delalay, M., Tiwari, V., Ziegler, A. D., Gopal, V., & Passy, P. (2019*). Land-use and land-cover classification using Sentinel-2 data and machine-learning algorithms: operational method and its implementation for a mountainous area of Nepal*. Journal of Applied Remote Sensing, 13(1), 014530-014530.

Dupuy, S., Le Mézo, L., & Gaetano, R. (2018). *Reunion Island: 2017 land use map*

Inglada, J. (2016). *Mapping of land cover from optical images. Remote Sensing Observation of Continental Surfaces: Agriculture and Forestry*, ISTE Editions, London.

Lawton G. (2020). *Supervised and unsupervised learning*, https://www.lemagit.fr/conseil/Apprentissage-supervise-et-non-supervise-les-differencier-et-les-combiner , Published on: 14 Oct 2020

Mutanga, O., & Kumar, L. (2019). Google Earth Engine Applications. Remote Sensing, 11(5), 591. https://doi.org/10.3390/rs11050591

Rakotoarison, T. R., Hajalalaina, A. R., & Safidinirina, E. N. (2021). Forest Dynamics with Sentinel 2 in Antanambe between 2005 and 2016 with the Snap Tool. *Advances in Remote Sensing*, *10*(3), 92-101.

Razafinimaro et al. (2022) *Land cover classification based optical satellite images using machine learning algorithms*, International Journal of Advances in Intelligent Informatics, Vol. 8, No. 3, November 2022, pp. 362-380, https://doi.org/10.26555/ijain.v8i3.803

Rish, I. (2001, August). *An empirical study of the naive Bayes classifier*. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).

Salperwyck, C., Lemaire, V., & Hue, C. (2014). Classifieur naïf de Bayes pondéré pour flux de données. In EGC (pp. 275-286).

Santosa, P. B. (2016). Evaluation of satellite image correction methods caused by differential terrain illumination. *Jurnal Forum Geografi*. Vol. 30, No. 1 (2016). https://doi.org/10.23917/forgeo.v30i1.1768

Shelestov, Andrii (02/2017). "Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping". Frontiers in Earth Science (2296-6463), 5 , p. 232994.

Soofi, A. A., & Awan, A. (2017). *Classification techniques in machine learning: applications and issues*. Journal of Basic & Applied Sciences, 13(1), 459-465, https://doi.org/10.6000/1927-5129.2017.13.76

Weston, J., & Watkins, C. (1998). *Multi-class support vector machines* (pp. 98-04). Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.