

# ***Interactive comment on* “Comparison of machine learning classification algorithms for land cover change in a coastal area affected by the 2010 Earthquake and Tsunami in Chile” by Matias I. Volke and Rodrigo Abarca-Del-Rio**

**Matias I. Volke and Rodrigo Abarca-Del-Rio**

matiasvolke@udec.cl

Received and published: 3 July 2020

Thank for this comment which allowed to improve the statistical significance of the work. For the process to achieve statistical confidence, classifiers are now iterated 100 times in 10 separate sub-sample data sets. We report the mean values and standard deviation in the manuscript (see section 3.- result) for each of the 10 training sample subgroups (see tables en figure 1 and 2). To evaluate the effect of training sample sizes as well as the performance of the classification algorithms on classification accuracies, we randomly divided the training sample data into 10 different data sets. The procedure

[Printer-friendly version](#)

[Discussion paper](#)



described in section 2.4.5. - Classification scheme" of the manuscript. These data sets correspond to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% of the total training data set. We made sure that each subset had the same proportions of training samples per soil cover class as the original set. To ensure that this occurs, a technique called "stratified holdout sampling" is used. This technique is implemented in the `CreateDataPartition()` function of the `caret` package. A fixed validation sample is reserved for the performance evaluation of all classifiers, corresponding to 30% of the initial total of the training samples. This sample is obtained in the same way as the training samples. However, it is reserved only for the validation of the algorithms trained in the different sizes of training sub-samples. Also applied the Wilcoxon signed-rank test to evaluate and assess the statistical significance of systematic pairwise differences between the ML models at the significant level  $\alpha = 5\%$ . The results of the p-values of the significance test are presented in the table in figure 3 below.

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2020-41>, 2020.

[Printer-friendly version](#)[Discussion paper](#)

Landsat																				
Algorithm	10%		20%		30%		40%		50%		60%		70%		80%		90%		100%	
	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde
<b>NB</b>	0.895	0.013	0.908	0.011	0.909	0.010	0.912	0.008	0.913	0.006	0.914	0.005	0.915	0.004	0.914	0.003	0.915	0.002	0.914	0.002
<b>KNN</b>	0.911	0.009	0.922	0.008	0.927	0.007	0.931	0.006	0.934	0.005	0.937	0.005	0.938	0.004	0.938	0.003	0.939	0.001	0.939	0.001
<b>MARS</b>	0.884	0.014	0.904	0.010	0.911	0.007	0.918	0.007	0.921	0.005	0.924	0.005	0.925	0.004	0.925	0.003	0.926	0.001	0.927	0.001
<b>GBM</b>	0.899	0.012	0.917	0.010	0.925	0.008	0.933	0.007	0.936	0.005	0.940	0.004	0.942	0.003	0.943	0.004	0.945	0.003	0.946	0.002
<b>SVM</b>	0.916	0.009	0.932	0.008	0.939	0.006	0.943	0.005	0.947	0.004	0.950	0.004	0.952	0.003	0.953	0.002	0.955	0.001	0.955	0.001
<b>RF</b>	0.917	0.010	0.927	0.007	0.933	0.006	0.936	0.006	0.939	0.005	0.940	0.005	0.941	0.004	0.943	0.004	0.943	0.003	0.942	0.002
<b>DNN</b>	0.925	0.011	0.939	0.009	0.934	0.009	0.934	0.008	0.938	0.008	0.945	0.007	0.946	0.005	0.948	0.004	0.949	0.001	0.949	0.001
<b>XGB</b>	0.904	0.016	0.920	0.012	0.927	0.010	0.933	0.007	0.939	0.006	0.942	0.005	0.946	0.004	0.947	0.003	0.951	0.001	0.952	0.001

**Fig. 1.** Table 4. Average overall accuracies and their coefficient of variation for ML algorithms applied in all training sample size (Landsat images).

[Printer-friendly version](#)

[Discussion paper](#)



Aster																				
Algorithm	10%		20%		30%		40%		50%		60%		70%		80%		90%		100%	
	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde	mean	stde
<b>NB</b>	0.924	0.013	0.923	0.007	0.931	0.007	0.934	0.011	0.932	0.007	0.933	0.005	0.933	0.005	0.932	0.002	0.934	0.001	0.934	0.001
<b>KNN</b>	0.925	0.014	0.940	0.010	0.954	0.006	0.957	0.005	0.964	0.004	0.966	0.002	0.967	0.003	0.970	0.002	0.971	0.002	0.971	0.002
<b>MARS</b>	0.942	0.010	0.952	0.021	0.949	0.011	0.959	0.022	0.959	0.021	0.967	0.004	0.967	0.007	0.969	0.014	0.969	0.002	0.970	0.002
<b>GBM</b>	0.935	0.010	0.958	0.008	0.965	0.004	0.968	0.004	0.970	0.002	0.973	0.003	0.974	0.002	0.975	0.002	0.975	0.001	0.975	0.001
<b>SVM</b>	0.944	0.012	0.962	0.008	0.967	0.005	0.971	0.006	0.975	0.003	0.976	0.004	0.976	0.003	0.977	0.001	0.979	0.001	0.979	0.001
<b>RF</b>	0.930	0.015	0.957	0.014	0.963	0.003	0.965	0.004	0.967	0.003	0.970	0.002	0.971	0.002	0.972	0.002	0.974	0.001	0.973	0.001
<b>DNN</b>	0.945	0.002	0.951	0.001	0.961	0.002	0.964	0.001	0.967	0.002	0.968	0.004	0.969	0.003	0.972	0.001	0.973	0.001	0.973	0.001
<b>XGB</b>	0.935	0.013	0.957	0.009	0.965	0.004	0.966	0.005	0.969	0.002	0.971	0.004	0.974	0.002	0.973	0.002	0.974	0.001	0.976	0.001

**Fig. 2.** Table 5. Average overall accuracies and their coefficient of variation for ML algorithms applied in all training sample size (Aster images).

[Printer-friendly version](#)

[Discussion paper](#)



Landsat				Aster				Landsat-Aster	
No	p.value	No	p.value	No	p.value	No	p.value	No	p.value
NB vs. KNN	<0.001	MARS vs. SVM	<0.001	NB vs. KNN	<0.001	MARS vs. SVM	<0.001	NB	<0.001
NB vs. MARS	<0.001	MARS vs. RF	<0.001	NB vs. MARS	<0.001	MARS vs. RF	0.064	KNN	<0.001
NB vs. GBM	<0.001	MARS vs. DNN	0.084	NB vs. GBM	<0.001	MARS vs. DNN	0.004	MARS	<0.001
NB vs. SVM	<0.001	MARS vs. XGB	0.033	NB vs. SVM	<0.001	MARS vs. XGB	0.037	GBM	0.006
NB vs. RF	<0.001	GBM vs. SVM	<0.001	NB vs. RF	<0.001	GBM vs. SVM	<0.001	SVM	0.006
NB vs. DNN	<0.001	GBM vs. RF	<0.001	NB vs. DNN	<0.001	GBM vs. RF	<0.001	RF	<0.001
NB vs. XGB	<0.001	GBM vs. DNN	0.084	NB vs. XGB	<0.001	GBM vs. DNN	0.084	DNN	<0.001
KNN vs. MARS	0.906	GBM vs. XGB	0.033	KNN vs. MARS	0.906	GBM vs. XGB	0.033	XGB	<0.001
KNN vs. GBM	<0.001	SVM vs RF	0.006	KNN vs. GBM	<0.001	SVM vs RF	0.006		
KNN vs. SVM	<0.001	SVM vs DNN	0.004	KNN vs. SVM	<0.001	SVM vs DNN	0.004		
KNN vs. RF	<0.001	SVM vs XGB	0.006	KNN vs. RF	<0.001	SVM vs XGB	0.006		
KNN vs. DNN	0.006	RF vs XGB	0.232	KNN vs. DNN	0.006	RF vs XGB	0.023		
KNN vs. XGB	<0.001	RF vs DNN	0.020	KNN vs. XGB	<0.001	RF vs DNN	0.088		
MARS vs. GBM	0.041	DNN vs XGB	0.084	MARS vs. GBM	0.041	DNN vs XGB	0.084		

**Fig. 3.** Table 6. Wilcoxon test results between different models for each image type and between images (significance achieved at  $p < 0.05$ ).

[Printer-friendly version](#)

[Discussion paper](#)

