Improving Land Cover Classification Using Genetic Programming

João E. Batista[1], Ana I. R. Cabral[2], Maria J. P. Vasconcelos[2], Leonardo Vanneschi[3], Sara Silva[1]

[1] LASIGE, Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016 Lisbon, Portugal
[2] Forest Research Centre, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017, Lisbon, Portugal
[3] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

Full Paper

LASIGE data and systems intelligence

Motivation:

Some mechanisms, such as the REDD+, developed by the UNFCCC, have the objective of avoiding deforestation and forest degradation while promoting its conservation, sustainable management and enhancement of the forest carbon stocks. However, for this kind of mechanisms to be effective, they require the ability to Measure, Report and Verify (MRV) carbon emissions and flows.

Goals:

To use the M3GP algorithm to create hyper-features that improve the performance of several state-of-the-art machine learning algorithms in the classification of land cover types in a satellite image.

Hyper-Features:

The hyper-features used in this work are mathematical expressions that combine satellite bands into more informative values. The indices used in Remote Sensing are particular cases of hyper-features.

Index	Hyper-feature
NDVI	(NIR – Red) / (NIR + Red)
NDWI	(Green – NIR) / (Green + NIR)
NBR	(NIR – SWIR2) / (NIR + SWIR2)

Created	Hyper-feature
Gw2 Run#11	NIR * (Green + NIR) / (SWIR2 + Red + 1)
Ao8 Run#18	Coastal aerosol / (Blue * SWIR2²)
Ao8 Run#18	NIR – SWIR1 + Cirrus – 2 * TIRS1
Ao8 Run#18	(Coastal aerosol * Blue + Coastal Aerosol * Green * SWIR1 + Coastal Aerosol * Green * Cirrus + Green * Red * NIR) / TIRS2
...	...

Algorithms:

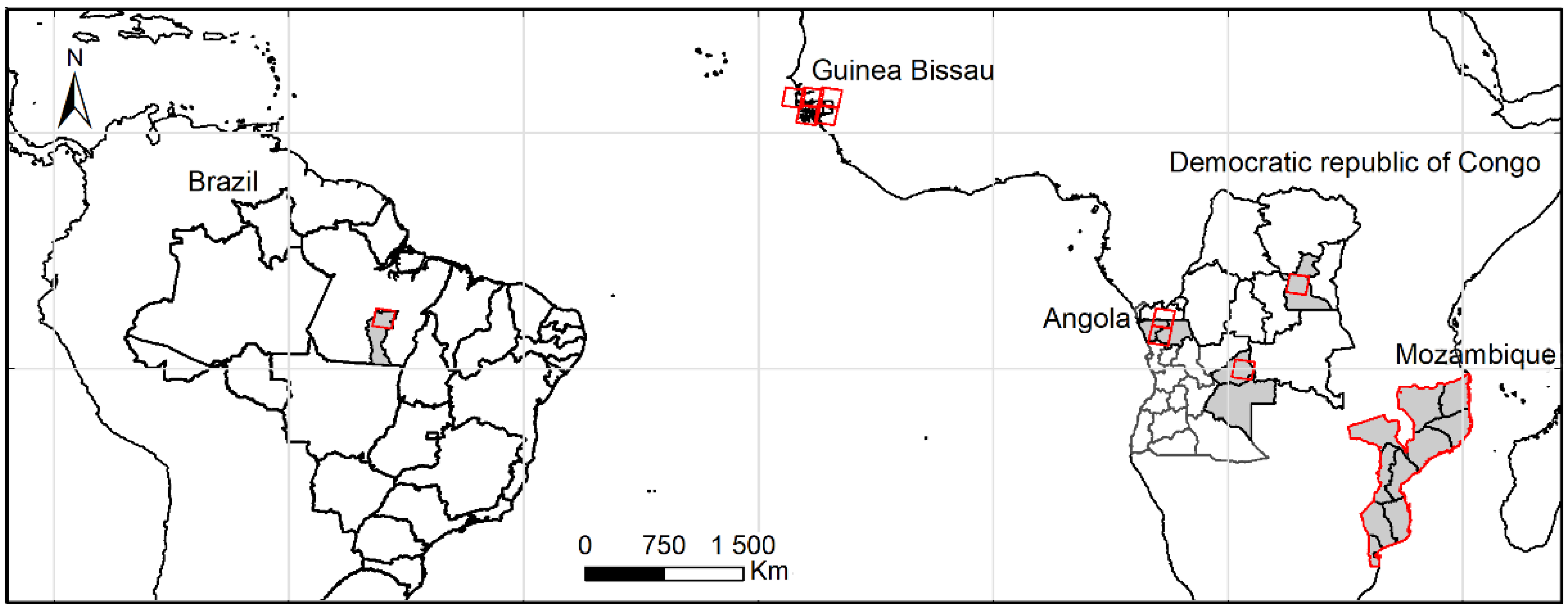
- M3GP: Was used to create hyper-features;
- Decision Trees (DT), Random Forests (RF), and XGBoost: Were used as classification models in each of the three test cases.

Test Cases:

- Original Features: The datasets contain only the original features;
- Original Features and Indices: The datasets contain the original features and the NDVI, NDWI and NBR indices;
- Original Features and Hyper-Features: The datasets contain the original features and created hyper-features.

Datasets and Study Areas:

Name	Country ISO Code	Scene Identifier Path / Row	Acquisition Date DD/MM/YYYY	Satellite	Classes	No. Bands No. Features	No. Pixels
Ao2	AO	177/67	09/07/2013	Landsat-8	2	7	3882
Br2	BR	225/64	28/02/2015	Landsat-8	2	7	4872
Cd2	CD	175/62	08/06/2013	Landsat-8	2	7	2849
Gw2	GW	204/52	13/05/2002	Landsat-7	2	7	4531
IM-3 IM-10	GW	203 / 51 52 204 / 51 52 205/51	From: 02/01/2010 To: 01/04/2010	Landsat-7	3 10	6	322 6798
Ao8	AO	182/64 65	18/06/2016	Landsat-8	8	10	2183
Gw10	GW	204/51 52 205/51	01/03/2019 24/03/2019	Landsat-8	10	7	5080
Mz6	MZ	Entire Country	From: 06/04/2016 To: 06/10/2016	Sentinel-2 A	6	10	190202



Results:

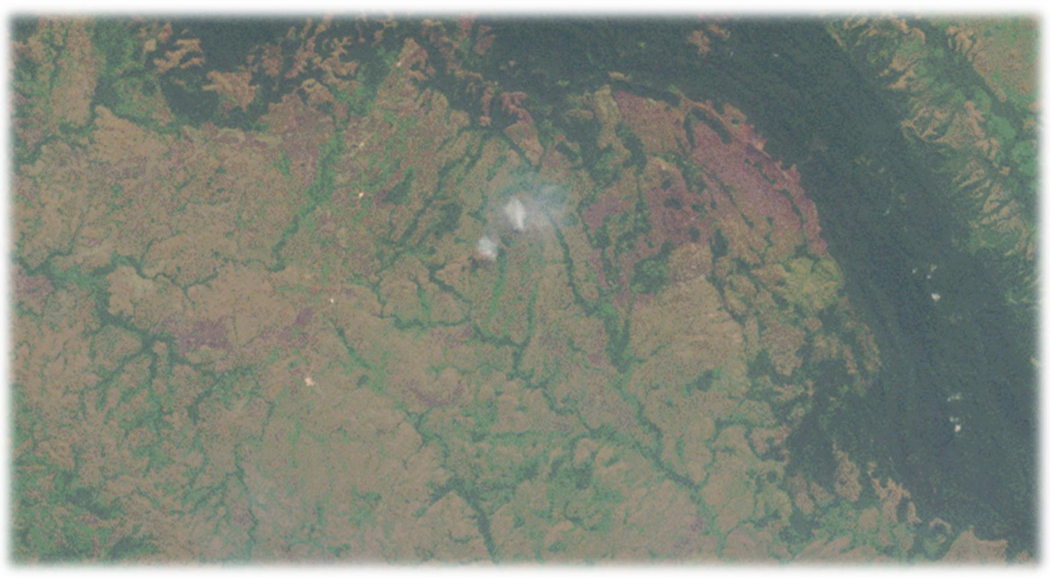
- Binary Classification: Neither the indices nor the hyper-features had a significant* effect in the accuracy of the classifiers;
- Multiclass Classification: Both the indices and the hyper-features significantly improved the accuracy of the models in the "multidate" datasets;
- Hyper-features vs Indices: The hyper-features significantly improved the accuracy in more cases than the indices and never degraded the accuracy of the models.

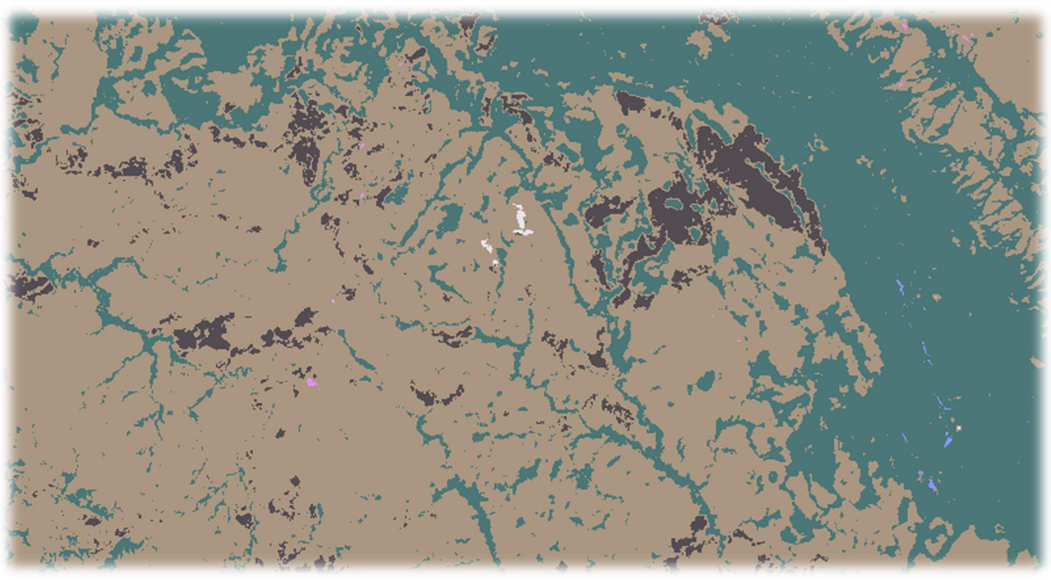
Did the indices improve the accuracy?	Ao2	Br2	Cd2	Gw2	IM-3	IM-10	Ao8	Gw10	Mz6
DT	---	---	---	---	---	---	---	YES	---
RF	---	---	---	---	---	YES	---	---	NO
XGBoost	---	---	---	---	---	YES	---	---	YES

Did the hyper-features improve the accuracy?	Ao2	Br2	Cd2	Gw2	IM-3	IM-10	Ao8	Gw10	Mz6
DT	---	---	---	---	---	YES	---	YES	YES
RF	---	---	---	---	---	YES	---	---	YES
XGBoost	---	---	---	---	YES	YES	---	YES	YES

* All comparisons were made using a *p*-value of 0.01, obtained using Kruskal-Wallis H-test on the results obtained in 30 runs.

Multiclass Classification in Angola (Scene 182/64 from 18/06/2016):





- Burnt
- Clouds
- Forest
- Savanna
- Urban
- Water

The image on the left was obtained using the satellite bands 2 (Blue), 3 (Green) and 4 (Red).

Conclusions and Future Work:

The hyper-features used in this work improved the performance of several classification algorithms in multiclass classification datasets obtained by compiling satellite images with different acquisition dates. This implies that the hyper-features can be useful when dealing with radiometric noise between images. This work will be extended to regression problems, such as the prediction of biomass in satellite images.

Acknowledgements: This work was partially supported by FCT through funding of LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020) and CEF (UIDB/00239/2020); projects BINDER (PTDC/CCI-INF/29168/2017), OPTOX (PTDC/CTA-AMB/30056/2017), PREDICT (PTDC/CCI-CIF/ 29877/2017), INTERPHENO (PTDC/ASP-PLA/28726/2017), GADgET (DSAIPA/DS/0022/2018), AICE (DSAIPA/DS/0113/2019); PhD Grant (SFRH/BD/143972/2019).