Decision Tree Ensemble vs. N.N. Deep Learning: efficiency comparison for a small image dataset

Jérôme Treboux
and Dominique Genoud
Institute of Information Systems
University of Applied Sciences, HES-SO Valais
Sierre, Switzerland
Email: (jerome.treboux, dominique.genoud)@hevs.ch

Rolf Ingold
Department of Informatics
University of Fribourg
Fribourg, Switzerland
Email: rolf.ingold@unifr.ch

Abstract—**This paper presents a study of the efficiency of machine learning algorithms applied on an image recognition task. The dataset is composed of aerial GeoTIFF images of 5 different vineyards taken with a drone. It presents the application of two different classification algorithms with an efficiency comparison over a small dataset. A Neural Network algorithm for classification through the TensorFlow platform will be explained first, and a Decision Tree Ensemble algorithm for classification through a machine learning platform will be explained second. This work shows that the accuracy of the Decision Tree Ensemble algorithm** (94.27 %) **outperforms the accuracy of the Deep Learning algorithm** (91.22 %). **This result is based on the final detection accuracy as well as on the computation time.**

I. Introduction

Nowadays, Deep Learning (DL) and Machine Learning algorithms are widely used to classify objects in images. These classification and image recognition processes allow the computer to understand the content and the context of an image. The process is composed of multiple processing layers to learn the representations of data with several levels of abstraction [1]. Deep networks can be trained in such way that they can discover intricate structures within large data sets by using, for example, the backpropagation algorithm [1].

The Deep Learning algorithm is trained over a large dataset to be performant enough to recognize patterns within the dataset, such as detecting objects within images or spoken words in speech recognition. The size of data available online is growing every day and are mainly accessible to everybody (open data). For example, the number of active social media accounts is about 2.078 Billion in 2015 [2]. Based on these kind of data sources, it is possible to analyse and extract various information based on the needs.

For specific subjects, generating datasets of this magnitude is impossible due to time, financial issues or limited information available regarding the specific subject. Our project is based on a real use case that is not generating enough data to let the Deep Learning algorithm be the most performant. Furthermore, as Criminisi [3] demonstrates, Decision Tree algorithms for computer vision have interesting accuracy since 2016. That is the reason why we are comparing the algorithms efficiencies on a specific dataset.

Our use case is based on precision agriculture, which aims to increase the productivity and maximize the yield of a crop. It can benefit the entire crop cycle through an application of the correct amount of inputs (such as water, fertilizers, pesticides or fungicides) at the exact time and place, or by detecting diseases in plants [4].

Today, precision agriculture has not been widely applied to vineyards or to complex pattern fields due to mixed, complex landscapes, and complex topographies, and is limited mainly to yield monitoring or remote sensing.

To provide a better management of vineyards or other highly diversified cultures, new services are currently under development. Examples of these services include automatic collection of data related to a culture, automatic spotting, designation of zones, and prediction of the evolution of a zone with a drone [4].

Cartographic data are currently used to detect specific and characterized natural zones, especially through satellite images [5][6].
These needs drove this research to detect vineyards and their potential weaknesses from infra-red images taken by UAVs. Following this detection, a new flight plan is generated for the drone, carrying treatment products for specific regions of the agricultural surface.

The image recognition process and object detection within aerial images are widely used. This process can be used for research purposes (for example, for cell detection [7], satellite image mapping [8], as well for everyday uses to find matching images like the service proposed by Google to find similar images). However, the image recognition from satellite and drone images is expensive.

The current article provides the state of the art in the first section. The dataset is then detailed, followed by the settings description. The fifth section presents the results. The algorithms are then compared and finally a conclusion is presented.

II. State of the art

A recent study of Geitgey [9] demonstrates the performance of the CNN and Deep Learning algorithms with bird's images

classification. The training dataset contains 12,000 birds' pictures and 52,000 images not containing birds. The training and testing processes are performed using TensorFlow, a deep learning platform provided by Google. The final accuracy of the CNN is of 95%, with a precision of 97.11% and a recall of 90.83% [9]. The main sources of error were based on the images containing planes, or bird species out of the ordinary (like hen or ostrich).

Many research projects are currently underway to improve the quality of the classification and detection of natural changes. To increase the quality of the image recognition, different algorithms are applied. The first methodology used is a pixel-based classification (PBC) and the second is an object-oriented classification (OOC) [10]. The PBC is based on conventional statistical techniques, it categorizes all the pixels into thematic classes using specific features such as spectral information. For each pixel, a spectral signature is estimated by the relative reflectance in different wavelength bands. With the new high-resolution (HR) dataset, this classification is now limited. The OOC is more broadly used and is slowly replacing the PBC. Rather than classifying each pixel, the OOC classifies objects present in the image. It is based, for example, on spatial-spectral characteristics of the high-resolution satellite data information. The OOC's development over the last decade was mainly based on rule-based and nearest neighbor (NN) classifiers.

The power of PBC lies in the combined spectral response from all training pixels. Therefore, the classification method ignores the impact of mixed pixels present with the OOC[11]. In contrast, the PBC cannot use the relationships between pixels or be applied to what the author called "hyper spatial" data [12]. The main weakness of PBC is the recognition of objects much bigger than a single pixel [13].

OOC's superior performance stems from its use of spectral information like land types, shapes, texture parameters and relationships between contexts. The inclusion of the vector-based GIS environments also enhances the performance, and the accuracy is improved by the usage of a large set of features such as the spatial, textural and contextual properties [14][15][16].

The third classification methodology for aerial images is the nonlinear scale-space filtering. The author's framework computes different scales of segmented images and is connected to a kernel-based learning machine for the classification of various earth-observations [17]. Creating this framework demonstrates that the multilevel object representation and support vector machine classifier possesses the desired qualitative properties for classification experiment (See Fig. 1). This eliminates the need for tuning several parameters during the segmentation. The current algorithms outperform the previous developments regarding the representation and the classification results.

III. Dataset

The dataset is composed of images from 5 vineyards in Valais, Switzerland taken by a drone. The file format to keep
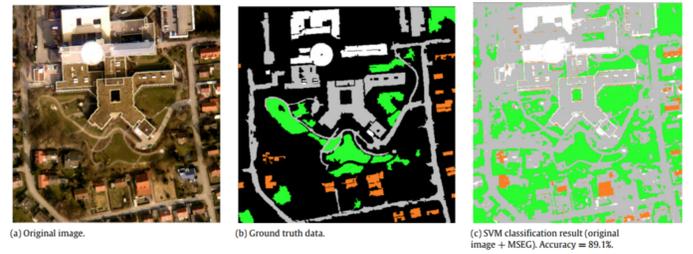

Fig. 1. Classification results with and without AML filtering. Green: vegetation, grey: asphalt, orange: tile roofs, white: bright roofs.

the geo information is GeoTIFF, created with the software Pix4D. Images are handled through a GIS software or a Python script to extract the relevant information (Bioformats for example) for further classification.

For a precise classification of specific zones, a subdivision of the images is necessary. This division results in 30x33 pixels tiles. Each tile contains mainly one of the 3 categories: (1) Road, (2) Vineyard or (3) Other (See Fig. 3). The size of the tiles was calculated through multiple experiments to fill two key criteria. Each tile must be large enough to be classifiable, but not too large so as to contain multiple classes. This represents 2'601 images per vineyards, which is 13'005 images for the full dataset. For each category, 2'082 images represent (1) Roads, 7'022 images represent (2) Vineyards and 3'903 represent the (3) Other category (See the summary in the Table I). Each image is manually labelled.

The dataset is divided into 2 main partitions for training and testing (See Fig. 2). To obtain the most accurate results, the train set is composed of 90% of the dataset taken with a balanced selection and theremaining 10% remain for the testing process.
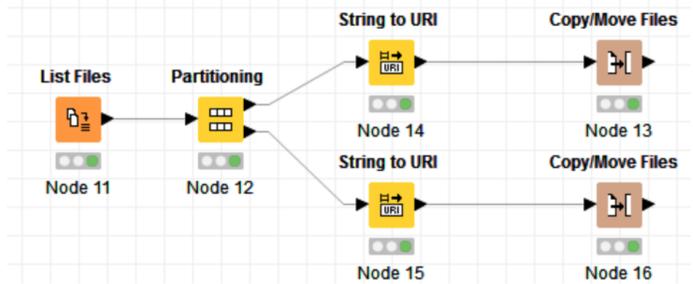

Fig. 2. Example of a process to divide the dataset into two subsets: train and test. This process is created on an open source platform for machine learning.

IV. Settings

This project is based on two types of settings for each tiles depending of the classification algorithm used. The settings influencing each algorithms are described below.

A. Deep Learning

The Deep Learning algorithm is applied using TensorFlow, a Deep Learning platform provided by Google. The dataset
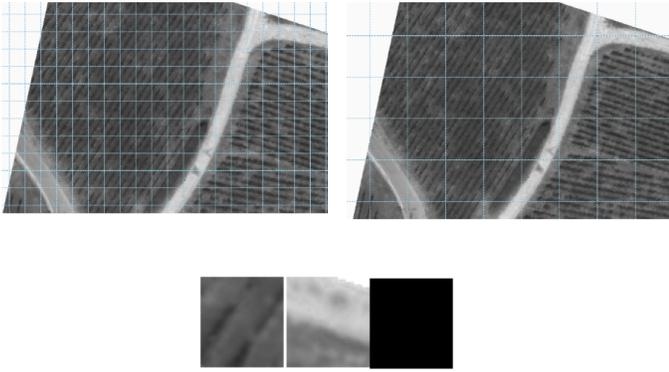
Fig. 3. Vineyard split with tiles of 30x33px (left) and 78x80px (right). On the left, some objects are on the same tile, but rarely a vineyard and other. On the right, objects are mixed,

is adapted to match the requirements of the TensorFlow platform. The classification is based on the most powerful model provided by Google, the Google Inception Model. This model is pre-trained to recognize over 1000 different classes of objects, but can be trained further to detect custom classes suited to our needs [18].

For the model retrain, the following parameters are applied:

- Bottleneck size: 2048
- Model depth: 3
- Model max width: 340
- Model max height: 340

The model is increased and adapted to include the new three classes in the main Inception Model. A new model is avalaible using the retrained graph based on Inception for validating and testing. The reduced model is set to use the new classes only. The result provides probability of each classe.

B. Decision Tree Ensemble

The Decision Tree Ensemble (DTE) is based on features extracted from each tile of the vineyard. This extraction process needs to be applied before training or testing. The following features are extracted from each tile:

- First order statistics: Min, max, mean, geometric mean, sum, variance, skewness, kurtosis...
- Tamura: Granularity, Contrast, Kurtosis of directionality, Standard Deviation Directionality, Max Directionality and Skewness.

TABLE I
FULL DATASET SUMMARY OF DRONE IMAGES AFTER SUBDIVISION INTO TILES. EACH CATEGORY CONTAINS THE NUMBER OF TILES MANUALLY LABELED.

Class	Number of tiles	Size of tiles
vineyard	7'022	30x33 pixels
road	2'082	30x33 pixels
other	3'903	30x33 pixels
Number of tiles per class	13'005	
Total number of tiles	2'601	

- Haralick: Statistical features based on gray-level co-occurrence matrix.

In total, 86 features are extracted from each images. To reduce the noise caused by too many parameters, an elimination process is set up (See Fig. 4). This elimination tests the importance and impact of each feature on the final accuracy. Some eliminated features are duplicates for example, or irrelevant for a black and white image.

Following the elimination, an overall error is presented for each combination of input parameters. The smallest and most powerful combination for this case uses only 16 features. This reduces the error to 0.047 (See Fig. 5).

Based on our iterative experiments, algorithm-fine tuning is done over similar dataset sampling. The final configuration of the Tree Ensemble Learner uses the information gain ratio as a split criterion. The tree depth and the minimum child node size are not limited. Hundred trees are used for the model and are generated on a stratified dataset sampling and using the square root method for attributes sampling.

V. RESULTS

The results below are obtained after multiple iterations of each algorithm and specific tuning. They present a detailed analysis for each of the two algorithms used. They show the difference in terms of accuracy, as well as stability over subsequent iterations.

A. Deep Learning

The Deep Learning has been trained multiple times to obtain the best validation accuracy. Each iteration had a different training dataset and tuning to maximise accuracy and stability. The finale size of the train to test ratio is adapted to 80% of the overall original dataset used for training with a stratified sampling. The accuracy obtained during the training process using these settings is of 93.5%.

Following the training process, a testing process is established. Indeed, errors are not highlighted with the TensorFlow platform. The definition of the classification accuracy type is defined to reach the research goals of having the best accuracy overall or reducing the false positive in a specific class to obtain the best unique class detection. In our case, the best accuracy overall is the main goal of the project.

The test process defined applied the trained model over each images of the testing dataset. The process is applied with a Python algorithm creating a comparison of the original source class to the predicted class. After multiple iterations, the final accuracy is of 91.2% with a standard error of $\pm1.32\%$. The misclassification is mainly present between vineyards and roads (See Table II). This missclassification is due to the presence of multiple classes on the same tile or a specific vegetation overly present in a class.

B. Decision Tree Ensemble

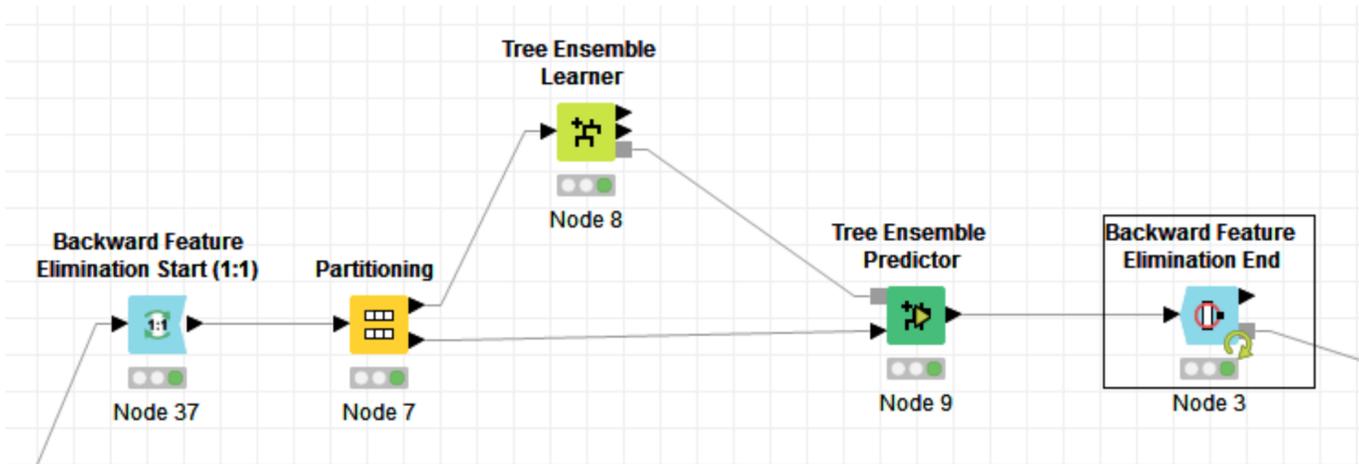Once the Decision Tree Ensemble training and fine-tuning is complete, the trained model is applied to the training set

Fig. 4. Workflow for features elimination process. To increase the quality of the final prediction.

Fig. 5. Backward Feature Elimination Filter. On the left, the error is linked to the number of features used. On the right, the column to be selected (depends on the left selection). The two selected features are the variance (intensity variation of the image's pixels) and the entropy (the amount of information present) of the image.

on each tile. A strict process is applied to have a comparison point at the end of the research and it is based on the following steps:

- Image meta data extraction: Image reader, image labelling and features extraction
- Feature elimination: Based on the model generated during the training process to keep and delete the same columns
- Prediction: Applying the model created before on the test set

TABLE II
THIS TABLE PRESENTS THE CONFUSION MATRICE OF THE DEEP
LEARNING CLASSIFICATION WITH THE OVERALL RESULT AND STANDARD
ERROR.

Class / Prediction	vineyard	road	other
vineyard	124	19	0
road	4	40	0
other	0	0	75
Accuracy	91.2 %		
Std error	±1.32 %		

- Evaluation: Generating a scorer to understand the performance of the classification

The final overall accuracy of the algorithm is of $94.275\,\%$ with a standard error of $\pm1.14\,\%$ (See Table III). The vineyards are very well detected and only a small portion is misclassified. This algorithm is very stable throughout all iterations and provide a similar accuracy. The misclassification is similar to the Deep Learning experiment and is due to the overlapping of objects on the same tile.

An exploration of the trees generated (See Fig. 6) is necessary to further understand the sources of error and be able to correct and adapt the classification process and algorithm.

TABLE III
THIS TABLE PRESENTS THE CONFUSION MATRICE OF THE DECISION TREE
ENSEMBLE CLASSIFICATION WITH THE OVERALL RESULT AND STANDARD
ERROR.

Class / Prediction	vineyard	road	other
vineyard	136	7	0
road	8	36	0
other	0	0	75
Accuracy	94.275 %		
Std error	±1.14 %		

VI. ALGORITHM COMPARISON

The performance of both algorithms are acceptable. However, the Tree Ensemble model is the most accurate for two reasons. Quite obviously, the overall accuracy is much more precise. Indeed, after multiple iterations, an accuracy of $94.275\,\% \pm 1.140\,\%$ shows a very precise recognition of objects from the tiles. Second, the number of missed branches of vine stock in tiles is almost non-existent, which increases the overall quality of the algorithm whose objective it is to detect the vineyards.

Furthermore, according to our experiments, we have observed that the number of tiles required to have a stable and
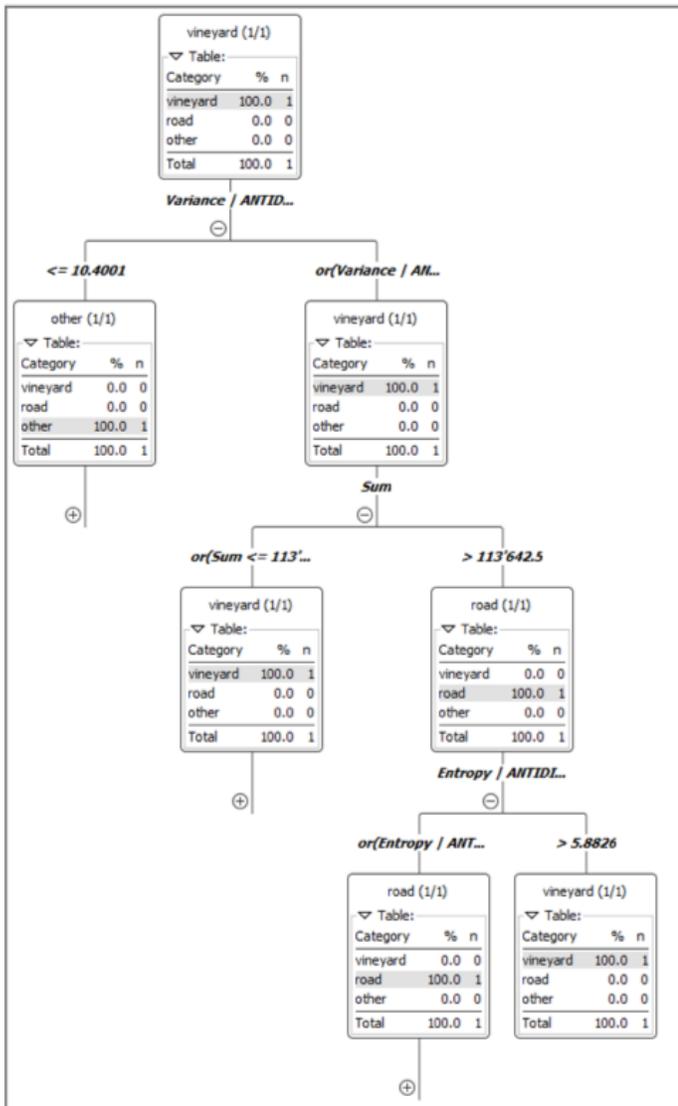
Fig. 6. Overvieew of the structure of the Decision Tree Ensemble model created on the dataset presented above. This extract of the structure shows that the first feature to separete the dataset is the VARIANCE / ANTIDIAGONAL which was selected during the features selection process.

accurate model is significantly smaller with a Decision Tree Ensemble algorithm than with Deep Learning. Using the DTE for small datasets allows us to produce a more reliable model for classification. Based on all possible features extracted from the images, once they are reduced with the feature selection, the processing time for prediction is very quick and can be run on small computational power machines, such as a tablets for drone pilots.

The Deep Learning algorithm is very simple to use as well since multiple platforms are available to exploit it. For a quick use without manipulation to obtain a first evaluation, these platforms (such as TensorFlow) can be used with an existing model (Inception in this case). Incresing the size of the Neural Network is possible but not very accurate. Indeed, in this research, we showed that the Deep Learning's accuracy is performant but not compared to a Decision Tree algorithm.

Moreover, it is difficult to understand the Neural Network model to see the potential errors and to correct them within these Deep Learning platform. Creating a new model from scratch, although more precise, would be very demanding in terms of processing power. Indeed, the accuracy is interesting, but the number of missed vineyards is important and should be reduced to be competitive against a Decision Tree algorithm.

Based on the previous presented accuracy of algorithms, a merging of the result from Deep Learning and Decision Tree will improve the overall accuracy. Indeed, the DL performs better on the road classification than the DTE. And the best result for the DTE is with the vine detection. Two possible solutions can be implemented:

1) Majority vote: With an implementation of a thrid algorithm, a majority vote is possible to increase the classification accuracy.
2) Threshold: Combining the two results of the DL and DTE, calculation of the mean value to be used as a threshold for classification.

VII. CONCLUSION

This paper shows a comparison of performances with advantages and disadvantages of the Deep Learning and Decision Tree Ensemble algorithms based on a specific use case. This research allows to classify objects present on the tile without further manipulation, based on the Deep Learning methodology and tree ensemble algorithms. It shows the possibility of algorithms with more "hands off" approaches, such as Deep Learning, and more "hands on" one for the Decision Tree Ensemble.

The comparison informs us that the performance of the Decision Tree is higher and more stable with a reduced computational performance and using specific reduced features. Even with a small dataset, the performance to detect vineyards without any further image manipulation reached 94.275%.

However, because of the reduced number of vineyard images present in the original dataset, it is not impossible that the model is not complete. This should be handled and controlled with more aerial vineyards images. These aerial images need to be taken in various environments, seasons and regions at different altitudes to be able to complete our model for our use case.

This work allows us to understand the possible improvements of Deep Learning methodology to reduce the dataset size keeping the performance as long as you adapt the algorithm. This improvement is part of the next steps of this research to understand and standardize the process for various type of images.

Furthermore, to improve vineyard detection and reduce errors of the image classification (reduce false positives), the size of the tiles need to be reduced. Indeed, each tile potentially has multiple objects present within it. This confuses the algorithm as the pattern and colors of each class are very similar.

Difference of the altitude from the drone's viewpoint can be added to differentiate vegetation on the ground from vineyards. This should improve the overall accuracy for both algorithms.

REFERENCES

[1] LeCun Y., Bengio J. and Hinton G., *Deep Learning*: Nature, 521, p.436-444, 2015

[2] Treboux J., Cretton F., Evéquoz F., Le Calvé A. and Genoud D., *Mining and Visualizing Social Data to Inform Marketing Decisions*: IEEE Advanced Information Networking and Applications (AINA), Switzerland, p.66-73, 2016

[3] Criminisi A. and Shotton J., *Decision forests for computer vision and medical image analysis*: Springer Science & Business Media, 2013.

[4] Agrofly, *High Precision Forecast with Agro Drones (UAV)*, SPSS, Sierre, Switzerland: CTI, 2016.

[5] Popkin G., *Satellite alerts track deforestation in real time*: nature.com, 530, p.392-393, 2016.

[6] Karakizi C., Oikonomou M. and Karantzalos K., *Detection and Vine Variety Discrimination from Very High Resolution Satellite Data*: Remote Sens, 8, p.235, 2016.

[7] MathWorks, *Detecting a Cell Using Image Segmentation*: [Online] Mathworks.com

[8] Schultz B., Immitzer M., Formaggio A. R., Del' Arco Sanches L., Barreto Luiz A. J. and Atzberger C., *Self-Guided Segmentation and Classification of Multi-Temporal Landsat 8 Images for Crop Type Mapping in Southeastern Brazil*: Remote Sensing, 14482-14508, 2015

[9] Geitgey A., *Deep Learning and Convolutional Neural Networks*: [Online] Medium.com, 2016

[10] Jawak S. D., Devliyal P. and Luis A. J., *A Comprehensive Review on Pixel Oriented and Object Oriented Methods for Information Extraction from Remotely Sensed Satellite Images with a Special Emphasis on Cryospheric Applications*: SciRes, 4, p.177-195, 2015

[11] Lu D. and Weng Q., *Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery*: Photogrammetric Engineering and Remote Sensing, 70, p.1053-1062, 2004

[12] Sridharan H. and Qiu F., *Developing an Object Based Hyperspectral Image Classifier with a Case Study Usin WorldView-2 Data*: Photogrammetric Engineering & Remote Sensing, 79, p.1027-1036, 2013

[13] Corcoran P., Winstanley A. and Mooney P., *Segmentation Performance Evaluation for Object Based Remotely Sensed Image Analysis*: International Journal of Remote Sensing, 31, p.617-645, 2010

[14] Chen Z., Ning X. and Zhang J., *Urban Land Cover Classification Based on WorldView-2 Image Data*: IEEE International Symposium on Geometrics for Integrated Water Resource Management, Lanzhou, 2012

[15] Aplin P. and Smith G., *Advances in Object Based Image Classification*: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37, p.725-728, 2008

[16] Liu D. and Xia F., *Assessing Object-Based Classification: Advantages and Limitations*: Remote Sensing Letters, 1, p.187-194, 2010

[17] Tzotsos A., Karantzalos K. and Argialas D., *Object-based image analysis through nonlinear scale- space filtering*: Journal of Photogrammetry and Remote Sensing, 66, p.2-16, 2011

[18] Google TensorFlow, *Image Recognition*: [Online] https://www.TensorFlow.org/tutorials/image_recognition