Sentinel-2 Image Scene Classification over Alentejo Region Farmland

Kashyap Raiyani[1], **Teresa Gonçalves**[1], **Luís Rato**[1,2], **Pedro Salgueiro**[1], **José R. Marques da Silva**[3,4]

[1]Departmento de Informática, Universidade de Évora, Portugal. [2]CIMA, Universidade de Évora, Portugal

[3]MED, Universidade de Évora, Portugal. [4]Agroinsider Lda., Évora, Portugal

(kshyp,tcg,lmr,pds,jmsilva)@uevora.pt

Problem Statement

- Satellite images - forecasting harvest dates, yield estimation, and manufacturing control.
- Given single satellite image over 5 days, is it possible to use all of them?
- It is essential to identify and remove atmospheric distorted images.
- Identifying - `Cloud`, `Shadow`, `Cirrus`, `Snow`, and `Water` coverage over farmland.

Building Image Scene Classifier

Dataset Creation

⇒ Holstein [1] created a database of manually labeled (6.6 million points) Sentinel-2 spectra.

⇒ 60 images as six classes `Water`, `Shadow`, `Cirrus`, `Cloud`, `Snow`, and `Clear-sky`.

⇒ 4 attributes: *product_id*, *latitude*, *longitude* and *class*.

⇒ We extended this database - adding Sentinel-2 13 bands values.

⇒ For comparison, added Sen2Cor scene classification:

Header	Column Value
Product ID	1 Column (78 character string)
Coordinates	4 Columns (latitude, longitude, east and, north)
Bands	13 Columns (Band 1 to 12 and 8A)
Tagged Class	1 Column (Manual tagged class value)
Sen2cor - SCL	1 Column (Scene classification class value)

Table 1: Structure of Extended Dataset.

Training and Evaluation

⇒ Implemented - Decision Tree (DT), Random Forest (RF) and Extreme Trees (ET) algorithms.

⇒ Compared to State-of-the-art Sen2Cor algorithm of the European Space Agency.

⇒ Recursively - used 1 image for testing and the rest 59 for training.

⇒ Presented the $F1_{avg}$ using Equation 1.

$$F1_{avg} = \sum_{p=1}^{60} (F1_p \times N_p) \div T \;\; with \;\; T = \sum_{p=1}^{60} N_p \tag{1}$$

⇒ $F1_p$ is the $F1$ value of the particular class within the product p. N_p is the number of points of the class within the product p, T is the total number of points of the class for all products and $p \in (1, 60)$ is the number of products.

Results

⇒ $F1_{avg}$ of 76.77% over all classes (using Extreme Trees), an improvement of 10% compared to Sen2Cor $F1_{avg}$ of 66.40%.

Class	DT	RF	ET	Sen2Cor	Support
Clear-sky	63.29	72.3	**74.16**	64.96	1694454 (25.56%)
Water	63.81	73.4	76.69	**80.73**	1071426 (16.16%)
Shadow	53.98	**63.96**	61.45	50.57	991393 (14.96%)
Cirrus	47.58	**56.63**	42.97	24.08	956623 (14.43%)
Cloud	65.25	75.08	**75.33**	75.04	1031819 (15.57%)
Snow	74.67	84.90	**87.00**	61.40	882763 (13.32%)
$F1_{avg}$	67.95	76.43	**76.77**	66.40	6628478 (100%)

Table 2: Evaluation Results: $F1_{avg}$ values of ML algorithms and Sen2Cor.

Application of Build Classifier

Study Area - Alentejo Region

⇒ Acquired 170 (5 days apart) ten corn parcels Sentinel-2 imagery between 05-01-2017 and 03-08-2019.

⇒ Figure 1 shows the image of the ten corn parcels (referred as parcel-1 to parcel-10 onwards).

Figure 1: Ten Corn Parcels from Alentejo Region (between (37°56'29.13" N , 8°22'21.95" W) and (37°55'32.44" N, 8°21'02.23" W) coordinates).

⇒ Crop growth can be measured by (Equation 2) the Normalized Difference Vegetation Index (NDVI) [2].

$$NDVI = (NIR - RED)/(NIR + RED) \tag{2}$$

⇒ Figure 2 shows the mean NDVI Value between 05-01-2017 and 03-08-2019 for parcel-1.

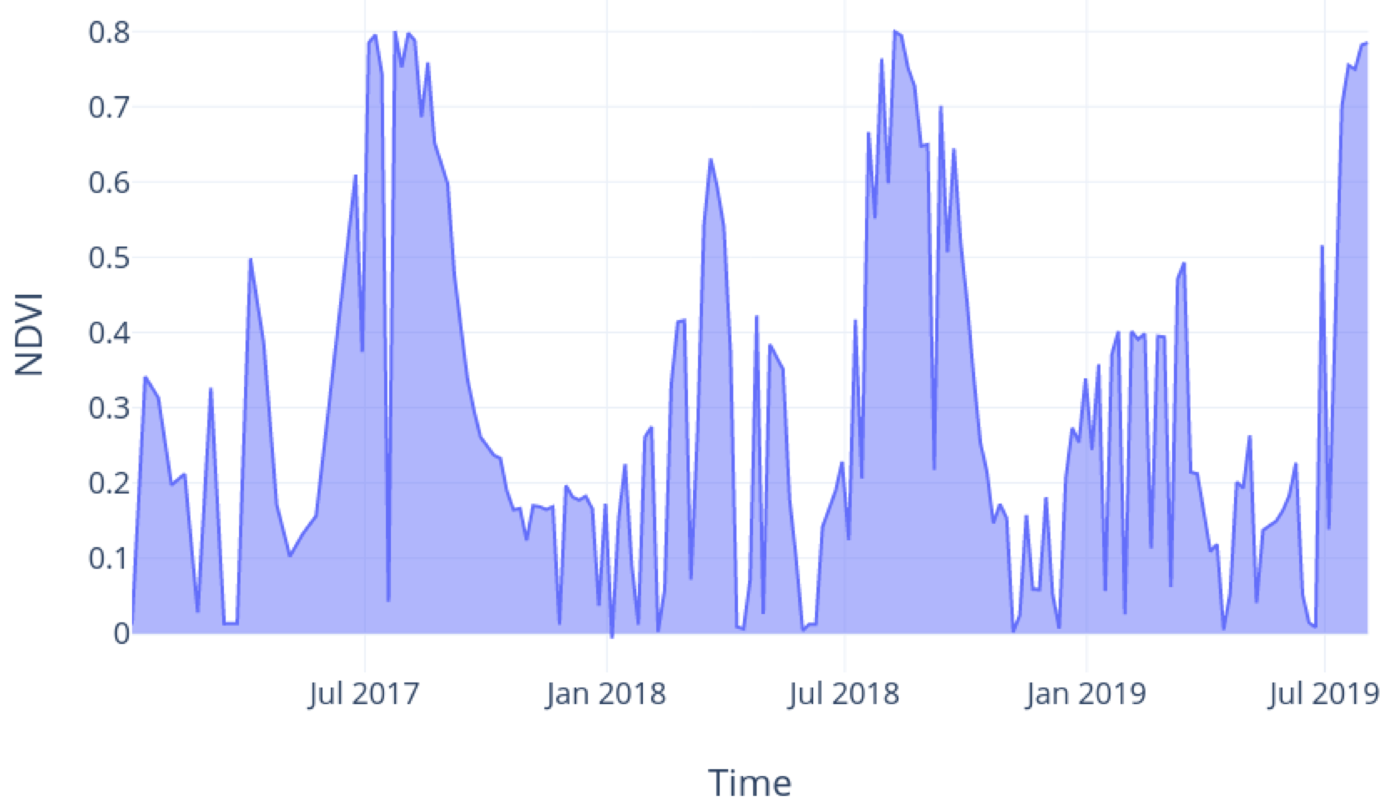

Figure 2: Parcel-1: Mean NDVI Value between 05-01-2017 and 03-08-2019.

⇒ Atmospheric disturbance can be observed as sudden dips in the NDVI values.

⇒ Losing crop growth and regain it within a range of 5 days is not possible.

Using Extreme Trees (ET) Classifier

⇒ Figure 1 was classified as no atmospheric disturbance image (`Clear-sky`) or image with disturbance (`Cloud`, `Shadow`, `Cirrus`, `Snow`, and `Water` coverage).

⇒ 0 if all points were classified as (`Clear-sky`) and 1 when all points were classified as atmospheric disturbance.

⇒ Figure 3 presents the calculated `Atmospheric Disturbance` between dates 14-06-2017 and 01-12-2017 (6 months).

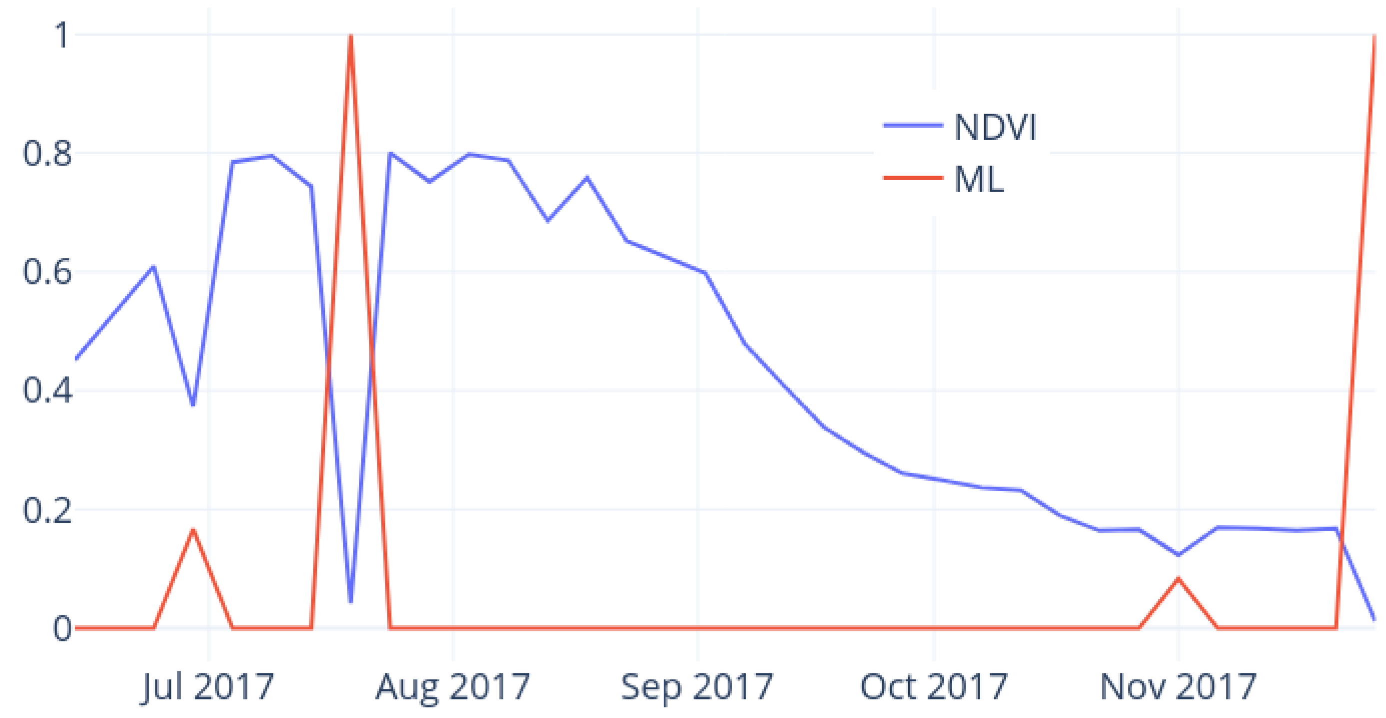

Figure 3: Parcel-1: Mean NDVI and Calculated Atmospheric Disturbance Value by Extreme Trees Machine Learning Model (between dates 14-06-2017 and 01-12-2017 (6 months)).

⇒ Red line for the ET model and blue line for mean NDVI.

⇒ The authors would like to state that 'NDVI value is not the sole parameter to find disturbance'.

⇒ This claim is supported by Figure 3 as on 08, 13, and 18 Aug'17, the mean NDVI ranges from 0.78 to 0.68 (a drop) to 0.76 but the value of atmospheric disturbance remains 0.0.

Conclusions and Future Work

- RF and ET are comparatively providing equivalent $F1_{avg}$ results (76.43% and 76.77%).
- Outperforming state-of-the-art Sen2Cor by 10% for image scene classification.
- Identifying - `Cloud`, `Shadow`, `Cirrus`, `Snow`, and `Water` coverage over Alentejo Region.
- Classified parcel images will help to remove atmospheric distorted images.
- Manually label individual data points for atmospheric disturbance.
- Compare the performance of the ML method to Sen2Cor.

Funding

This work was supported by the NIIAA (Núcleo de Investigação em Inteligência Artificial em Agricultura) project, Alentejo 2020 program (reference ALT20-03-0247-FEDER-036981).

References

[1] André Hollstein, Karl Segl, Luis Guanter, Maximilian Brell, and Marta Enesco. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images. *Remote Sensing*, 8(8):666, 2016.

[2] JW Rouse, RH Haas, JA Schell, and DW Deering. Monitoring vegetation systems in the great plains with erts. *NASA special publication*, 351:309, 1974.