

Evaluating Classification Algorithms for Improved Wastewater System Calibration

Mohammed Chachan Younis¹, Edward Keedwell², Dragan Savic³, Anthony Raine⁴

^{1,2} Department of Computer Science. ³ Centre for Water Systems.

College of Engineering, Physics and Mathematics. University of Exeter

⁴ BSc (Hons) C.WEM MCIWEM. Pell Frischmann

¹mcy201@exeter.ac.uk

ABSTRACT

Hydraulic models provide an approximate model of rainfall collection (storm), wastewater collection (foul), and combined (both rainfall and wastewater collections) network performance, capturing the large scale element of the system, but such systems require calibration with real world data to achieve reliable and accurate results. However, many factors surrounding real-world network and catchment characteristics are unknown and can influence the hydraulic performance of the network. Consequently, calibrating wastewater models to accurately reflect real-world conditions is a time consuming and complex process. This paper employs a two stage urban runoff forecasting approach with a combination of image classification techniques and modelling using InfoWorks ICM. The image classification section consists of the automated processing of the satellite image as an aid to modelling wastewater in the urban environment by classifying the land-cover as pervious and impervious segments of water from rainfall events. By classifying such areas with an urban catchment, and by using the Wallingford PR Equation across all tests of Area 1 (roads), 2 (roofs) and 3 (permeable area), we explore the potential for a partially automated network system calibration process of a wastewater network.

Keywords: *Urban drainage, InfoWorks ICM, Image classification*

1 BACKGROUND

To begin to simulate urban flow systems, it is necessary to model the flow of water over the drainage basin; by determining the relevant drainage, rainfall, and elevations (levels) of the landscape. Throughout man-made water flow systems, modelling successfully imitates and allows the prediction of the dynamics of temporary surface retention (ponds), which illustrate the flow across the urban catchment area along designated water flow highways [1]. Software simulating urban hydrological models, for example, InfoWorks or SWMM, create simulations of flows of underground pipelines and surface runoff; additionally, it can also be utilised to estimate inundation conditions. A contemporary addition to the range of models provided by the range of image classification techniques is overland flow; by utilising precise high-resolution digital elevation model (DTM) data, which shows the height of the surface, it is possible to forecast urban modelling successfully. A recent modification to these models is a description of overland flow, which enables the modelling of urban flooding using high-resolution accurate DTM data.

This paper employs a two stage urban runoff forecasting approach as a combination of image classification techniques and the InfoWorks ICM (Integrated Catchment Modeling) model. The proposed approach aims to classify the satellite image into pervious and impervious segments, and using the result of this classification with an uncalibrated InfoWorks ICM model to improve the calibration (the process of making the model fit data through the network of a particular real life) of a wastewater network.

The image classification section consists of the automated processing of the satellite image as an aid to modelling wastewater in the urban environment by classifying the landcover as permeable and impermeable segments of water from rainfall events. By classifying such areas with an urban catchment, and using a DTM, we explore the potential for a partially automated network system calibration process and determine the landscapes elevation. A typical sample with the urban catchment of image size 3456x4992 pixels in the associated DTM, and the image for the village of Feock, Cornwall in the southwest England, used in this study is given in Figure 1 on the left. On the right is the ground truth (GT) image of Feock, which was generated manually.

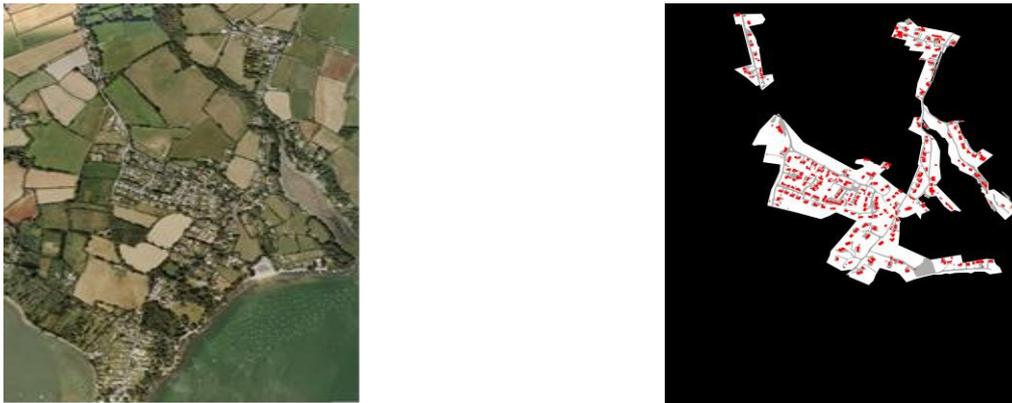


Figure 1. Feock test image: Satellite raw image (left). Ground truth image (right)

Remote sensing image classification is a challenging problem because of many factors, such as:

- the complexity of the landscape;
- remotely sensed data;
- and image processing.

Therefore classification approaches have a major impact on the success of this classification process. There are a vast majority of methods to analyze and classify images, including parametric statistical methods or non-parametric techniques such as neural networks, fuzzy inference systems and fuzzy neural systems. One of the recent methods suggested for land cover classification of images is ensemble methods, which is a family of algorithms that has been used in many data mining applications. However, these methods are not fully explored for analyzing remotely sensed images [2,3]. In this work, we used Classification Trees (CT) and Random Forests (RF) as a member of ensemble methods. We analyzed accuracy, reliability and training size dependency of CT and RF as ensemble methods in pixel based satellite image classification by using different tree sizes. The results show that, by using properly dimensioned RF, efficient, accurate and reliable classification of the image is possible to obtain permeable and impermeable areas.

The paper is organized as follows. In section II, we briefly review modelling approaches in assessing urban flooding and the design of urban drainage system. Section III presents a brief overview of CT and RF used for satellite image classification. Section IV discusses the experimental work in which we used a MatLab based performance evaluation of CTs and RFs, and this section presents the results obtained. Section V concludes the work with main highlights and some possible future research areas.

2 MODELING URBAN DRAINAGE

2.1 Hydraulic models

Hydraulic models provide an approximate model of wastewater network performance, capturing the large scale element of the system; however, these systems require adjustments according to real world data, to maximise accuracy in measured outcomes. There are urban hydrological models, such as InfoWorks ICM which can simulate underground pipeline flow, surface runoff flow, and estimate inundation conditions. These models have been widely used in urban flood planning and management, but not used so much in urban flood forecasting [4].

2.1.1 InfoWorks ICM Model

InfoWorks ICM software by Innovyze was adopted, in this study, to improve the calibration of a wastewater network. By using the image classification technique with InfoWorks ICM, increasingly, engineers can work with a sole, unified model; by incorporating a range of environmental variables in a hydrological system. By adding drainage assets such as manholes, sewers and other wastewater ancillaries in combination with natural and man-made water channels to create a model in which the catchment and the floodplain are one, as they are in real life. The additional ability to factor-in bridges, sluices, weirs and pumps allows the creation of models that are more complete and more accurate than ever before [5]. The software used for the creation of the model was version 7.5.4 of InfoWorks ICM.

2.1.2 Permeable and impermeable areas

In order to predict the behaviour of rainfall runoff, it is important to identify the runoff surfaces which are divided into two main categories: impermeable (for example, roads or roofs); and permeable (for example grass). This identification helps in the indication of the amount of runoff volume, which is the part of the rainfall which does not absorb through the soil or is lost through either depression storage (puddles), interception, evaporation or evapotranspiration.

3 OVERVIEW OF IMAGE CLASSIFICATION

Commonly, conventional methods to analyse and classify images include parametric statistical methods or non-parametric techniques namely neural networks, fuzzy inference systems and fuzzy neural systems. Established statistical approaches are often pixel-based and tend to include the following: the maximum likelihood classifier, minimum distance classifier, and clustering techniques. In comparison with conventional classifiers, neural networks supply a robust alternative due to the fact they afford robustness and tolerance. Additionally, clustering algorithms such as fuzzy K-means and fuzzy C-means have been utilised for image analysis and classification. These are highly popular owing to their ability to produce higher classification accuracy than traditional methods most notably for small training datasets [6-8].

Categorical data sets, a notable example being land cover classification, are used in the creation of CT. A classification tree comprises of a set of tree-structured decision tests working by means of a divide-and-conquer way. Accordingly, each leaf node has an associated label, which will be assigned to instances falling into this node. A predicted outcome is that when a series of feature tests are conducted starting from the root node, that the result is attained when a leaf node is reached [9]. Most recently, ensemble methods have been proposed for land cover classification; in this instance, several classifiers are trained; their results are combined through a process of voting. Conventional learning approaches can be contrasted with ensemble methods; the former attempts to construe one learner from training data, whereby the latter tries to construct a set of learners and combine them. Additional terms for ensemble learning are committee-based learning and learning multiple classifier systems [10].

Amongst popular, efficient and respected techniques is RF, which stands out amongst the multitude of ensemble approaches; due to its boosting and bagging methods. The technique is based on an ensemble of tree classifiers; it grows many classification trees and the input vector is classified by each single tree contained within the forest. The RF Method has clear advantages; not only is nonparametric, it also demonstrates a way to highlight the importance of individual variables in classification; additionally, it is suitable for multisource classification [11]. A response variable (e.g. land cover, percent tree cover) is computed by the RF Method, by creating many (usually into the hundreds) different decision trees (the forest of trees), following on from this, each object to be modelled down each of the decision trees. The response is then calculated by evaluating the responses from all the trees in the forest. Regarding classification, whichever class is predicted most regularly results in being the one assigned to the corresponding object. The key to the success of RF is how it creates each of the decision trees that make up the forest [12].

4 EXPERIMENTAL EVALUATION

4.1 Data and Study Area Image

The study was carried out in a small village, Feock, in the UK and the model set up based on the underground pipeline network, the DTM and a digital map of the village which contains spatial information outlining roads, buildings, residential areas, vegetation, watercourses, and estuary shoreline. Parameters for the model were derived based on the properties of the subcatchment and the storm sewer conduits. The small village Feock was used as the main test image for performance evaluation of all CT and RF algorithms.

4.2 Method

As a first step, we defined test and training areas, and selected parts in the training area were used for training the algorithms, and the selected remaining part of the image was used as the inputs for testing the algorithms and performance evaluation, along with confusion matrix results. The parts have been labelled according to the classification objective as two main classes: permeable and impermeable. In this performance comparison of algorithms work we used the following 3 classes: buildings, roads, and vegetated ground in both training and test areas.

The classification depending upon constructing pixel-based features was carried out by using the classification tree and random forest methods, and the relevant tools available in the MATLAB Statistics Toolbox. The classification accuracy for all the classification results was calculated by using the reference points of the validation areas, i.e., excluding the training area. The accuracy estimates mean accuracy of individual classes, and the overall accuracy of the classification were all included. Using the test image, we evaluated the performance of the 5 methods in terms of accuracy and reliability, along with mean testing and training errors, by changing tree size and also training set and pruning level parameters. These methods are CT and RF with 10, 20, 50 and 100 trees.

4.3 Results

As one of the evaluation steps, the confusion matrices (count cases) for corrected data as opposed to predicted GT data, percentages and mean training/testing error (incorrectly predicted pixels) for the Feock test image were obtained, as given in Tables 1 and 2.

Table 1. CT and RF Classification Performance

Classification Method	Mean training error	Mean testing error
CT	8.30%	16.90%
RF-10	2.10%	15.10%
RF-20	1.30%	14.60%
RF-50	0.90%	14.40%
RF-100	0.80%	14.20%

Table 2. Confusion matrix for the three classes

		Feock data			
		predicted value			
Real value	RF 100	Housing	Vegetation	Roads	Total of GT
	Housing	1036865 (4.4%)	360564 (1.4%)	971071 (3.4%)	2368500 (9.2%)
	Vegetation	322458 (1.3%)	16101759 (63.6%)	1020553 (2.6%)	17444770 (67.5%)
	Roads	908545 (4.8%)	2457072 (10.0%)	2656913 (10.3%)	6022530 (23.3%)
	Total of predicted	2267868 (10.5%)	18919395 (74.9%)	4648537 (14.6%)	25835800 (100.0%)

In terms of actual conclusion from these tables, observing the mean training error in Table 1, indicates that RF50 has almost same error value as RF100 however RF50 is half the size of RF100 meaning less computational demand and less time to train. The mean testing error, in the same table, shows that RF100 has better accuracy among all the classification methods mainly because of the greater capacity (since it contains much more detail) it has to describe data. Based on the overall relevant results it can be concluded that the best performance among all is the one provided by RF100. Table 2 illustrates the confusion matrix description of the performance of the classification model of RF100. The confusion matrix contains the number of pixels in each situation and the process is repeated for all the known pixels of the GT to get the confusion matrix. Each row represents a pixel of real values, and the columns represent predicted values. Note that the table contains pixels from 10 runs/rounds (in order to get a more precise figure), where the Feock image size is 3456x4992 pixels with total number of pixels 17252352, but only 2583580 known pixels x10 rounds = 25835800.

4.4 Calibrating a model

The idea of the current study is to classify the land-cover of the satellite urban areas images as permeable and impermeable segments, then using the result of this image classification with an uncalibrated InfoWorks model to improve the calibration of the wastewater network. By classifying such areas with an urban catchment, we explore the potential for a partially automated wastewater system calibration process. There is only 12 runoff surfaces we can have in any model, but we can assign at least 99 runoff parameters which are assigned in the Land use table, which effectively means we can have hundreds of differing runoff surfaces if we really wanted to, but modellers usually keep things simple and rarely differentiate between different urban/highway surfaces etc. In this study, we calibrate the model based on our classification results by changing the original runoff areas 1, 2, and 3 in the Subcatchment table of the model. Each subcatchment can be categorised into as many as 12 different contributing surface types (runoff areas). However, traditionally only the first three are used where: Runoff Area 1 = Roads; Runoff Area 2 = Roofs; and Runoff area 3 = Permeable Areas (for example grass). The other 12 are called Runoff Area (1-12) Absolute and define the area in hectares or acres (depending on which catchment area unit is being used) [5]. The Runoff Area fields define how much of the subcatchment belongs to the particular runoff surface type which can be assigned in own unique runoff characterises using various runoff models and coefficients.

The Wallingford Runoff Model (which historically has been used to predict runoff from urban catchments in the UK, although the New PR Equation and even more recently the UKWIR runoff

models are being applied as a replacement to the New UK PR Model) is used for predicting runoff from both impervious (Areas 1 Paved and 2 Roof) and pervious (Area 3) areas uses ‘Contributing Area’ (the area that drains into the system we are modelling) and not the ‘Total Area’ (the full area of the subcatchment, including those parts that do not drain into the system we are modelling) of the subcatchment. The Wallingford PR Equation uses a regression equation to predict the runoff coefficient depending on the density of the development, the soil type and antecedent wetness of each subcatchment [5]. In current work, we investigated three simulations of varying runoff areas which assigned to the Wallingford Runoff Model in a winter rainfall event, Figure 2:

Figure 2 (up) was run with original calibrated model: In this part, we ran the InfoWorks simulation of Feock map without any change to the runoff areas parameters, where the software considered road, roof, and pervious surfaces have runoff indices of 10, 20, and 21 respectively. Figure 2 (middle) was run with uniform subcatchments: to make a general judgment about the whole Feock area, we used the predicted values (percentages) of our algorithm results with the ground truth results to see how much improved the results by feeding the model with new runoff areas. Figure 2 (bottom) was run with differential subcatchments: another simulation is investigated to know how much better can we make to the model’s output is by use smaller subcatchments. In this case, we divided Feock map to 54 polygons, and then use the result of every subcatchment individually with corresponding GT and the actual raw image.

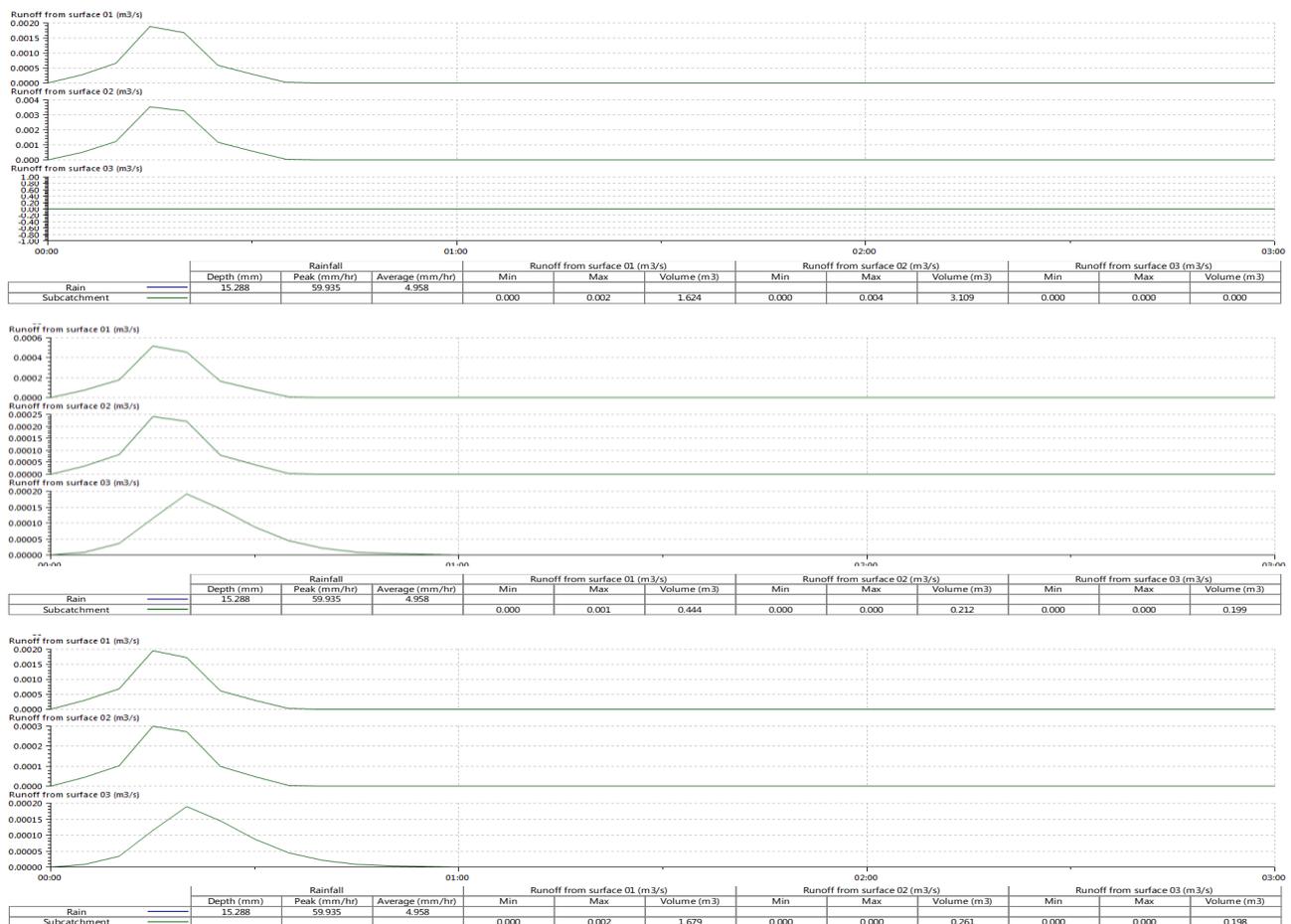


Figure 2. Rainfall profile of subcatchment ID no. SW82383501 from the three runoff surfaces (paved, roof and pervious areas): Run the simulation with original calibrated model (top). Run the simulation with uniform subcatchments (middle). Run the simulation with differential subcatchments (bottom)

Table 3. Summary results from simulation of event M5-30 storm period of all subcatchments

Input case	Total rainfall m ³	Total runoff m ³	Total inflow m ³	Total outflow m ³
Original	109.2	82.8	90.7	60.4
Predicted	109.2	14.9	22.7	17.3
Subcatchments	109.2	16.9	24.7	21.1

4.5 A discussion of the results

In this study, InfoWorks ICM was used and the results showed that if we have more impervious areas then we have more flow in the pipes because, obviously, more rainfall runoff is generated. With regard to our classification software, quite a lot of the example study area is pervious in reality, but the InfoWorks model presents this study area as mostly impervious. The graphed results in Figure 3 for a selection of subcatchment ID no. SW82383501 allows results from selected objects to be compared.

The simulation results show that the proposed approach is a promising method for obtaining more accurate calibrations of wastewater systems through a partially automated methodology, reducing the requirement for engineers to perform this calibration manually. The results in Table 3 show the simulation with the original calibrated model has much more runoff than we have in our model. When we do the total percentage application with our model, there is a small difference between the applications of pervious/impervious uniformly to the entire model, and the application to individual subcatchments. The explanation for this may be that we took a small portion (Feock) as a case study, which is a very small example section of the Truro drainage area study map. There is about 6 times less runoff in our predictions compared with the original calibrated model. This vagary is justified due to the runoff from the surrounding green areas, which includes runoff from surrounding fields and/or slopes. This catchment model was originally verified in the Autumn of 2001 which was an extremely wet period (circa 1 in 200 year 3 month event) in this part of Cornwall. This is likely to have led to additional runoff being generated from surrounding pervious areas which wouldn't normally occur.

Other significant information from table 3 refers to total inflow. Inflow is an object reference for how water enters the system. The total inflow defines inputs (base flow, foul flow, runoff, infiltration, and inflow) to the subcatchment. We can find the value of foul flow from the total inflow by subtracting total runoff from total inflow, which is 8 m³ in all three cases. The final column in Table 3 concerns total outflow. The object referenced total outflow defines outputs (base flow, foul flow, runoff, infiltration, and outflow) from the subcatchment. These outputs should be less than the inputs because of processes such as evaporation, losses to groundwater through the streambed etc. Table 3 illustrate this clearly.

5 CONCLUSIONS

In this work, we evaluated and compared the performance of machine learning classifiers, namely CT, and RF, in classifying images, using a pixel-based classification procedure. The results showed that RF was superior to CT. Each of these classifiers has their associated advantages and disadvantages; therefore, the choice of the appropriate one may be case dependent. A proper RF seems to achieve relatively high mean accuracy levels with a relatively small amount of training samples, but the setting of pruning parameters and training sizes are important for all methods.

One of the main concluding points is that the performance results get better with the higher number of trees; where CT or one tree is the worst performing among all the compared methods. However, the performance improvement above a tree count of 50 is slow, which indicates that the maximum

necessary complexity has already been reached, and further increase in the tree count just makes the algorithm slower without much improvement in performance.

Even though the additional runoff in the original calibrated model contrasted with our predictions, it is probably that our approach is valid for many urban situations, especially during summer storms when the runoff from grassed areas (verges / gardens etc.) tends to be minimal.

An objective for any classifier is a high accuracy; therefore, classifier robustness (as being independent of the image used for the same category of images: for example, for urban land cover classification) performance improvement is necessary. We observed that the classifier performance might change with image size/resolution and image content (number and distribution of objects in given image), i.e. the number of objects detected in the image and its partition into training and testing sizes are important. Apart from the number of training samples used in classification, the parameters of the classification algorithms/methods might have an impact on the overall classification accuracy as well. Consequently, appropriate setting of parameters for different machine learning classifiers and choice of proper training sample size are important factors which can be the subject of future research.

Generating a classification system to see where the subcatchments are in the predicted image to generate results without needing to run every polygon of the satellite / GT images individually, as was the case in the current study, would be less time consuming. This enhancement would be future research.

References

- [1] R. Allitt, J. Blanksby, S. Djordjević, Č. Maksimović, and D. Stewart, "Investigations into 1D-1D and 1D-2D urban flood modelling," *In WaPUG Autumn Conference*, vol. 25, 2009, pp. 1-12, Blackpool, UK.
- [2] S. Yan-yan, and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, 2015, pp. 130-135.
- [3] P. Du, G. Li, W. Zhang, X. Wang, and H. Sun, "Consistency measure of multiple classifiers for land cover classification by remote sensing image," *In International Workshop on Multiple Classifier Systems*, 2009 Jun 10, pp. 398-407, Springer Berlin Heidelberg.
- [4] L. Jiang, YA. Chen, and HU. Wang, "Urban flood simulation based on the SWMM model," *Proceedings of the International Association of Hydrological Sciences*, vol. 6, no. 368, 2015 May, pp. 186-191.
- [5] Innovyze. *InfoWorks ICM (Integrated Catchment Modeling) v.7.5.4*. User manual references, Wallingford, UK, 2017.
- [6] A. D. Kulkarni, and L. Barrett, "Random Forest Algorithm for Land Cover Classification," *Computer Science Faculty Publications and Presentations*, 2016.
- [7] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, 2011 May, pp. 247-259.
- [8] E. Zarogianni, TW. Moorhead, and SM. Lawrie, "Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level," *NeuroImage: Clinical*, vol. 31, no. 3, 2013 Dec, pp. 279-289.
- [9] R. Kohavi and J. R. Quinlan, Will Klossgen and Jan M. Zytkow, editors, "Decision-tree discovery", in *Handbook of Data Mining and Knowledge Discovery*, pp. 267-276, Oxford University Press, 2002.
- [10] P. Moacir and Jr Ponti, "Combining classifiers: from the creation of ensembles to the decision fusion," in *Conference on Graphics, Patterns and Images*, 2011, pp. 1-10.
- [11] A. Kulkarni and B. Lowe, "Random forest algorithm for land cover classification," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 4, pp. 58-63, 2016
- [12] N. Horning, "Random Forests: An algorithm for image classification and generation of continuous fields data sets," *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, New York, USA. 2010.