Land cover classification based optical satellite images using machine learning algorithms

Arisetra Razafinimaro [a,b,c,1,*], Aimé Richard Hajalalaina [a,b,c,2], Hasina Rakotonirainy [a,b,c,3], Reziky Zafimarina [d,4]

[a] Ecole de Management et d'Innovation Technologique(EMIT), University of Fianarantsoa, Madagascar
[b] Laboratoire d'Informatique et des Mathématiques Appliqués au Développement (LIMAD),Fianarantsoa, Madagascar
[c] Informatique-Geomatique et Mathématiques Appliquées (IGMA),Fianarantsoa, Madagascar
[d] Higher Polytechnic School of Antananarivo, University of Antananarivo, Madagascar
[1] tokyarisetra@gmail.com; [2] rhajalalaina@yahoo.fr; [3] hasinarakotonirainy@emit.mg; [4] zojaona@yahoo.fr
* corresponding author

ARTICLE INFO

Article history
Received February 28, 2022
Revised June 27, 2022
Accepted July 18, 2022
Available online November 30, 2022

Keywords
Satellite images optical
Machine learning
Knowledge
Land cover
Reliability

ABSTRACT

This article aims to apply machine learning algorithms to the supervised classification of optical satellite images. Indeed, the latter is efficient in the study of land use. Despite the performance of machine learning in satellite image processing, this can change but depends on the nature of the satellite images used. Moreover, when we use the satellite, then the reliability of one classifier can be different from the others. In this paper, we examined the performance of DT, SVM, KNN, ANN, and RF. Analysis factors were used to investigate further their importance for Sentinel 2, Landsat 8, Terra Modis, and Spot 5 images. The results show that the KNN showed the most interesting accuracy during the analysis of medium and low-resolution images with spectral bands lower or equal to 4, with a higher accuracy of about 93%. The RF completely dominated the other analysis cases, where the higher accuracy was about 94%. The classification accuracy is more reliable with high-resolution images than with the other resolution categories. However, the processing times of high-resolution images are much higher. Moreover, higher accuracy was often achieved with more expensive processing times. Besides, almost all machine learning algorithms suffered from the Hugs phenomenon during the analyses. So, before the classification with machine learning, some preprocessing is needed.

1. Introduction

Over the last four decades, satellite image classification has become a rapid solution to map land cover and quantify land use [1]. Indeed, it is a part of the experimental sciences using computer algorithms. Due to its widespread application on various topics and the evolution of satellite images in terms of spatial and spectral resolution, their classifications recommend efficient algorithms. In general, these algorithms result in reliability, like a higher accuracy. Artificial intelligence's intervention with machine learning helps renew the way of general classifying data. Its algorithms respond mainly to the precision and time optimization of the treatments of voluminous data as satellite images have benefited. In other words, it is designed to overcome the limitations of classical algorithms such as maximum likelihood, minimum distance, and thematic mapper [2], [3].

Then, many experiments have been conducted by scientific researchers on these algorithms applied to various land-use themes. For example, Sesnie *et al.* [4] applied KNN to the Finnish forest inventory and then confirmed the effectiveness of this algorithm than the classical algorithms like the maximum

true likelihood. The study of [5] distinguished the Costa Rican rainforest's type and structure, which proved that the SVM is slightly superior to the RF. Also, [6] performed a distribution study of 35 plants, from the experience, RF proved that it gave a better result than the SVM and the ANN. Similarly, [7] conducted a multispectral satellite image processing of Landsat TM. The forest species classification of [8] in Australia showed that the SVM also produced a higher result.

It is undeniable that machine learning has transformed the science of satellite image classification. In addition, all comparisons between classical classification algorithms and machine learning have proven their reliability effectively. However, the lack of knowledge about their reliability for optical satellite images hinders the construction of a truly relevant method taking into account the image category. This provokes problems not only with the choice of algorithms adapted to the image used but especially with the formation of a generalized method for processing optical satellite images for land use study. For this reason, we will conduct in-depth analyses of supervised machine learning algorithms through optical satellite images: Sentinel 2, Landsat 8, Terra Modis, and Spot 5. To accomplish it, we will compare them using hypothesis testing, such as the study of the accuracy with the number of spectral bands to the method used with or without radiometric correction, the number of samples, their sensitivities to noise as well as their processing times. These are to build relevant syntheses and transform them into a knowledge base and experiments.

This article is organized as follows. In the second section, we will briefly discuss the methods. In this, we will generally talk about the machine learning method and the characteristics of satellite images to use. Then, in the third section, we will illustrate the results obtained, to which we will give interpretations. Furthermore, we will discuss each result. In the end, we will give the conclusion and the corresponding perspectives to this article.

2. Method

2.1. Machine learning approach

In this study, we use a method based on machine learning classification. Indeed, machine learning is a derivative of artificial intelligence. Then, it proposes three categories of learning: supervised, unsupervised and semi-supervised machine learning [9]. It should be noted that we particularly use the supervised mode throughout this study. This consists of two phases, named the training phase and the testing phase [10]. On the one hand, the learning phase ensures the construction of the prediction model using the data samples to which the classification model and the expected label of the final result are specified. With supervised learning, we manually create the necessary samples, also called areas of interest. Thus, this phase produces a prediction model necessary for the next phase.

On the other hand, the test phase, called the classification phase, consists of using the prediction model acquired in the previous phase. In this phase, new data are applied, which will be the optical satellite images to be classified to produce the classification. Then, the validation confirms whether the classification results are acceptable or not. Fig 1 shows this method.

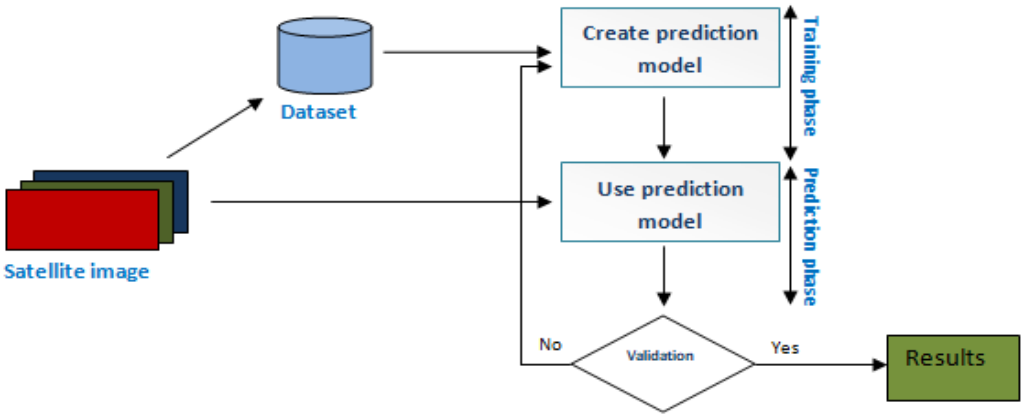

Fig. 1. Methods for classification images satellites using machine learning

This method will be applied during the verification of various assumptions. It includes assumptions related to the efficiencies of the classifiers to the optical satellite image resolutions, the number of spectral bands, and the preprocessing effects (image correction). Also, we will use this method to measure the processing times of each classifier and the sensitivity to the Hugs phenomenon. In order to verify this phenomenon, we will perform processing with 50, 100, 150, 200, 250, and 300 samples, which will be used during the creation of the prediction model.

2.2. Characteristics of satellites images used

In satellite image processing, the effectiveness of a classifier should depend on the type of image used. This can be caused by the type of image resolution and image size. Consequently, it is important to take sample images to verify this hypothesis. We will take an image from Sentinel 2, Landsat 8, Modis, and SPOT 5. These images are delivered by different satellites with different spatial resolutions and are fully adapted to the study of land cover.

2.2.1. A Landsat 8 OLI

In satellite image consists of 8 bands of 30m and a single panchromatic band at 15. Table 1 shows their characteristics.

Table 1. Characteristic of Landsat 8 OLI 8 spectral bands at 30m spatial resolution

N° band	Canal	Spectral band (nm)	Resolution (m)
B1	Arosol	0,433- 0,453	30
B2	Blue	0,450- 0,515	30
B3	Green	0,525- 0,600	30
B4	Red	0,630- 0,680	30
B5	NIR	0,845- 0,885	30
B6	SWIR 1	1,560- 1,660	30
B7	SWIR 2	2,100- 2,300	30
B9	Cirrus	1,360- 1,390	30

2.2.2. Terra Modis

Terra Modis image has a very low spatial resolution, but we take two bands with 250m spatial resolution in this study. Table 2 shows their characteristics.

Table 2. Characteristic of 2 Terra Modis spectral bands at 250m spatial resolution

N° band	Canal	Spectral band (nm)	Resolution (m)
B1	VIS	0.62-0.67	250
B2	NIR	0.84-0.87	250

2.2.3. Sentinel 2

A Sentinel 2 image is composed of 13 spectral bands, but in this study, we choose the 4 bands with 10m resolution particularly. Table 3 represents their characteristics.

Table 3. Characteristics of Sentinel 2 satellite image at 10 m of our study area

N° band	Canal	Spectral band (nm)	Resolution (m)
B2	Blue	492..4	10
B3	Green	559.8	10
B4	Red	664.6	10
B8	NIR	832.8	10

2.2.4. Spot 5 HRG 2

Like Sentinel 2, a Spot 5 HRG2 image also has 4 spectral bands with 10m spatial resolution. We will use these bands in this study. Table 4 shows their characteristics.

Table 4. Characteristic of SPOT 5 HRG 2 satellite image at 10m resolution

N° bande	Canal	Spectral band (nm)	Resolution (m)
B1	Green	0.50–0.59	10
B2	Red	0.61–0.68	10
B3	PIR	0.78–0.89	10
B4	MIR	1.58–1.75	10

We have combined the spectral bands in Table 1 to Table 4. As a result, their combinations have allowed us to produce multispectral images. First, the B2, B3, B4, and B8 bands described in Table 1 have given Fig 2(a). Then, the spectral bands B1, B2, B3, B4, B5, B6, B7, and B9 quoted by Table 2, have permitted us to produce Fig 2(b). Then, the spectral bands B1 and B2, represented by Table 3, have given us Fig 2(c).In the same way, the spectral bands B1, B2, B3, and B4, quoted in Table 4, have produced Fig 2(d).

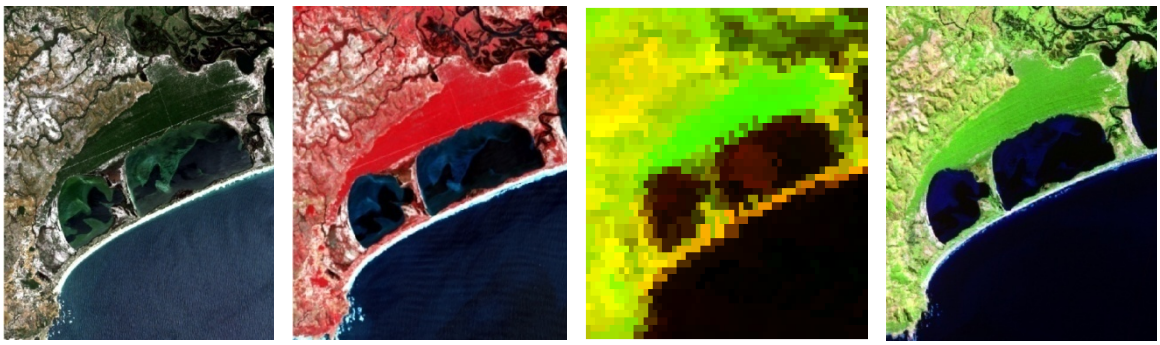

Fig. 2. Sentinel 2 image at 10m (a), Landsat 8 OLI at 30m (b), Terra Modis at 250m(c) and Spot 5 HRG at 10m (d)

2.3. Algorithmes of classification

The DT (Decision Tree) is one of the supervised machine learning methods. Indeed, this algorithm is considered a simple method because it is a nonparametric method [11]. That is, it does not require any assumption about the independence of features during classification. In addition, the DT has a structure based on a root node called the head node and then internal nodes, terminal nodes, or leaves as well. Each node has only one parent node but can have at least two descendant nodes. Note that the value of each leaf represents the classes that make the classification up. Moreover, the DT procedurally defines the classification by recursively data partitioning to be classified on the branches of the obtained tree or prediction model. It assigns a label for the dataset according to the observation based on the node of the leaf. Mathematically, each leaf represents a class. Its distribution is considered as follows:

$$P_j(C_k|x) \text{ [12]} \quad (1)$$

-The SVM (Support Vector Machines) or Wide Margin Separator is a machine learning algorithm discovered by [13], [14], [15]. Indeed, this algorithm looks for the weight vector w to separate the positive and negative examples of the sample data used during the learning phase [16]. Moreover, the SVM is an improvement of the perceptron algorithm invented by Rosenblatt in which the hyperplane is constructed especially for the signs and the distance calculation system of each point. On the one hand, the margin is obtained by the wayof the minimum distance calculation of the samples. On the other hand, the maximization of this margin leads to the method of large margin separators. Also, the SVM can solve the linear and nonlinear situation as the data size increases as well. According to [17], nonlinear data is the most used problem which is described as pattern:

$$\max L_D = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \quad (2)$$

Under the constraints:

$$\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*) = 0 \text{ et } \alpha_i, \alpha_i^* \in [0, C]$$

Whose weights are calculated by:

$$w = \sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)\Phi(x_i),$$

And the model by :

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\,(x_i, x)+\beta$$

-The KNN (K nearest neighbor) method is a supervised classification algorithm with a particularity compared to learning methods. It is based on the learning phase because it does not need this phase. It relies directly on the comparison of entities to classify and their characteristics to which distance calculation is performed. As a result, each entity to be classified assigned to the class meeting the basic principle of nearest neighbors [18]. The KNN is also categorized to the nonparametric method [19], [20]. It canalso be noted that the Euclidean distance is widely used to calculate the distance between the pixel and the reference pixel set [21] as well.

$$d_{ir} = \sqrt{\sum_{f=1}^{N_f} \delta_f(y_{f(i)} - y_{f(r)})}^{\,2} \quad (3)$$

hence $y_{f(i)}$ is the pixel value i for feature f and $y_{f(r)}$ is the value of the reference zone r for the characteristic f.

-The ANN (Artifical neural Network) is a supervised classification method modeling the basic concept of an animals nervous system to recognize patterns or objects [22]. Indeed, its basic architecture is based on the primitive function, which can discriminate classes. In addition, it applies the interconnections of neurons in the form of layer organization, where one layer has neurons. Moreover, the ANN has hidden nodes where each neuron in a layer can communicate with neurons in all adjacent layers. Besides, the number of input layers is greater than the number of output layers, where there is one neuron for each input variable, while one is intended for one output class. It is suitable to state that the number of neurons in the hidden layers and the fast creation of the hidden layers ensure the ability of the system to solve an input problem in a low time. According to [2], ANN can be slow to train, it's possible to produce non-optimal classification, and so easy to over-train. Mathematically, a neuron is described as a pattern [23]:

$$a = f(w_1 p_1 + w_2 p_2 + \cdots + w_R p_R + b) \quad (4)$$

where P1, P2...PR are the inputs, w1, w2...wR are the synaptic weights of neuron L, b are the bias,f is the activation function, and a is the neuron output.

-The Random Forest is a classification method created by Breiman in 2001[24], which gathers classifiers to improve the performance of the decision tree in terms of optimization. Then, this classifier assigns a highly special system. Because it uses a majority voting system to predict the classes, its subsets are created by itself. Each decision tree also is generated independently, without any pruning. Furthermore, this classifier employs an error estimation system known as OOB or "Out-Of-Bag" [2]. The new unlabeled input data is classified by all the decision trees created. Thus, each tree votes for class membership, while the membership class with the maximum number of votes will be selected.

3. Results and Discussion

3.1. Result

Different promising results have been obtained during our analysis, so we will interpret them by using different angles. These are in order to draw knowledge relevant to the use of optical satellite images.

3.1.1. Precision analysis

This analysis focuses on the use of machine learning algorithms for the classification of our optical satellite images. Indeed, this allows us to answer the question: "Which algorithms would give relevant results with higher accuracies depending on the image used? We will thus define 2 categories of treatments, named the first part is focused on the image classification of 2 bands for Modis and 3 bands for SPOT 5, Landsat 8 and Sentinel 2 on the machine learning algorithms DT, SVM, KNN, ANN and RF. In the second part, we will classify the 4 bands of SPOT 5, Landsat 8 and Sentinel 2.

3.1.1.1. Two (2) and Three (3) band analysis

During our first analysis of the machine learning algorithms applied to our Sentinel 2 image of 3 spectral bands (B2, B3, B4), we obtained their best results presented by Fig 3.

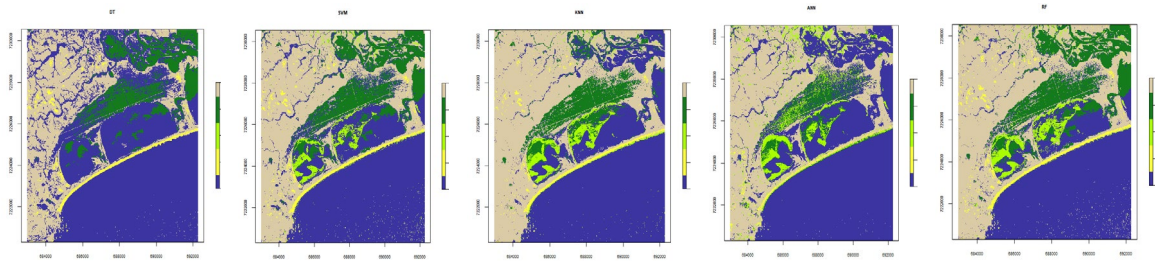

Fig. 3. Sentinel 2 3bands classification results with DT, SVM, KNN, ANN and RF

Then, we also applied the same principle and the same algorithm to the Sentinel 2 image, our Landsat 8 image of 3 spectral bands (B2, B3, B4). We obtained the results shown in Fig 4.

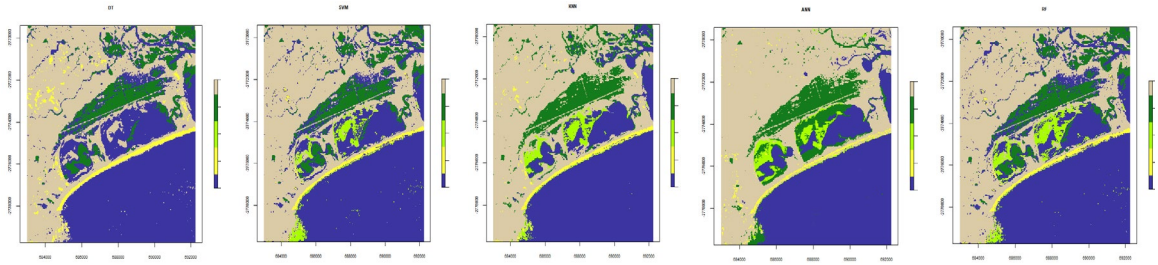

Fig. 4. Landsat 8 3bands classification results with DT, SVM, KNN, ANN and RF

Similarly, we analyzed the Terra Modis image of 2 spectral bands at 250 m spatial resolution on machine learning algorithms. These allow us to know each algorithm's efficiency in the face of an insufficient number of spectral bands and a low-resolution image. Fig 5 represents the best results acquired during the Terra Modis analyses.

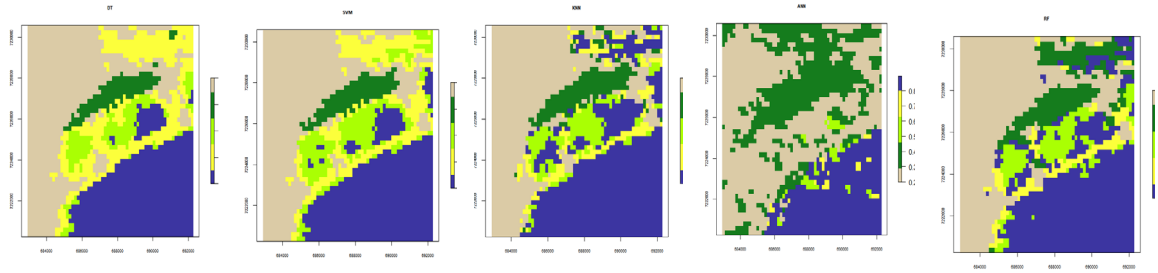

Fig. 5. Terra Modis classification results of 2 bands with DT, SVM, KNN, ANN and RF

Furthermore, we also apply machine learning algorithms to classify a SPOT 5 HRG 2 image of 3 spectral bands (B1, B2, B3) with a spatial resolution of 10m. The best results during the analyses are shown in Fig 6.

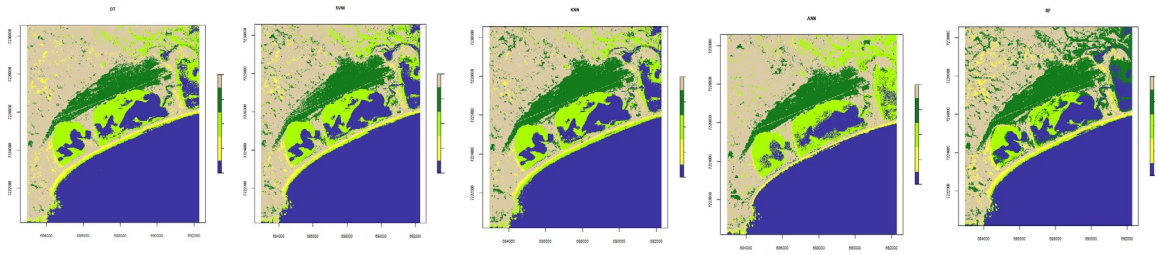

Fig. 6. Classification results of SPOT 5 HRG 2 of 3 bands with DT, SVM, KNN, ANN and RF

These results are associated with Kappa indices showing the precision on which their reliability is justified, and the differences between them are readable quantitatively. Table 5 shows the kappa index corresponding to each result of the algorithm as a function of the image served.

Table 5. Kappa index of the results obtained during the analysis of 2 and 3 spectral bands

Images	DT	SVM	KNN	ANN	RF
Sentinel 2	89.5	91	92	90.6	92.9
Landsat 8	89.2	91.1	91.8	90.3	91.1
Modis	88.5	90.1	91.2	88.8	91
SPOT 5	91.2	93.1	92.5	91.2	93.7

Table 5 allows us to promote our analysis by using the diagram. The latter illustrates the accuracy given by each Machine learning algorithm on each image. Fig 7 represents this diagram.

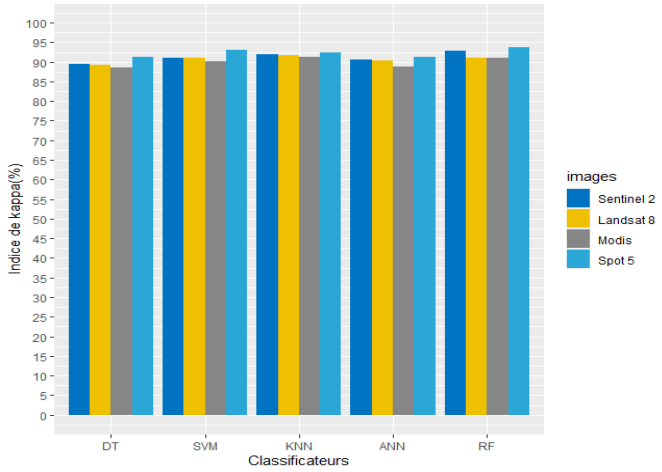

Fig. 7. Bar graph of the analysis of 2 and 3 spectral bands of optical satellite images

In general, each algorithm could give an accuracy above 90%. During the analyses on the Sentinel 2 image, it was the RF algorithm that gave an accuracy of almost 93%. Then, the KNN gave high accuracy on the Landsat 8 analyses with an accuracy of 92%. Besides, the KNN always found the best accuracy, 91.1%, with Terra Modis. Also, the RF showed an accuracy of almost 94% for the Spot 5 image. Moreover, the latter is largely superior to the accuracy obtained with other images.

3.1.1.2. analysis of four (4) bands

During the second type of analysis, we will check the efficiency of each Machine Learning classifier if it works with 4bands satellite images. Due to the availability of this number of bands having the same spatial resolution on Sentinel 2, Landsat 8 from the OLI sensor, and SPOT 5 from the HRG 5 sensor, we will use them. For the analysis of Sentinel 2, we especially use the spectral bands B2, B3, B4, and B8 because only these bands have the same spatial resolution of 10m. The best results during this analysis are shown in Fig 8.

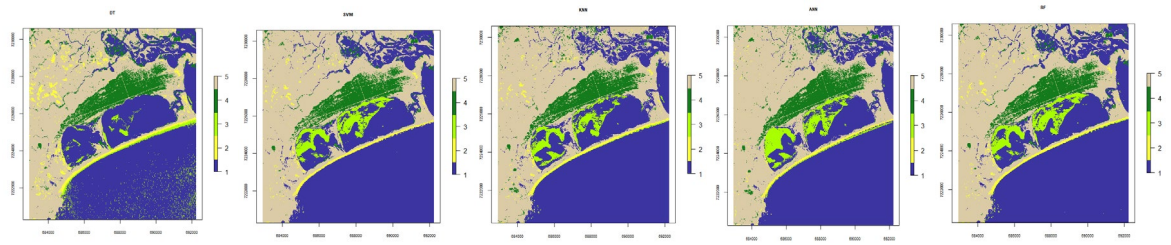

Fig. 8. Sentinel 2 4bands classification results with DT, SVM, KNN, ANN, and RF

Next, we will work on the first four bands of Landsat 8 to verify this reliability. We randomly selected the visible bands to be used, which are B1, B2, B3, and B4, without considering their efficiencies according to our objective. On the one hand, this traditional selection method may impact our results' reliability. On the other hand, it permits us to identify the best algorithm that does not perform. In any case, the best results obtained during the analyses are presented in Fig 9.

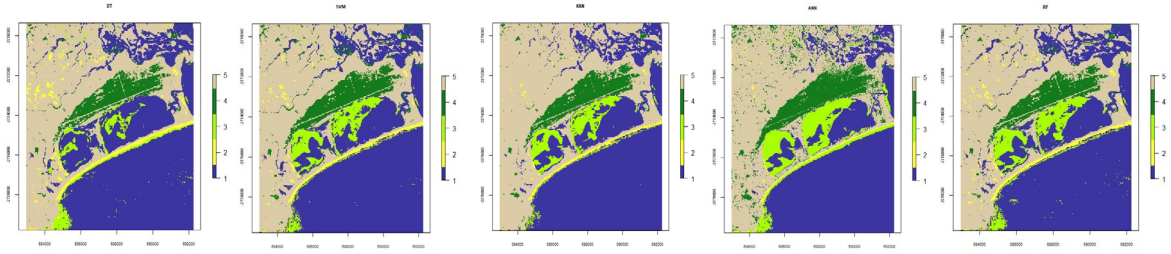

Fig. 9. Landsat 8 OLI 4bands classification results with DT, SVM, KNN, ANN, and RF

Then, we also analyze the SPOT 5 HRG 2 image using its first four bands, namely bands B1, B2, B3, and B4. These bands have particularly a spatial resolution of 10m. The best results acquired are presented in Fig 10.

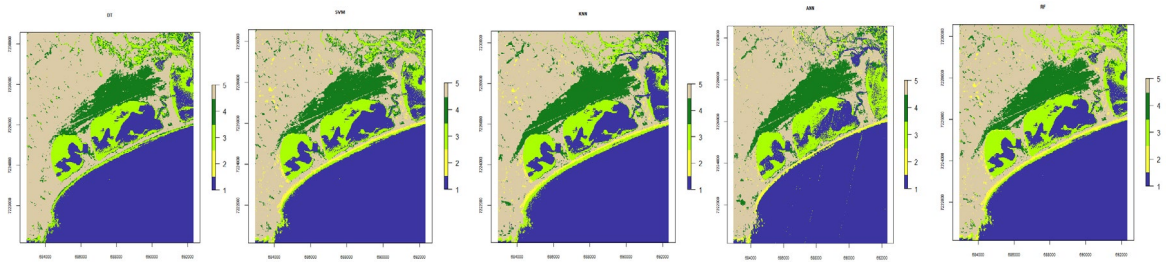

Fig. 10. Classification results of Spot 5 HRG2 4bands with DT, SVM, KNN, ANN, and RF

It is so familiarly noted that the best results of each algorithm are obtained following validations by using the google earth image, which obtains their kappa indices quantifying and their accuracy compared to the reality on the ground. Table 6 shows the kappa indices corresponding to each image and each classifier.

Table 6. Image Kappa Index during 4bands analysis with DT, SVM, KNN, ANN, and RF

Images	DT	SVM	KNN	ANN	RF
Sentinel 2	90.9	93.3	93.1	92.9	93.8
Landsat 8	90.5	92.5	92.9	92.3	92.6
SPOT 5	90.6	93.9	93	93	94.2

We have found a bar graph showing the nature of these values. This constantly compares each classifier's contribution to the optical satellite image used. Fig. 11 illustrates more knowledge of each algorithm's reliability during the analysis.

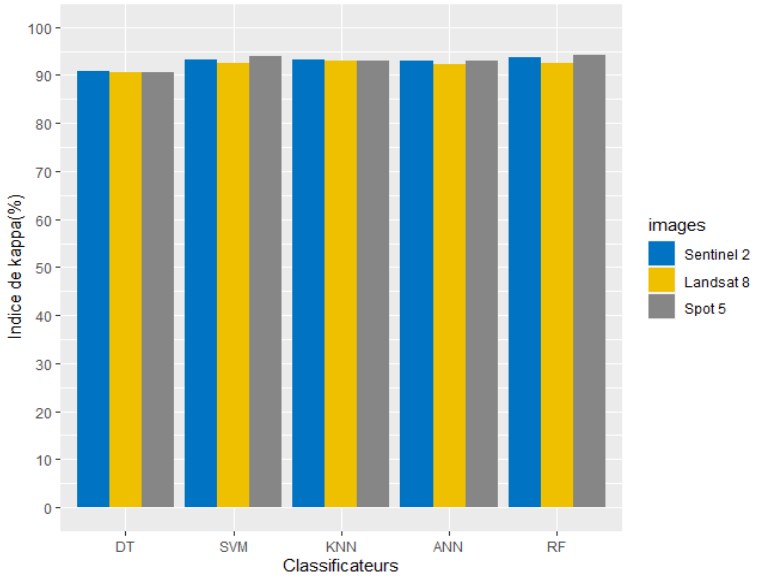

Fig. 11. Bar graph of the analysis of 4 spectral bands of optical satellite images

According to this diagram, all accuracies are above 92% except for DT, which is less than 91%. In addition, the RF is still the record holder on Sentinel 2 and SPOT 5 images, but we also found that the differences between the RF and SVM are smaller and insignificant. In addition, KNN is still the leader on Landsat 8, with a small difference between RF and SVM, and ANN is still in the middle.

3.1.2. Capacity analysis in terms of big data

This kind of analysis consists of checking each previously used machine learning algorithm on the point of: "Which algorithms would persist in the face of a large number of variables?".That is, we will measure the ability of DT, SVM, KNN, ANN, and RF on heavy, multi-band, multispectral data processing. To accomplish this, we will process the 8 spectral bands with a spatial resolution of 30m of Landsat 8 from the OLI sensor. Then, the best results will be compared to the best results obtained before. The analyses on each algorithm allow us to produce the results presented in Fig 12.

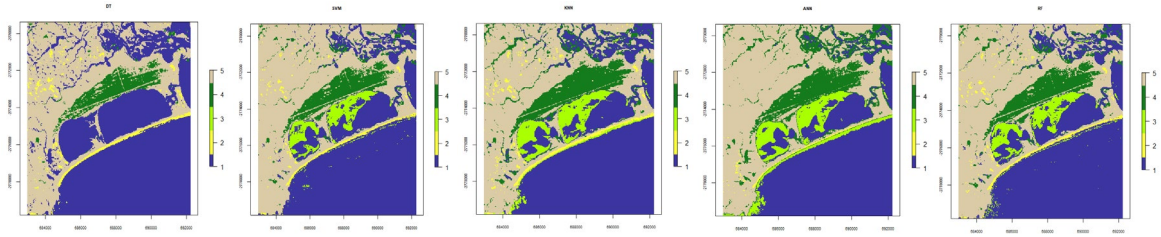

Fig. 12. Landsat 8 OLI classification results of 8 bands at 30m with DT, SVM, KNN, ANN and RF

Table 7 shows the Kappa index in relation to the number of bands and the algorithms used during the experiments. For the kappa index of 3 and 4 Landsat bands, we take the precisions on the previous results.

Table 7. Kappa index of Landsat 8 OLI 3.4 and 8 bands at 30m with DT, SVM, KNN, ANN and RF

Nbr bandes	DT	SVM	KNN	ANN	RF
3bandes	89.2	91.1	91.8	90.3	91.1
4bandes	90.8	92.5	92.9	92.3	92.6
8bandes	89	91.9	91.4	91.5	92.3

In order to illustrate not only the trend of these data presented in Table 7, but also to draw the maximum knowledge from their nature, we will draw a bar graph. Fig 13 shows the diagram reflecting the reliability of Landsat 8 OLI in 3 bands, 4 bands, and 8 bands.

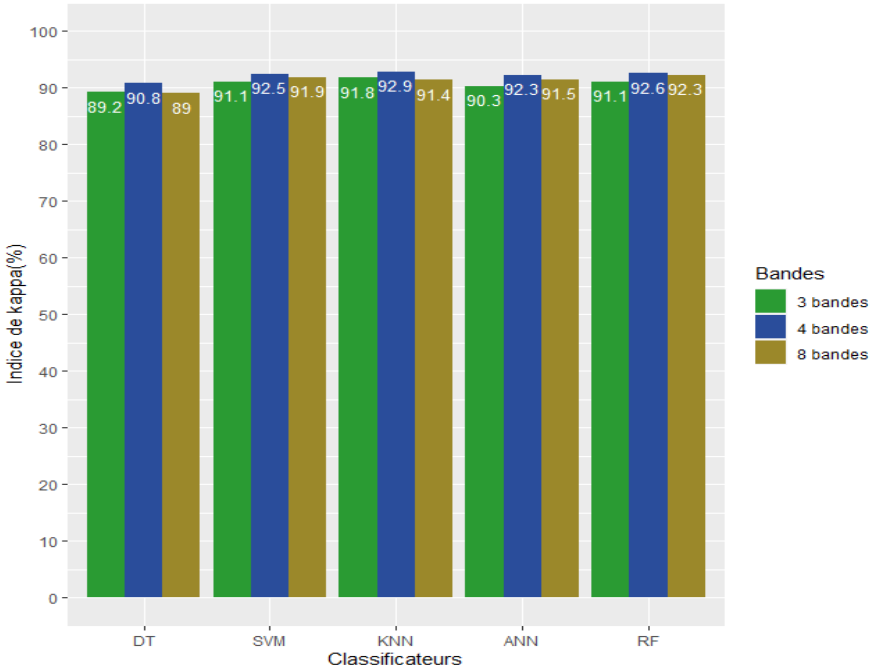

Fig. 13. Bar chart of the analysis of 3,4 and 8 spectral bands of Landsat 8 OLI satellite image

3.1.3. Analyze the contribution of preprocessing to the classifier

In this section, we will analyze whether the preprocessing impacts the performance of the Machine Learning classifier and answer the question: "Could the preprocessing change the efficiency of the classifier? To do this, we will take the case of a radiometric correction of type "DOS" by applying it to our Landsat 8 OLI image of 8 bands. Then, we will compare the results the previous results to extend the level of synthesis. Fig 14 shows the results during DOS correction classified.

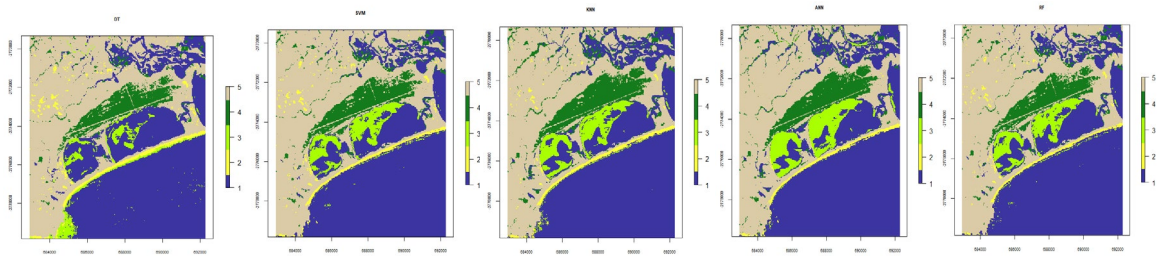

Fig. 14. Landsat 8 classification results of 8 DOS bands at 30m with DT, SVM, KNN, ANN and RF

These results are validated by using the google earth image. Thus, their Kappa indices are presented in Table 8. The latter compares the efficiency of the results of the DOS correction level and the reliability of the results of DN correction level distributed.

Table 8. Kappa index of Landsat 8 OLI 8 bands DOS at 30m with DT, SVM, KNN, ANN and RF

Correction	DT	SVM	KNN	ANN	RF
DN	89	91.9	91.4	91.5	92.3
DOS	90.8	92.6	92.5	92.4	93.3

Table 8 allows us to make a graphical representation. It reflects the difference between the two levels of correction. This is to improve the way of interpreting these results. Fig 15 shows the bar chart of the

reliability acquired on the Landsat 8 OLI classification of 8 bands at 30m with a DOS and DN correction level.

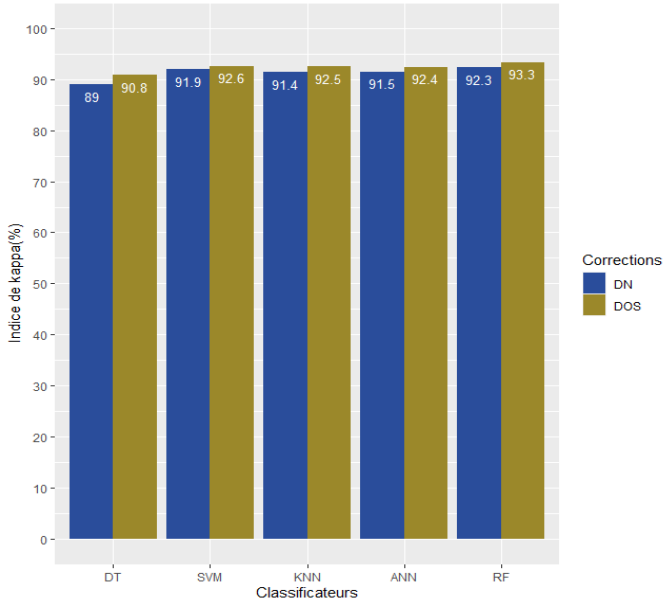

Fig. 15. Bar chart of the analysis of the contribution of a Landsat 8 OLI satellite image correction

For the DT, the DOS radiometric correction generated almost 2% more reliability than the correction level distributed by the supplier, while the SVM, KNN, ANN, RF found an extra 1% of reliability between the two correction levels.

3.1.4. Sensitivity analysis of the hugs phenomenon

First, the sensitivity to the hugs phenomenon is considered one of the. Indeed, this phenomenon leads to a significant decrease in the accuracy of the classification result. Moreover, this problem is related to the machine learning algorithm. Therefore, the amount of samples adequate to a classifier remains a big problem if it suffers from this phenomenon. Thus, the question arises: "Which classifiers are susceptible to the Hugs phenomenon?". In order to answer this question, we will perform processing using 50, 100, 150, 200, 250, and 300 samples when creating the prediction model to which the 3bands, 4bands images of Sentinel 2, Landsat 8, and SPOT 5, and also 8bands of Landsat 8 will be used.

3.1.4.1. Sensitivity to the three (3) bands Hugs phenomenon

For the three (3) band images, the analyses gave us the accuracies of each number of samples and algorithm presented in Table 9.

Table 9. 3bands optical image analysis of the sensitivity to the Hugs phenomenon

	Sentinel 2					Landsat 8					SPOT 5				
	DT	*SVM*	*KNN*	*ANN*	*RF*	*DT*	*SVM*	*KNN*	*ANN*	*RF*	*DT*	*SVM*	*KNN*	*ANN*	*RF*
50	85.6	86.2	86	87.2	87	85	86.2	88.6	86	87	84	90.1	89.9	89.2	90.8
100	85.9	87.3	89	89.4	88.4	88	87.5	89.3	87.7	89	85	90.7	90.5	89.6	91.5
150	86.3	88.5	90	90.6	89.2	89.2	88	90	88.6	89.5	87	91.5	90.9	90.1	91.8
200	89.5	89.8	90.5	89.5	90.5	89	89.4	90.5	89.4	90.7	88	92	91.7	90.5	92.1
250	88	90.2	91	89.2	91.1	88.3	90	91	90	91.1	89	93.1	92.5	91.2	93.7
300	87	91	92	89	92.9	87	91.1	91.8	90.3	90.3	91.2	92.3	92.3	90.6	92.3

Then, we drew the histograms corresponding to each image of 3 bands used to examine this phenomenon of Hugs well. The obtained diagrams are thus presented in Fig 16.

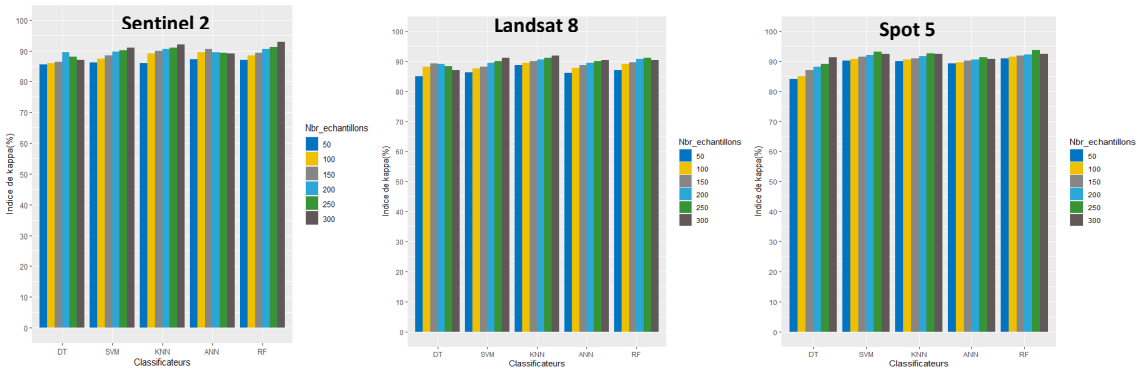

Fig. 16. Sensitivity analysis diagrams for the Hugs phenomenon on Landsat 8, Spot 5, Sentinel 2 to 3 bands images

In Sentinel 2 3bands image processing, DT and ANN present Hugs phenomenon, while SVM, KNN, and RF are not susceptible to this phenomenon. Then, the DT and the RF present Hugs phenomenon, but the DT is the most affected in Landsat 8 image, while the DT is not affected by this phenomenon in SPOT 5, while the SVM, KNN, ANN and RF are less affected.

3.1.4.2. Sensitivity of four (4) bands Hugs phenomenon

The appearance of this phenomenon could be different in the 4 bands than in the three previous bands. The analyses on each 4 bands image and the number of samples show the Kappa indices obtained (Table 10).

Table 10. 4 bands optical image analysis of the sensitivity to the Hugs phenomenon

	Sentinel 2					Landsat 8					SPOT 5				
	DT	*SVM*	*KNN*	*ANN*	*RF*	*DT*	*SVM*	*KNN*	*ANN*	*RF*	*DT*	*SVM*	*KNN*	*ANN*	*RF*
50	85	91.3	91.3	89	91	85.1	91	90	90.3	90	85	91.3	91.5	90.7	91
100	85.7	92.3	92	91	91.8	87.5	91.5	91.1	90.5	90.8	86	92.3	91.8	91	91.4
150	87	92.6	92.3	92.9	92	90.5	92.5	91.8	90.8	91	87.5	92.6	92	91.5	91.9
200	90.9	93	92.4	91.7	92.8	90	92.3	92	91.2	91.5	88.8	93.9	92.3	93	93
250	89	93.8	92.6	90.5	93.2	89.8	92.2	92.4	91.5	92.1	90.6	93.6	92.5	92.2	93.8
300	88		93.1	88.3	93.8	89.2	92	92.9	92.3	92.6	88.1	93.4	93	91.3	94.2

The graphical representation of these values (Table 10) will allow us to analyze this phenomenon and its relationship with the number of samples, the image used and the algorithm used. Fig 17 represents this phenomenon in 4 bands.

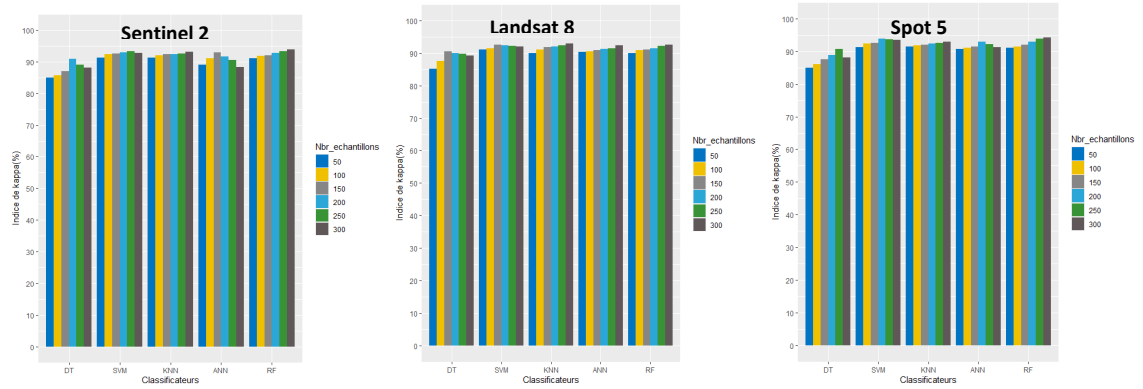

Fig. 17. Sensitivity analysis diagrams for the Hugs phenomenon on Landsat 8, Spot 5, Sentinel 2 4 bands images

Under the Sentinel 2 image, the DT, SVM, and ANN present the Hugs phenomenon, while the KNN and RF are Stable. Then, this phenomenon is proven in Landsat 8 with the DT and SVM algorithms, while it is negative with the KNN, ANN, and RF algorithms. In the SPOT image, the research is common to that of Sentinel, like the phenomenon is present on the DT, SVM, and ANN while it is absent on the KNN and RF.

3.1.4.3. Sensitivity of 8 bands Hugs phenomenon

A large number of spectral bands could accentuate the presence of the Hugs phenomenon on certain algorithms. Thus, to verify its presence, we will classify the images (DN and DOS) of 8 bands of Landsat 8 on the different algorithms against the different number of samples. In addition, the radiometrically corrected image with a correction level distributed by the supplier and the image with a DOS correction level will be compared. Hence, their kappa indices are presented in Table 11.

Table 11. 8bands optical image analysis of the sensitivity to the Hugs phenomenon

	Landsat 8 DN					Landsat 8 DOS				
	DT	*SVM*	*KNN*	*ANN*	*RF*	*DT*	*SVM*	*KNN*	*ANN*	*RF*
50	86	90.1	89	89.67	91	86	91	90.3	90.9	91
100	89	90.4	89.6	90	92.3	88.3	91.5	90.9	91.1	91.4
150	88.7	91	90	90.3	91.9	88.9	91.8	91.2	91.3	92.5
200	88	91.5	91.4	90.7	91.4	90.8	92.4	91.6	91.7	93.3
250	87.3	91.9	91.1	91	91.2	89.7	92.4	92	92.2	93
300	87	91.3	91.5	91.5	90.5	88	92.6	92.5	92.4	93.1

In order to make relevant deductions from this comparison of the hugs phenomenon, we will graphically represent this table. Fig 18 illustrates the comparison of Landsat 8 with a DN correction level compared to DOS.

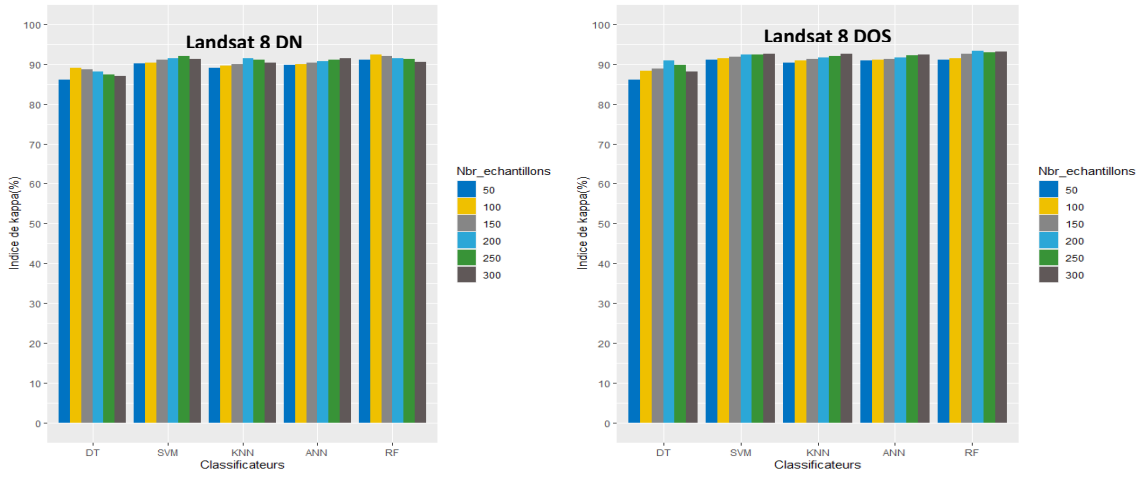

Fig. 18. Diagrams of sensitivity analysis to the Hugs phenomenon on the Landsat 8 image corrected DN and DOS

From Fig 18, the analysis of Landsat 8 DN shows that DT, SVM, KNN and RF represent Hugs phenomenon of which DT and RF are the most susceptible, while SVM and KNN are less susceptible. ANN resists the Hugs phenomenon because among of the 5 algorithms it is the only one that does not present any of this phenomenon. Then, the analysis on Landsat 8 DOS reflects that the DT still represents this phenomenon even after the radiometric correction but in a reduced way, while the SVM, KNN, ANN and RF show an increasing accuracy with respect to the number of samples.

3.1.5. analysis of processing times

The execution time is one of the criteria to select an algorithm. Indeed, this could put the difference from one algorithm to another. Therefore, we will conduct this analysis in which we can measure their

processing costs. Hence, the question arises: "Which algorithms are the most optimized? To answer this question, we will analyze through a common hardware resource the 3 spectral band satellite image processing time of each algorithm. It can be noted that we take the execution time of the classification phase for each algorithm particularly.

3.1.5.1. Processing time obtained in relation to the best accuracy

After classifying our images, named Senitnel 2, Spot 5 and Landsat 8, we also measured the processing time of each algorithm. During this analysis, we particularly took the processing time where we got the highest accuracy on each algorithm. Fig 19 illustrates the bar chart showing the processing time corresponding to each algorithm and the images used.

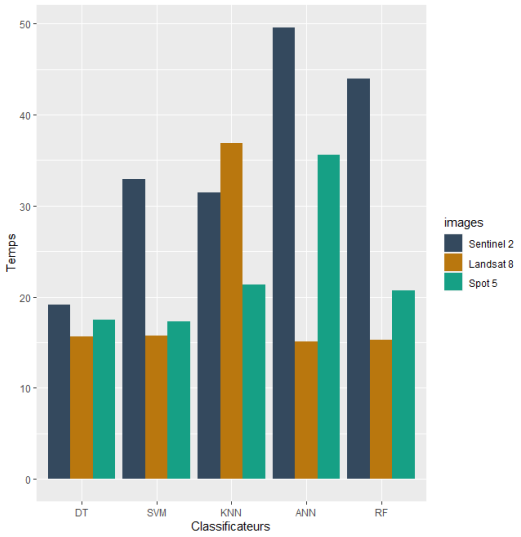

Fig. 19. Time analysis diagrams of 3bands satellite image processing

According to this diagram, in DT, the processing times are less long than the other algorithms, with Sentinel 2 processing being the longest, followed by Spot and then Landsat 8. In addition, Landsat 8 has the longest processing time during processing with KNN, then Sentinel 2 and Spot 5. On the other hand, the processing with ANN shows the longest time compared to the other algorithms, of which Sentinel 2 is the first most long, then the SPOT and Landsat 8. In general, the Sentinel 2 image processing is the longest, then the Spot and Landsat 8.

3.1.5.2. Processing time about the number of samples

Analyzing the processing time on different samples (50, 100, 150, 200, 250, 300) allows us to verify if there is a significant relationship between the number of samples and the algorithm. Thus, the results obtained on each algorithm are shown in Fig 20.

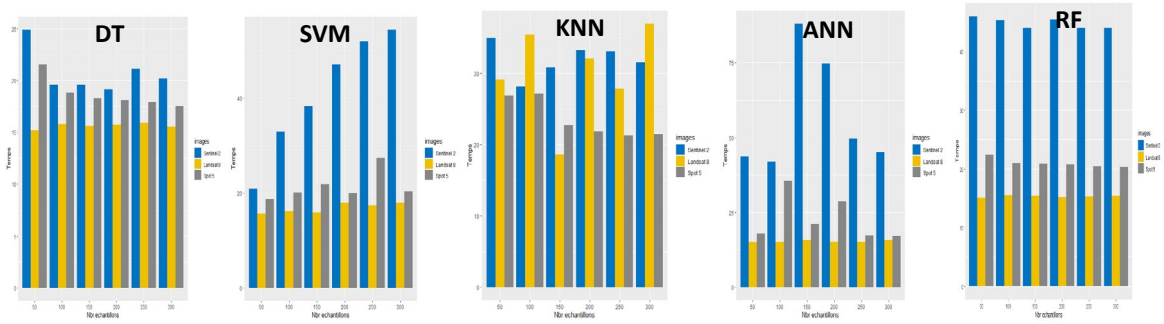

Fig. 20. Analysis diagrams of processing times to the number of samples

In DT, the processing time of Sentinel 2 and Spot 5 images is a little varied while more or less stable with Landsat 8. Subsequently, with SVM, the processing time of Sentinel 2 increases when the number

of samples also increases, while the processing time is a little varied for Spot 5 and Landsat 8. With KNN the processing time is completely varied for each image. In addition, with ANN the processing time is varied on Sentinel 2 and Spot 5, while it is stable on Landsat 8. In RF the processing times are constant for each satellite image used.

3.2. Discussion

3.2.1. Discussion for precision analysis

First, the accuracy is higher on high-resolution images than the medium and low-resolution images. They were especially pronounced in the SPOT 5 image, while the lower accuracy occurred in the Terra Modis image. Indeed, this differentiation could be caused by the amount of information or feature, the number of variables (spectral band), and the number of pixels in an image. Second, RF shows much higher accuracies on high-resolution images, while KNN is reliable for medium and low-resolution images. The efficiency of an algorithm could explain this to the number of features and related information to the image. According to [25], [21] KNN does not make any assumptions about the variables. Thus we can consider that this is the reason why KNN could give impressive accuracy on the case of 2bands image. Similarly, they also confirmed that KNN is favorable for Landsat satellite images. But, Sentinel 2 and SPOT 5 images had more information than Landsat 8 and Terra Modis as well. The work of [5], [2] also confirmed this unprecedented efficiency of RF on high dimensional data. In terms of accuracy, the RF algorithm is the leader, while the SVM is qualified in second. On the one hand, the DT algorithm showed significant inefficiency. This situation is normal for the RF to the fact is an improved version of DT in a Bagging technique. On the other hand, the accuracy is significantly improved for all algorithms during the use of 4bands compared to the 2bands and 3bands spectral. For example, SVM, which has shown significant improvement in 4bands, is comparable to RF. In short, a sufficient number of features contributes greatly to the reliability of a classifier.

3.2.2. Discussion for big data capacity

In general, each algorithm used suffers from a drop in accuracy during image analyses with large amounts of the varactor. This problem is significant, especially for DT and KNN. At least two factors can explain this. On the one hand, the bands used have much noise, which can lead to the inefficiency of an algorithm. On the other hand, this reliability failure occurs because the implemented algorithm is sensitive to noise and inefficient in terms of large quantities of processed variables. Due to this inefficiency, KNN became fourth in terms of reliability on Landsat 8 satellite image of OLI sensor with 8 spectral bands while RF takes the first position, followed by SVM and ANN. The works of [26],[27],[28] confirmed that KNN presents limitations in the framework of satellite image processing like the number of reference plots and the quantity of measurement. These factors could be considered to cause the problem of KNN. Undeniably, the capacity of each algorithm between 4 and 8 spectral bands remains unknown. But, we confirm that all the algorithms used have notable limit in terms of the quantity of variable or spectral bands, but the rate depends on the classifier. Moreover, using many more variables does not mean a real solution to the classification using a Machine learning classifier. Therefore, it would be necessary to limit the number of bands or improve a classifier's performance. These are considerably adequate solutions in order to make a classifier perform well.

3.2.3. Discussion of the contributions of preprocessing to the classifier

Usually, a significant improvement in accuracy found on each Algorithm appreciates to the use of the DOS radiometric correction. Indeed, this could be due to two reasons. On the one hand, preprocessing, such as radiometric correction, makes the machine learning algorithm efficient. On the other hand, the preprocessing could eliminate some noises related to the spectral bands, thus making the classification easy. Then, this improvement is significant, especially on the DT, it means that the less performing classifiers get much more out of using this method compared to the algorithms known as quite performing. In addition, RF is still ranked the best classifier compared to the others. Due to the positive results during this analysis, preprocessing is thus one of the way to improve the efficiency of the classifier in front of the important number of variables or the spectral band. Hence, the large number of spectral bands could turn into more reliable but this depends not only on the efficiency of

the classifier as well as the method implemented. The work of [26] confirms the contribution of radiometric correction in this study of land use, in which they said that the classifiers could benefit more performance through radiometric correction. Of course, this analysis was performed specifically on the Landsat 8 OLI 8bands image, but this research is reasonably valid on any optical satellite image.

3.2.4. Discussion for the sensitivity to the Hugs phenomenon

The DT, SVM, ANN and RF algorithms are commonly susceptible to the Hugs phenomenon when processing 3bandS optical images, and the KNN is stable. This problem occurs due to two factors. On the one hand, this can be caused by the algorithm's efficiency in differentiating two classes, and then as more samples are added, irrelevant features may be present. Therefore, it would be difficult for the classifier to separate two so near classes. On the other hand, this could also be a problem of the variables used that are not relevant, which leads to the inefficiency of the algorithms synthesized between the variables. However, the stability of KNN could be explained by its performance in terms few numbers of variables as well as its characteristic that does not pose synthesis in terms of variables. As a result, the 4 bands did not significantly change the situation for DT, SVM, and ANN, while RF and KNN seem stable and improved accuracies. On the other hand, the insensitivity of ANN to the Hugs phenomenon during the analyses on sample numbers using the Landsat 8 image of correction level DN would mean stability on large and uncorrected images, while DT, SVM, KNN, and RF are all victims of this phenomenon. A radiometric correction makes all algorithms stable except DT. Thus, we can draw that it is necessary to correct the image, which participates in the elimination of noise and which intervenes in the stability of a classifier. In addition, the selection of bands would be a solution to eliminate noise in the images to be used and to avoid the Hugs phenomenon. This would easily help the user during the learning process without knowing how effective it can be.

3.2.5 Discussion for processing times

Sentinel 2 image processing times are regularly much higher than other images, while Landsat image processing times are generally lower. This could be caused by the image size since we processed an area of the same size, but the numbers of pixels are different. Our Sentinel 2 image has 1015 pixels, equivalent to 779 pixels on SPOT 5 and 108 on Landsat 8. Therefore, the more the number of pixels, the more the execution time is costly. In addition, the ANN showed non-optimal processing time and then RF, while the DT is optimal. These could be explained by the complexity of the structure of the Algorithm for ANN; then it is so challenging to implement. Besides, creating its prediction model takes much time as the size or the "size" is specified. According to [29], ANN can be slow to train; it is possible to produce non-optimal classification, and very easy to over-train. Certainly, the non-optimal processing time could usually mean difficulties to the classifiers at the time of processing. However, sometimes, this could practically translate into the reliability of the prediction model, for example, in the case of RF, which shows higher accuracy but with non-reduced processing times. , The processing time does not depend on the number of samples but mainly on the size of the image to be processed, the prediction model, and the classifier's efficiency. In most cases, lower accuracies were obtained with less long processing time. Thus, reliable classifiers are almost non-optimal. Therefore, some optimizations would be needed to find reliability in terms of processing time.

4. Conclusion

To sum up, we performed the reliability analyses of machine learning algorithms such as DT, SVM, KNN, ANN, and RF. Then, the satellite images Sentinel 2, Landsat 8 from the OLI sensor, Terra Modis and Spot 5 HRG 2 were used, which are included in the optical satellite image category. In addition,a study area was chosen in which 5 object categories were identified: water, sand, akata, forest and grassland. Furthermore,5 hypotheses were mainly employed to verify the effectiveness of these machine learning algorithms. These include accuracy analysis, analysis of the capacity of the algorithms, analysis of the impact of preprocessing on each classifier, analysis of Hugs' phenomena, and analysis of processing times, to propose algorithms suitable for satellite image categories. The results allowed us to identify that the KNN is adequate for Landsat and Terra Modis satellite images with less than four bands or

medium and low-resolution satellite images. Moreover, the RF classifier is a good candidate for other categories of satellite images. In the same way, preprocessing is also essential to eliminate noise in the spectral bands. Also, all the machine learning algorithms present the Hugs phenomenon as well. Moreover, we have observed that the highest accuracy is almost obtained with towering processing times. On the one hand, these findings confirm that the land cover result's reliability depends not only on the algorithms but also on the images used. Moreover, much preprocessing is required before the classification of optical satellite images. On the other hand, this knowledge helps us to categorize the optical satellite images from the point of view of producing more accuracy and optimization with the help of machine learning. Some perspectives are related to this research paper. First, since some classifiers could not give reliable results, it is essential to bring improvements, such as using multi-classifiers techniques. Second, since we have found throughout these intense analyses that machine learning algorithms are all sensitive to noise related to spectral bands, it is necessary to use variable selection techniques to give even more precision to the results. Moreover, reliable results were obtained with the time of towering treatments. It is thus essential to use optimization algorithms like the techniques of metaheuristic research in order to optimize the parameters or to seek optimal spectral bands. These are in the context of improving the performance of a classifier to find even more reliability and optimization in the study of land use. In addition, our findings will use to generalize optical satellite image processing chains to ensure the reliability of the treatment results.

Declarations

Author contribution. All authors contributed equally to the main contributor of this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] J. Knorn, A. Rabe, V. C. Radeloff, T. Kuemmerle, J. Kozak, and P. Hostert, "Land cover mapping of large areas using chain classification of neighboring Landsat satellite images," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 957–964, May 2009, doi: 10.1016/J.RSE.2009.01.010.

[2] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, May 2018, doi: 10.1080/01431161.2018.1433343.

[3] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors 2018, Vol. 18, Page 18*, vol. 18, no. 1, p. 18, Dec. 2017, doi: 10.3390/S18010018.

[4] E. Tomppo and M. Katila, "Satellite image-based national forest inventory of Finland for publication in the IGARSS'91 digest," *Dig. - Int. Geosci. Remote Sens. Symp.*, vol. 3, pp. 1141–1144, 1991, doi: 10.1109/IGARSS.1991.579272.

[5] S. E. Sesnie, B. Finegan, P. E. Gessler, A. M. S. Smith, R. B. Zayra, and S. Thessler, "The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees," *http://dx.doi.org/10.1080/01431160903140803*, vol. 31, no. 11, pp. 2885–2909, 2010, doi: 10.1080/01431160903140803.

[6] A. C. Lorena *et al.*, "Comparing machine learning classifiers in potential distribution modelling," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5268–5275, May 2011, doi: 10.1016/J.ESWA.2010.10.031.

[7] M. Li, J. Im, and C. Beier, "Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest," *GIScience Remote Sens.*, vol. 50, no. 4, pp. 361–384, Aug. 2013, doi: 10.1080/15481603.2013.819161.

[8] X. Shang and L. A. Chisholm, "Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2481–2489, 2014, doi: 10.1109/JSTARS.2013.2282166.

[9] A.Ter-Sarkisov," Detection and segmentation of lesion areas in chest CT scans for the prediction of COVID-19," *Sci. Inf. Tech. Lett.*, vol. 1, no. 2, pp. 92-99, Nov. 2020, doi: 10.31763/sitech.v1i2.202.

[10] A. Septiarini, H. Hamdani, M. S. Sauri, and J. A. Widians," Image processing for maturity classification of tomato using otsu and manhattan distance methods," *J. Informatika*, vol. 16, no. 3, pp. 118-126, Sept. 2022, doi: 10.26555/jifo.v16i1.a21985.

[11] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, Aug. 2003, doi: 10.1016/S0034-4257(03)00132-9.

[12] C. E. Brodley and M. A. Friedl, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 399–409, Sep. 1997, doi: 10.1016/S0034-4257(97)00049-7.

[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," *Proc. Fifth Annu. ACM Work. Comput. Learn. Theory*, pp. 144–152, 1992, doi: 10.1145/130385.130401.

[14] C. Cortes, V. Vapnik, and L. Saitta, "Support-vector networks," *Mach. Learn. 1995 203*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[15] V. N. Vapnik, "The Nature of Statistical Learning Theory," *Nat. Stat. Learn. Theory*, 1995, doi: 10.1007/978-1-4757-2440-0.

[16] -Stéphane -CANU, "Apprentissage et noyaux : séparateur à vaste marge (SVM)," *Rev. l'Electricité l'Electronique*, vol., no. 07, p. 69, 2006, doi: 10.3845/REE.2006.062.

[17] F. Lauer *et al.*, "Méthodes SVM pour l ' identification To cite this version : HAL Id : hal-00110344 Méthodes SVM pour l ' identification," 2006 Available at : https://hal.archives-ouvertes.fr/hal-00110344.

[18] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992, doi: 10.1080/00031305.1992.10475879.

[19] R. E. McRoberts, M. D. Nelson, and D. G. Wendt, "Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique," *Remote Sens. Environ.*, vol. 82, no. 2–3, pp. 457–468, Oct. 2002, doi: 10.1016/S0034-4257(02)00064-0.

[20] H. Franco-Lopez, A. R. Ek, and M. E. Bauer, "Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method," *Remote Sens. Environ.*, vol. 77, no. 3, pp. 251–274, Sep. 2001, doi: 10.1016/S0034-4257(01)00209-7.

[21] E. Tomppo and M. Halme, "Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach," *Remote Sens. Environ.*, vol. 92, no. 1, pp. 1–20, Jul. 2004, doi: 10.1016/J.RSE.2004.04.003.

[22] P. M. Atkinson and A. R. L. Tatnall, "Introduction Neural networks in remote sensing," *http://dx.doi.org/10.1080/014311697218700*, vol. 18, no. 4, pp. 699–709, 2010, doi: 10.1080/014311697218700.

[23] A. A. Azmer, N. Hassan, S. H. Khaleefah, S. A. Mostafa, and A. A. Ramli, "Comparative analysis of classification techniques for leaves and land cover texture," *Int. J. Adv. Intell. Inform.*, vol. 7, no. 3, pp. 357–367, Nov. 2021, doi: 10.26555/ijain.v7i3.706.

[24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[25] R. Haapanen, A. R. Ek, M. E. Bauer, and A. O. Finley, "Delineation of forest/nonforest land use classes using nearest neighbor methods," *Remote Sens. Environ.*, vol. 89, no. 3, pp. 265–271, Feb. 2004, doi: 10.1016/J.RSE.2003.10.002.

[26] T. Koukal, F. Suppan, and W. Schneider, "The impact of relative radiometric calibration on the accuracy of kNN-predictions of forest attributes," *Remote Sens. Environ.*, vol. 110, no. 4, pp. 431–437, Oct. 2007, doi: 10.1016/J.RSE.2006.08.016.

[27] A. A. Mehdawi and B. Bin Ahmad, "K-nearest neighbor method for classification of forest encroachment by using reflectance processing of remote sensing spectroradiometer data," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 15, pp. 2881–2885, 2013, doi: 10.19026/RJASET.6.3799.

[28] Q. Meng, C. J. Cieszewski, M. Madden, and B. E. Borders, "K Nearest Neighbor Method for Forest Inventory Using Remote Sensing Data," *http://dx.doi.org/10.2747/1548-1603.44.2.149*, vol. 44, no. 2, pp. 149–165, Apr. 2013, doi: 10.2747/1548-1603.44.2.149.

[29] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, May 2018, doi: 10.1080/01431161.2018.1433343.