Remote Sensing of Environment xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example

Li Luo, Giorgos Mountrakis *

Department of Environmental Resources and Forest Engineering, State University of New York College of Environmental Science and Forestry, United States

ARTICLE INFO

Article history:
Received 12 August 2009
Received in revised form 11 January 2010
Accepted 16 January 2010
Available online xxxx

Keywords:
Hybrid classifiers
Impervious surfaces
Intermediate inputs
Intermediate results
Landsat ETM+
Partial classification

ABSTRACT

With the constant proliferation of computational power, our ability to develop hybrid classifiers has improved. Hybrid classifiers integrate results from multiple algorithms and often improve classification accuracy. In this paper, a hybrid classification framework was used to evaluate two research hypotheses: i) can manipulated results from prior classifiers ("intermediate inputs" (IIs)) improve classification accuracy in subsequent classification steps. and ii) is there an optimal dataset proportion for creation and usage of intermediate inputs. These additional intermediate inputs were based on spatial and texture statistics calculated on a partially classified image. The implementation of intermediate inputs on an impervious surface classification task using a 2001 Landsat ETM+ image from central New York was demonstrated. The results suggested that there was an average accuracy improvement of 3.6% (maximum 6.6%) by using intermediate inputs. These improvements were proved statistically significant by a *Z*-test and tended to increase as classification difficulty increased. The experiments in this paper also showed that there was an optimal point that balanced the number of pixels and pixel classification accuracy from prior steps used to produce intermediate inputs. Additionally, some traditional problems such as separation of impervious surfaces and soil were successfully tackled through intermediate inputs. The concept of the intermediate inputs may easily apply to other sensors and/or ground features.

1. Introduction

Impervious surfaces, defined as water impenetrable surfaces such as rooftops, roads, parking lots, sidewalks, and other man-made surfaces, have become a key indicator in urban environmental studies (Arnold & Gibbons, 1996; Schueler, 1994). Accurate estimation of imperviousness is of high significance for hydrology, land use planning, resource management and ecosystem studies. Remote sensing imagery provides a cost-efficient alternative to ground-based mapping and thus has been increasingly employed for impervious surface estimation.

Previous research explored classification approaches in order to estimate imperviousness based on their spectral and/or spatial characteristics. Typical classification methods include multivariate regression models, spectral mixture models, machine learning models and integration with geographical information systems. Early in 1980, Forster (1980) applied multiple regression analysis to derive linear equations relating each Landsat band with percentage of land use in the Sydney metropolitan area. Later, spectral mixture analysis was developed to map the imperviousness (Lee & Lathrop, 2005; Lu & Weng, 2006; Phinn et al., 2002; Powell et al., 2007; Wu & Murray, 2003). In these studies, imperviousness can be estimated using a linear summation of endmembers (spectral signatures of 'pure' materials). Bauer et al. (2005) developed a second-order polynomial regression model to estimate the relationship between the imperviousness and "tasseled cap" greenness by using Landsat TM imagery. Yang et al. (2003) presented a practical step-wise multivariate regression model for landscape impervious estimation through the synergistic use of ETM+ and high-resolution imagery. Machine learning models such as decision tree classifiers (Crane et al., 2005; Dougherty et al., 2004; Herold, 2003; Yang et al., 2003) and neural networks (Hu & Weng, 2009; Iyer & Mohan, 2002; Lee & Lathrop, 2006), have been widely applied to estimate pixel-based or subpixel imperviousness. Weng (2007) provided an excellent overview of impervious surface estimation using remote sensing techniques.

The majority of current approaches towards impervious surface detection employ a single classifier. Some studies have demonstrated that combining multiple classifiers can produce more accurate results than a single classifier approach (Breiman, 1996; Hansen & Salamon, 1990; Krogh & Vedelsby, 1995). However, only a few studies have focused on hybrid classifiers for estimating imperviousness. Steele (2000) constructed combinations from the spatial, *k*-nearest neighbor, and linear classifiers for land cover mapping of two Landsat TM scenes. The results showed that the combination of multiple classifiers produced substantial increased accuracies compared to a single classifier approach. Liu et al. (2004) presented a hybrid classifier method using a two classifier combination consisting of a decision tree and a fuzzy ARTMAP neural network. This approach combined results

* Corresponding author.
E-mail address: gmountrakis@esf.edu (G. Mountrakis).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.01.008

from these two classifiers using a hybrid scheme and derived a confidence map providing users with additional information on classification accuracy. Coe et al. (2005) adopted a hybrid approach which combined an object-oriented and a pixel-based classification approach to detect impervious surfaces through integration of various sources. Franke et al. (2009) applied a hierarchical multiple end-member spectral mixture analysis (MESMA) approach to map urban land cover. This approach was consisted of four levels of complexity ranging from the simplest level consisting of only two classes, impervious and pervious, to 20 classes that differentiated material composition and plant species. Mountrakis et al. (2009) developed an expert-based multi-process system to estimate the distribution of impervious surfaces. This method was composed of a series of processing steps and involved multiple classifiers. In each step, a different algorithm was applied to classify part of the dataset. Mathematically simple and computationally efficient classifiers were adopted during the prior steps. As the number of unclassified pixels decreased and the classification difficulty increased, complex and computationally intensive classifiers were progressively introduced.

In this paper, we tackle the problem of binary impervious surface classification. If any portion of a pixel occupies a constructed impervious surface on the ground, that pixel is assigned to the impervious class. A progressive classification methodology was used to assess two research questions:

i) Can we process results from prior classifiers (creating "intermediate inputs" (IIs)) to assist classification in subsequent steps and thus improving overall classification accuracy?
ii) Is there an optimal dataset proportion for creation and usage of intermediate inputs?

2. Study area and datasets

The study area is the region covered by a Landsat Enhanced Thematic Mapper Plus (ETM+) scene (path 15, row 30) acquired on April 8th, 2001. It is located in central New York covering an area of approximately 173 km×177 km. The Landsat ETM+ imagery with 30 m spatial resolution and 6 bands (blue, green, red, near IR and two mid IR bands) is used as the data source in this research.

Reference data was collected from 19 sites throughout the entire scene. Samples within the 19 sites are representatives of land cover classes in the study area and were produced from high-resolution aerial Emerge imagery (1999) or digital orthophoto quarter quads (DOQQ) (2001). Using on-screen interpretation and digitization techniques on the DOQQ images, 18 sites containing land cover classes of bare soil, crop, snow and wetlands were selected. Site 19 covered the entire city of Syracuse, NY with land cover classes of bare soil, forest, grass, water and impervious surface. Reference data within this site were produced from Emerge imagery with 3 bands (near infrared, red and green) and 0.61 m spatial resolution. Although there is a 2 year gap in the acquired dates between Emerge imagery and Landsat image, the land cover distribution remained steadily across the city of Syracuse due to minimal socioeconomic changes. As a result, the Emerge image acquired in 1999 is a reliable reference source for the classification of 2001 Landsat ETM+ image in this research. A land cover map with five land cover classes (tree, grass, bare soil, impervious surface and water) was created by Myeong et al. (2001) from the Emerge image of site 19 by using a "hybrid" or "guided clustering" method (Bauer et al., 1994) and was directly used to derive the impervious reference. Landsat pixels overlapping any impervious pixels on the derived land cover map were assigned in the imperviousness class; all others were assigned as non-impervious. Although the overall classification accuracy of the derived land cover map is about 84.8%, the difference in spatial resolution (a single Landsat pixel contains app. 2500 Emerge pixels) justified the reference data in Site 19 as accurate enough for model training. The reference data from all the 19 sites contains 101,919 pixels in impervious class and 280,334 pixels in non-impervious class. These pixels were then randomly divided into calibration dataset and validation dataset with a ratio of 3:7.

3. Methodology

3.1. General approach of IIs incorporation

As illustrated in Fig. 1, the general procedure of incorporating IIs is based on spatial analysis of partial classification results derived from a priori classifier. The a priori classifier can be either a single classifier or a collection of classifiers. Furthermore, it is not restricted to a specific classification method as long as partial results are supported. An accuracy threshold is set up to determine the proportion of the obtained partial classification results from the entire dataset. Then IIs — additional inputs created on the partially classified image — complement traditional imagery information and assist the classification of the remaining data through a posterior classifier.

3.2. A case of implementation

In this paper, we tackle the problem of binary impervious surface classification. If any portion of a pixel occupies the constructed impervious class, that pixel is assigned to the impervious class. The motivation for a binary classification results from the imperviousness overestimation of large scale subpixel classifiers in rural areas (Homer et al., 2004; Yang et al., 2003), assigning imperviousness where no imperviousness exists. This proves to be quite significant for environmental and ecological applications where minor imperviousness changes in the low imperviousness range have considerable impacts (Forman & Alexander, 1998; Klein, 1979; Paul & Meyer, 2001). Therefore, our method does not directly compete with subpixel algorithms; instead, it complements them as someone could selectively perform subpixel analysis using our results as a preprocessing binary filter.

A hybrid multi-process classifier was used to implement the concept of intermediate inputs. The tested multi-process classifier includes a series of classifiers (five) acting collectively as the a priori classifier (see Fig. 1). Partial classification results were produced from the a priori classifier leading to specific IIs creation. An experiment was set up to incorporate and evaluate these IIs by testing an a posteriori classifier with and without IIs.

3.2.1. Hybrid multi-process classifier

The potential benefits of building an integration framework for collaborative operation of multiple algorithms in remote sensing applications have been recently presented (Mountrakis, 2008). In this paper, an expert-based system developed to support integration of multiple classifiers was applied for impervious surface detection (Mountrakis et al., 2009). The classification process is comprised of 5 classification steps (Fig. 2). In each step, parts of the dataset were classified while the rest were forwarded to subsequent steps. Using the 6 bands and their spatial distribution, numerous bands combinations, normalizations, and spatial filtering masks were produced as candidate inputs. For example, NDVI is one input which was calculated from the normalized difference between band 3 and band 4. For the first four classification steps, simple classifiers utilizing only two dimensions of the dataset were applied in each step targeting at the extraction of either impervious or non-impervious pixels. Each classifier selected two inputs with the best impervious/non-impervious separation from the candidates.

As low input dimensionality classifiers were successively added after the first four steps it became progressively more difficult to classify leftover pixels using only two inputs. To address the high complexity, a neural network structure was selected as the fifth and

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

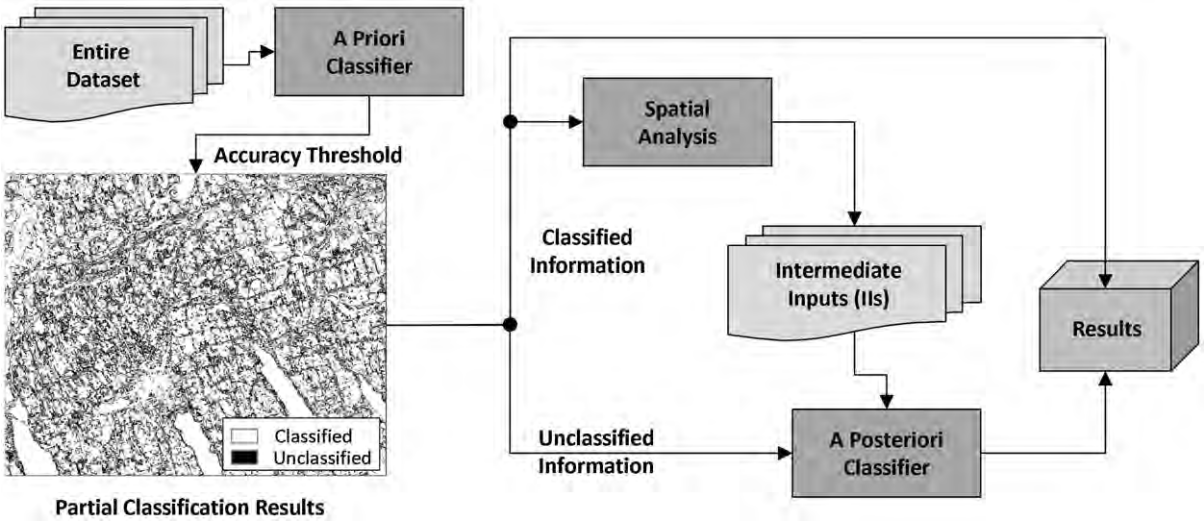

Fig. 1. The framework of IIs incorporation.

final step of the process. 4800 multilayer perceptron neural networks were trained using the backpropagation algorithm with 10 randomly selected inputs from more than 200 combinations of normalized differences and texture statistics between the 6 ETM+ bands. The node number for each hidden layer was randomly selected: between 8 and 16 for the first hidden layer, and from 0 to 6 for the second hidden layer. A higher range of node numbers in the input layer and hidden layers was not tested in order to avoid overfitting. The network output layer contained two nodes, each comprised of a logistic function, with one representing the impervious and the other the non-impervious class. The range of the two nodes was continuous between 0 and 1. Pixels were assigned to the corresponding class of larger valued node. Among the 4800 candidate architectures and inputs, the neural network with the best overall accuracy was selected. The random search employed was an efficient method to assess the values of intermediate inputs. The large number of candidate networks was used to ensure the validity of our statistical comparisons. In a typical implementation of our methodology, a lower number of simulations and inputs would be sufficient. Furthermore, processing time can be improved using advanced input selection methods, such as genetic algorithms. Since the purpose of this paper was to demonstrate the novelty and value of intermediate inputs rather than produce a full-scale implementation, processing time optimization is reserved for future work.

In the first four steps, all pixels extracted in each step were classified as either impervious or non-impervious class. Steps 1 to 3 identified pixels as non-impervious class and pixels extracted in step 4 were identified as impervious class. However, step 5 is the last step of the multi-process method in which all leftover pixels were classified. Therefore, step 5 identified both impervious and non-impervious pixels. For more details of the classifiers' structure in this hybrid multi-process model, please refer to Mountrakis et al. (2009). Table 1 shows the progressive classification of the validation dataset expressed as proportion of classified data in this step for the impervious (ISA) class and the non-impervious (NonISA) class.

3.2.2. Intermediate inputs

In the aforementioned hybrid classifier, we identified a process where partial results were created with all pixels falling into three categories: impervious, non-impervious, and unclassified, where unclassified represented leftover pixels that would be classified in subsequent step(s). That process is explained in details in Section 3.2.3. Using all classified

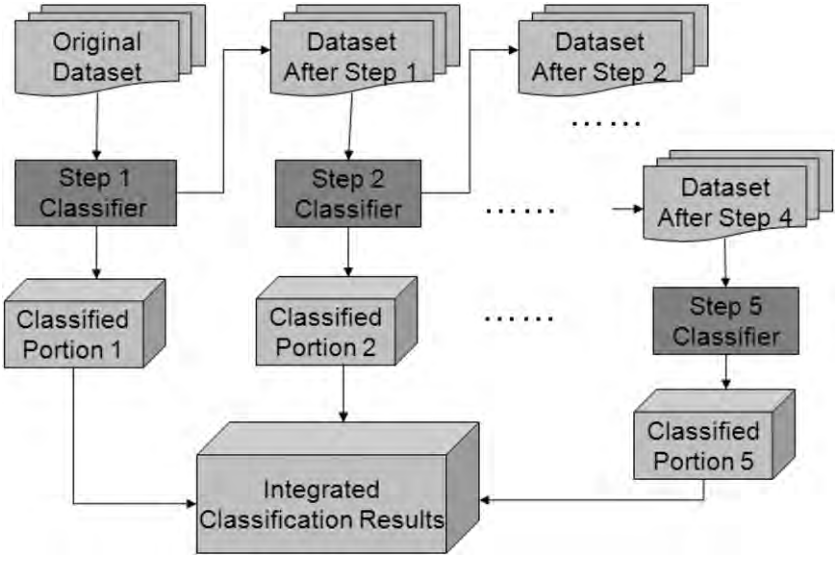

Fig. 2. The multi-process hybrid classification framework.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

Table 1
The extraction of validation dataset during the first 5 steps.

Steps	Absorption per class in reference data (%)		
	ISA	NonISA	Overall
Step1	0.17	7.52	3.23
Step2	1.41	14.14	6.72
Step3	2.12	23.58	11.07
Step4	50.71	0.58	29.82
Step5	45.59	54.18	49.16
Total	100	100	100

impervious and non-impervious pixels up to that step, statistics on the unclassified pixels were produced, creating "intermediate inputs" (IIs). The driving hypothesis is that these IIs can be involved in the classification of a subsequent step and potentially improve classification accuracy. 48 different IIs of three types were produced, namely texture-based (25), distance-based (11), and road-targeted (12) IIs.

3.2.2.1. Texture-based statistics. IIs were calculated by utilizing texture-based statistics. Mean and variance of classified pixels within a rectangle neighborhood centered on each unclassified pixel were obtained for five different spatial neighborhood sizes (3×3, 5×5, 7×7, 9×9, and 11×11 pixels) leading to 10 IIs. Additional texture statistics such as contrast, energy, and homogeneity, which were obtained from co-occurrence matrices of the classified pixels were also produced using the following equations (Chen et al., 1998):

$$Contrast = \sum_i \sum_j (i-j)^2 P_d(i,j) \quad (1)$$

$$Energy = \sum_i \sum_j P_d^2(i,j) \quad (2)$$

$$Homogeneity = \sum_i \sum_j \frac{P_d(i,j)}{1+|i-j|} \quad (3)$$

where P_d is the co-occurrence matrix in a rectangle neighborhood (3×3, 5×5, 7×7, 9×9, and 11×11) of the classified map, i is the row number and j is the column number in P_d. As a result, there were 15 additional IIs: 5 inputs for contrast, energy and homogeneity, respectively. All texture-based statistics were normalized and calculated using only classified pixels.

3.2.2.2. Distance-based statistics. Distances from each pixel to the nearest impervious pixel and the nearest non-impervious pixel were calculated. The ratio between the distance to the closest impervious pixel and the distance to the closest non-impervious pixel for each unclassified pixel was used as one intermediate input. Distances to impervious and to non-impervious of all pixels within a certain neighborhood size (3×3, 5×5, 7×7, 9×9, and 11×11) were also averaged to produce 10 additional IIs. In all distance calculations, the unclassified pixels were treated as pixels belonging to the opposite class.

3.2.2.3. Road-targeted statistics. In our early visual investigations, the algorithms missed isolated road pixels. To compensate for that, 12 IIs designed to detect roads are added. Based on road structure as elongated, line-shaped features, these inputs implemented rectangle line detection masks which are centered on each unclassified pixel of variable size (5×5, 7×7, and 9×9) in the horizontal, vertical and two diagonal directions. The unclassified pixels within each rectangle mask are not included during the filtering.

3.2.3. Strategy for incorporation of IIs

The primary objective of the research in this paper is to investigate whether the concept of IIs is beneficial. In addition, we are interested in identifying the portion of the partially classified image that optimizes their potential benefits. It is a balancing act of two opposing forces: as the portion of the classified image used for IIs production was increased, the classification accuracy of the already classified image portion decreases.

To study this balance, a gradually increasing contribution was used for the neural network developed in step 5 of the hybrid multi-process classification method introduced in Section 3.2.1, while an additional step (step 6) was added to assess IIs. As a result, classifiers in steps 1 to 5 were used as the a priori classifiers to produce partial classification results. Based on these partial classification results, IIs were derived and utilized to assist classification in the a posteriori classifier in step 6. Accuracy thresholds were applied on step 5, starting from 89% to 99% with an increment of 1%. Using the calibration dataset, these accuracy thresholds were translated into node thresholds for the outputs of the neural network comprising step 5. These thresholds were slightly different for each of the two classes (Fig. 3). The "winner takes all" rule was applied in cases when a pixel satisfied both thresholds; the pixel was assigned to the node with the strongest response. For example, in the case of a 94% accuracy threshold, all pixels with impervious node response larger than 0.76 were selected and assigned to the imperviousness class (Fig. 3). Pixels with response higher than 0.74 at the non-imperviousness node were also extracted and classified as non-impervious. Pixels satisfying both thresholds were assigned to the one with the strongest response. Fig. 3 shows the relationship between different accuracy thresholds for step 5 and the corresponding thresholds for the neural network outputs (logistic nodes) of step 5.

The use of variable accuracy thresholds for step 5 resulted in variable proportions of classified data, where higher accuracies led to lower proportions of classified data for the current step and therefore a larger leftover dataset. As a result of the 11 accuracy thresholds (89% to 99% with 1% interval), 11 datasets were created expressing the leftover pixels after step 5. Fig. 4 shows the percentage of the remaining dataset after step 5 that is carried over to step 6, depending on the accuracy threshold of step 5. It is important to emphasize the tradeoff when the accuracy threshold decreases. On one hand, when the accuracy threshold in step 5 decreases, a higher number of pixels are classified within that step, but on the other hand, step 6 becomes more challenging, since step 6 will include pixels with lower discriminatory power.

In Fig. 5, the experiment setup for testing IIs was summarized. As was mentioned, there were 11 datasets derived from 89% to 99% variable accuracy threshold for step 5. For each of the 11 datasets the two best neural networks were identified in terms of overall accuracy,

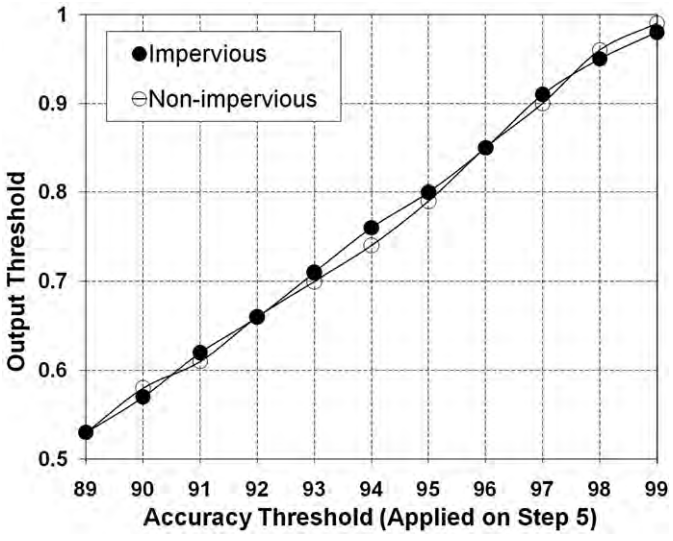

Fig. 3. Network output thresholds for variable accuracy thresholds in step 5.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

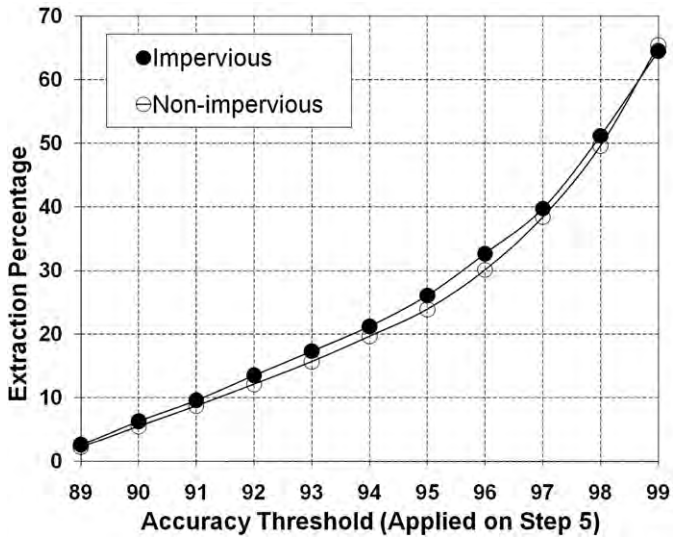

Fig. 4. Remaining dataset portion after step 5 application for different accuracy levels.

one without and the other with IIs, creating a total of 22 neural networks.

The evaluation of the IIs performance was done using step 6 as a guide. First, 11 networks (mirroring 89% to 99% accuracy thresholds) were trained with 10 randomly selected inputs, without inclusion of IIs. After the 11 best networks were identified along with their respective 10 inputs, 11 new neural networks were trained using the same identified 10 inputs plus IIs. The idea was to isolate any improvements by IIs. Because we could not determine a priori the influence of the 48 possible intermediate input candidates, a random search method was adopted in the networks to select 5 of them as additional inputs. Therefore, each of these networks was trained with 15 inputs in total: the fixed 10 from its corresponding best network without IIs, and the randomly selected 5 IIs.

After the IIs concept was implemented, the process using exclusively step 6 was evaluated to assess direct improvements of the IIs. Then results from steps 5 and 6 were combined in order to study the cumulative effect of IIs and whether an optimal application threshold did exist. The spatial footprint and classification results on the central New York Landsat scene were also produced.

4. Results

4.1. IIs and neural network selection

11 different neural networks using IIs as possible candidates for step 6 were identified. Table 2 presents the five IIs selected to build each of the 11 optimal neural networks along with the number of nodes in their hidden layers [layer 1, layer 2]. The accuracy thresholds applied on step 5 are also listed. Looking further into the selected IIs, the road statistics have a strong presence. Contrast and homogeneity statistics along with distances to features also play an important role. Of special interest is the dependency of IIs selection and pixel size considering future implementations on imagery of different spatial resolution.

As Table 2 indicates there were cases where the same II was selected twice for the same accuracy threshold (e.g. accuracy thresholds 92% and 97%). This duplication suggests that using four IIs produced higher accuracy than five in our tests. The neural network in that case would ignore the duplicate input, actively using only four inputs. Furthermore, there is a reasonable expectation that the same selected IIs would not appear in two adjacent accuracy thresholds because various accuracy thresholds would result in significantly different corresponding training datasets and the higher accuracy dataset would contain a substantial higher number of training pixels (see Fig. 4).

4.2. Direct assessment of IIs benefits using step 6

Results obtained exclusively from step 6 with and without IIs were compared by calculating overall accuracies. The purpose of this

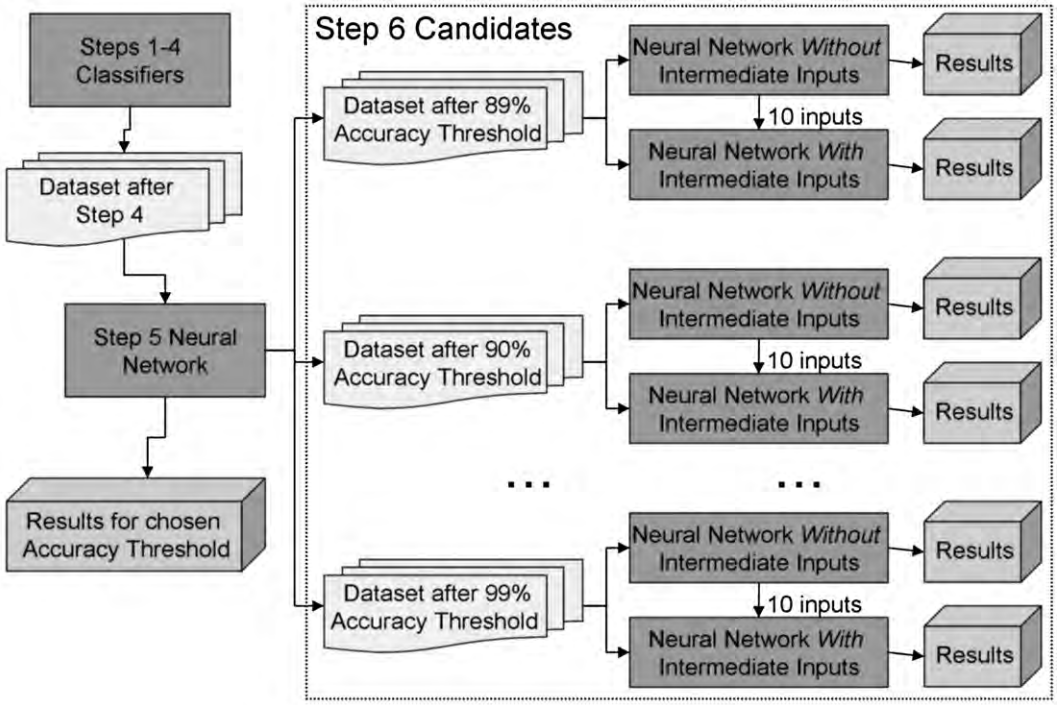

Fig. 5. The framework for testing IIs.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

Table 2
IIs and network structure.

Step 5 accuracy threshold	Selected IIs	Hidden layer number of nodes
89	Contrast (9×9) Contrast (5×5) Road feature (diagonal direction, 9×9) Average value of distance to impervious (5×5) Average of classified points (7×7)	[8, 5]
90	Average value of distance to impervious (5×5) Variance of classified points (5×5) Road feature (vertical direction, 9×9) Average value of distance to non-impervious (9×9) Average value of distance to non-impervious (7×7)	[12, 0]
91	Homogeneity (9×9) Homogeneity (5×5) Road feature (diagonal direction, 9×9) Road feature (horizontal direction, 7×7) Average value of distance to impervious (5×5)	[10, 0]
92	Average value of distance to impervious (3×3) Road feature (diagonal direction, 9×9) Average value of distance to impervious (3×3) Energy (9×9) Contrast(7×7)	[11, 0]
93	Average value of distance to non-impervious (11×11) Average of classified points (9×9) Average value of distance to impervious (5×5) Average value of distance to non-impervious (5×5) Road feature (vertical direction, 7×7)	[9, 0]
94	Average value of distance to impervious (5×5) Average value of distance to non-impervious (5×5) Average value of distance to non-impervious (9×9) Road feature (diagonal direction, 5×5) Road feature (horizontal direction, 5×5)	[9, 0]
95	Variance of classified points (3×3) Average value of distance to non-impervious (9×9) Average of classified points (5×5) Average value of distance to impervious (7×7) Homogeneity (3×3)	[12, 2]
96	Average value of distance to non-impervious (9×9) Average value of distance to impervious (3×3) Variance of classified points (9×9) Road feature (vertical direction, 5×5) Contrast (5×5)	[13, 1]
97	Contrast (5×5) Average value of distance to impervious (7×7) Contrast (7×7) Contrast (7×7) Average of classified points (11×11)	[11, 2]
98	Variance of classified points (5×5) Average of classified points (5×5) Road feature (horizontal direction, 5×5) Average value of distance to impervious (11×11) Average of classified points (11×11)	[11, 0]
99	Road feature (vertical direction, 7×7) Average value of distance to non-impervious (11×11) Homogeneity (11×11) Road feature (horizontal direction, 9×9) Contrast (11×11)	[11, 3]

comparison was to assess the benefits of IIs as classification difficulty increased, by lowering the accuracy threshold. All tests were performed by simulating step 5 on the remaining validation dataset from steps 1 through 4 and using different accuracy thresholds on the neural network outputs from step 5.

The results are shown in Fig. 6 and Table 3. Two observations are of special interest:

i) Classification accuracies for both algorithmic setups with and without IIs decrease, as accuracy thresholds for step 5 decrease, since classification complexity increases.
ii) Algorithms with IIs provide a substantial and consistent benefit. The most encouraging result is that this benefit increases as the classification difficulty increases: we see an average improvement of 3.6% in the overall accuracy metric.

To further assess the contribution of the IIs, a *Z*-test between Kappa statistics with and without IIs was conducted. The null hypothesis is: Kappa without IIs = Kappa with IIs. The *Z*-score between Kappas was calculated as (Cohen, 1960):

$$Z\text{−score} = (K_1 - K_2) / \sqrt{(ASE(K_1))^2 + (ASE(K_2))^2} \quad (4)$$

where K_1, K_2 represent Kappa values and $ASE(K_1)$, $ASE(K_2)$ represent Asymptotic Standard Errors (*ASE*s) of the two models. A *p*-value was calculated using the *Z*-score expressing the confidence level that these two models can produce identical results. The confidence level was set to the high value of 0.01. The results are shown in Table 4 and the null hypothesis can be rejected for all accuracy thresholds except for 99%, which is still at a very low level (0.0173).

4.3. Overall assessment of IIs benefits combining step 5 and step 6

The previous assessment evaluated the incorporation of IIs as complexity increased. However, it did not take into account the variability in total pixels contained in each of the 11 datasets. In order to assess overall benefits, performance from steps 5 and 6 were combined. The same first four steps were kept for all models and three algorithmic approaches were compared:

i) Model 5: The optimal neural network for step 5 assuming there was no step 6 to follow (i.e. all pixels should be classified by step 5);
ii) Model 6R (Regular): 11 combinations of step 5 and 6, where step 6 did not use IIs;
iii) Model 6II (with IIs): 11 combinations of step 5 and 6, where step 6 incorporated IIs.

The overall accuracies for the three models were calculated (Fig. 7 and Table 5). These values are directly comparable to each other, since they were produced by the use of the same validation dataset. Accuracy of Model 5 is constant at 87.74%, since no accuracy threshold was applied. Model 6R expresses potential benefits associated with adding another neural network as a post-process while also increasing the input dimensionality. Model 6II incorporates the additional neural network and increased dimensionality but also supports IIs. It should be noted that Model 6II contains the exact same

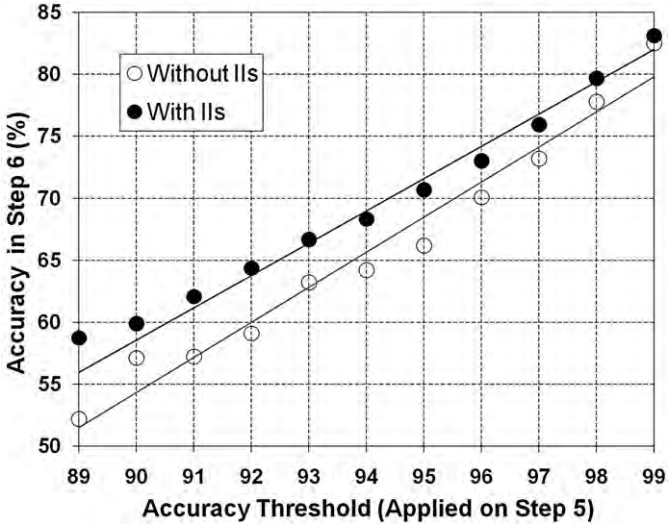

Fig. 6. Performance of step 6 for various accuracy thresholds on step 5.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

Table 3
Overall accuracy at each accuracy threshold for step 6.

Accuracy threshold for step 5		89	90	91	92	93	94	95	96	97	98	99
Overall (%)	Without II	52.18	57.12	57.23	59.10	63.21	64.21	66.17	70.07	73.21	77.78	82.52
	With II	58.74	59.89	62.06	64.36	66.66	68.32	70.69	73.02	75.93	79.68	83.12
	Improvement	6.56	2.77	4.83	5.26	3.45	4.11	4.52	2.95	2.72	1.90	0.60

Table 4
Z-test for step 6 between with and without II models.

Accuracy threshold for step 5	89	90	91	92	93	94	95	96	97	98	99
Z-score	3.386	2.457	5.133	6.716	5.000	6.719	8.299	6.293	6.424	5.612	2.112
p-value	0.0004	0.0070	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0173

inputs selected for each corresponding network in Model 6R plus 5 randomly selected IIs. We did not examine the case of randomly selecting not only IIs but also regular inputs to allow direct model comparisons, even though it could potentially increase further the II benefits.

The results clearly showed an advantage for the IIs as expressed in Model 6II. The overall accuracy has improved over 1.3% from Model 5 and 0.9% over Model 6R. These improvements may not seem significantly in terms of absolute numbers, but keeping in mind the classification complexity, we consider these to be substantial improvements. These improvements are also image-dependent, for instance, the presence of many bare soil areas, which typically cause spectral confusion with impervious surfaces, can affect the results.

Most importantly, from our hypothesis perspective, it can also be inferred that there is an optimal point where the benefits of Model 6II are maximized. The shape of the overall accuracies suggests that this optimal point exists and in these experiments it is the 96% accuracy threshold. That threshold expresses the balance between including a sufficient amount and accurate enough pixels to produce the IIs. The 96% threshold is image-dependent, but it can easily be identified in other classification applications as this process is automated.

4.4. Final model selection and classification

Benefits of IIs were maximized at an accuracy threshold of 96%. Detailed accuracy metrics of this threshold were further developed, using accuracy assessments (Congalton, 1991; Story & Congalton, 1986) for all three models combining the effects of steps 5 and 6. As illustrated in Table 6, Model 6R and Model 6II produced higher accuracies compared to Model 5. Therefore, introducing an additional step improves accuracy metrics. By using IIs, the producer's accuracy of NonISA and user's accuracy of ISA were almost identical while the improvements were concentrated on producer's accuracy of ISA and user's accuracy of NonISA. This translates into a reduction of ISA pixels mistakenly classified as NonISA.

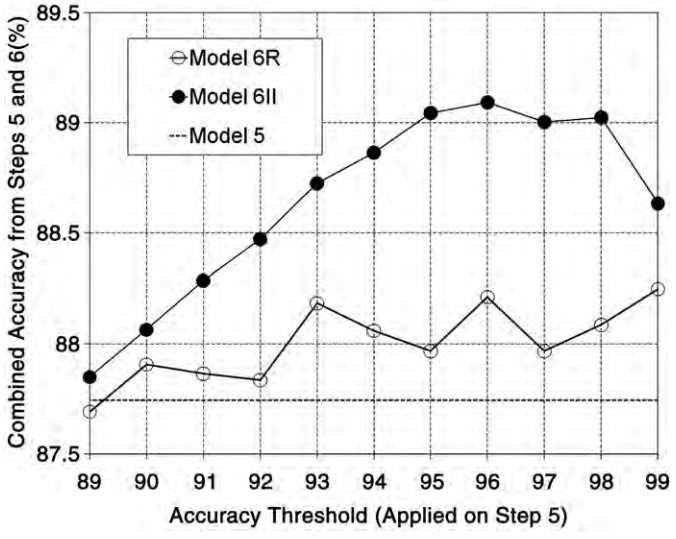

Fig. 7. Combined performance of steps 5 and 6 for various accuracy thresholds on step 5.

The classification results of the entire study area using Model 6II at 96% threshold are represented in Fig 8. The simulated region includes the reference data used for training and validation of the algorithm. In order to evaluate further, classification results for the three models are visualized within a selected representative area (red rectangle in Fig. 8) and graphically shown in Fig. 9. Fig. 9a denotes the classification results using step 5 as the final step (Model 5). Fig. 9b and Fig. 9c illustrate the classification results by using an additional step 6 without IIs (Model 6R) and with IIs (Model 6II), respectively. Fig. 9d is also included to demonstrate the spatial footprint of the steps under consideration. In the case of Model 5, both blue and red pixels in Fig. 9d are classified using step 5.

Focusing on the circles in Fig. 9 revealed that numerous soil pixels were misclassified as impervious surfaces by using Model 5 (Fig. 9a) and Model 6R (Fig. 9b). Model 6II with IIs was successful at limiting such misclassifications (Fig. 9c), especially the soil pixels close to roads. This is a significant contribution of the method, since current algorithms have a difficulty differentiating soil and impervious classes due to their lack of spectral separability. It shows the ability of IIs to add spatial context in the classification process. While there is a significant improvement on successful soil identification, a limited number of impervious pixels, namely road pixels were missed. This was mostly driven by: i) the limited amount of ISA pixels corresponding to roads; roads spatial footprint was small compared to other ISA features (e.g. houses and parking lots), and ii) the algorithmic performance metric used to select the winning neural network was the overall classification accuracy. It is a tradeoff that we are willing to accept, considering the overall classification improvement. Furthermore, missing road pixels could be identified with a post-process using morphological operators or using external datasets, but it is beyond the scope of this paper.

5. Discussion and conclusions

The concept of intermediate inputs could be seen as a context classification; classification methods that take into account the labeling of neighbors when seeking to determine the most appropriate class for a pixel (Richards & Jia, 1999) with several successful implementations (for example Benediktsson et al., 2005; Binaghi et al., 2003; Melgani & Serpico, 2003; Tilton et al., 1982). Similar to context classification, introducing intermediate inputs in the remote sensing image classification process takes advantage of spectral/

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

Table 5
Combined overall accuracy at each specified accuracy level for steps 5 and 6.

Accuracy threshold for step 5		89	90	91	92	93	94	95	96	97	98	99
Overall (%)	Model 5	87.74 (threshold independent)										
	Model 6R	87.69	87.90	87.86	87.83	88.18	88.06	87.97	88.21	87.97	88.08	88.25
	Model 6II	87.85	88.06	88.29	88.47	88.73	88.86	89.05	89.09	89.00	89.02	88.63

spatial context information from neighboring pixels. It exploits the relationship between adjacent pixels. However, there is a significant difference from typical context classifiers. The intermediate inputs concept uses algorithms to classify a scene to a predetermined accuracy level, which acts as the basis to explore spatial relationships. In other words, it is a posteriori process, since we already have an accurate and partially classified scene. In contrast, typical context classifiers, through either preprocessing or postprocessing procedures, act on the entire image without incorporating detailed partial classification results.

The statistical evaluation suggested that using intermediate inputs improved the classification compared to traditional classification models. The overall accuracy showed an average improvement of 3.6% (maximum 6.6%). Furthermore, the *Z*-test between Kappa statistics with and without intermediate inputs illustrated that the improvements were statistically significant. These improvements tended to increase as classification difficulty increased, a significant advantage of the proposed method. It should be noted that the aforementioned statistical improvements are site dependent. Our methodology takes over where spectral separability between classes is limited, for example, between types of soil and impervious surfaces. Intermediate input improvements are highly dependent on the proportion of these inseparable types in the overall image. Considering that this is typical challenge among numerous existing classifiers, intermediate inputs have the potential for significant advances. Furthermore, the use of intermediate inputs was not explored to its full potential. Our statistical comparisons constrained all intermediate input-based models to the usage of the exact same non-intermediate inputs in order to attribute any improvements specifically to intermediate inputs. If numerous non-intermediate inputs are also examined along with candidate intermediate inputs accuracy could improve further.

The incorporation of intermediate inputs requires a two-step classification process: the a priori classifier to provide a partially classified image; and the a posteriori classifier to integrate image-based inputs with intermediate inputs. In this paper the a priori classifier was composed of five steps in order to achieve even higher accuracy. Scientists implementing the intermediate inputs concept could avoid such complexity by selecting a simple and efficient a priori classifier depending on their accuracy requirements. Instead the optimization process could include an assessment of the obtained results and misclassifications in order to create targeted intermediate inputs: for example in our experiment, spatial masks were created as intermediate inputs in an attempt to connect broken road pixels.

From the computational load perspective, the incorporation of intermediate inputs brings in additional computing circles. We demonstrated that an optimal threshold did exist in a conclusive statistical manner using an extensive number of neural network combinations. In future implementations, lower number of neural network combinations need to be tested now that the proof of concept exists, while adaptive methodologies such as genetic algorithms can be implemented to test even fewer combinations. Since the classification process for each accuracy threshold was independent of each other, parallel computing techniques could improve simulation times.

Another advantage of the intermediate inputs concept is its independence of algorithmic methodology. In the experiments in this paper, neural networks were used as the underlying methodology, however a wide range of methods is supported. The only constraint applies to the a priori classifier; it should have the ability to create partial results for a given accuracy threshold. There are no constraints for the a posteriori classifier. This flexibility allows incorporation of technique in numerous applications. It could potentially be generalized to other ground feature classifications such as forest, rivers, etc and even applications outside the remote sensing domain.

In this paper, intermediate inputs were applied on coarse resolution imagery (~30 m Landsat ETM+ imagery). The concept of intermediate inputs can also be implemented to higher resolution imagery, where shape and texture intermediate inputs are expected to become more prominent in the classification process, since spatial dependencies are more pronounced in higher resolution imagery. This work proved the concept for intermediate inputs through an impervious classification illustration. Future implementations will be

Table 6
Detailed accuracy comparison of the three models for step 5 and 6.

		Model 5	Model 6R	Model 6II
Producer's	ISA	89.94	90.42	92.41
	NonISA	85.16	85.61	85.19
User's	ISA	87.70	88.09	88.01
	NonISA	87.80	88.36	90.51
Overall		87.74	88.21	89.09

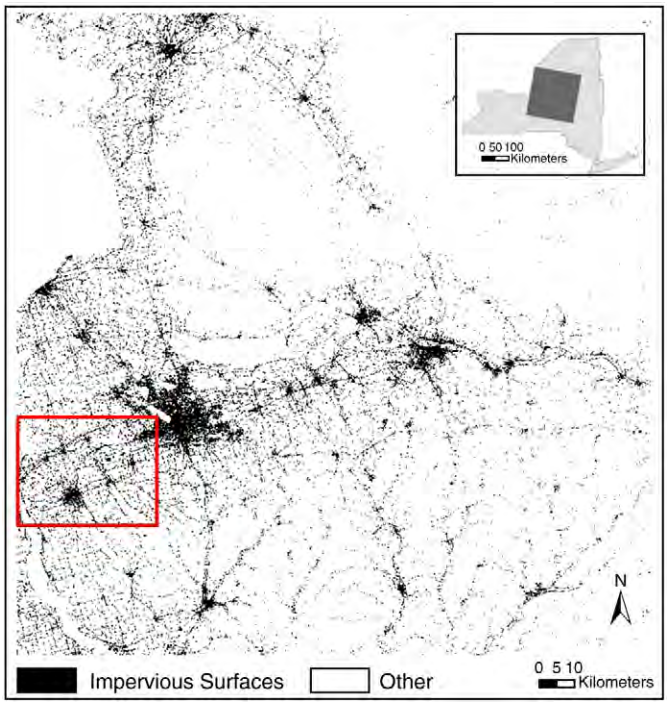

Fig. 8. Classification map using 6II model at 96% threshold; Red rectangle denotes a selected representative area analyzed further in Fig. 9. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

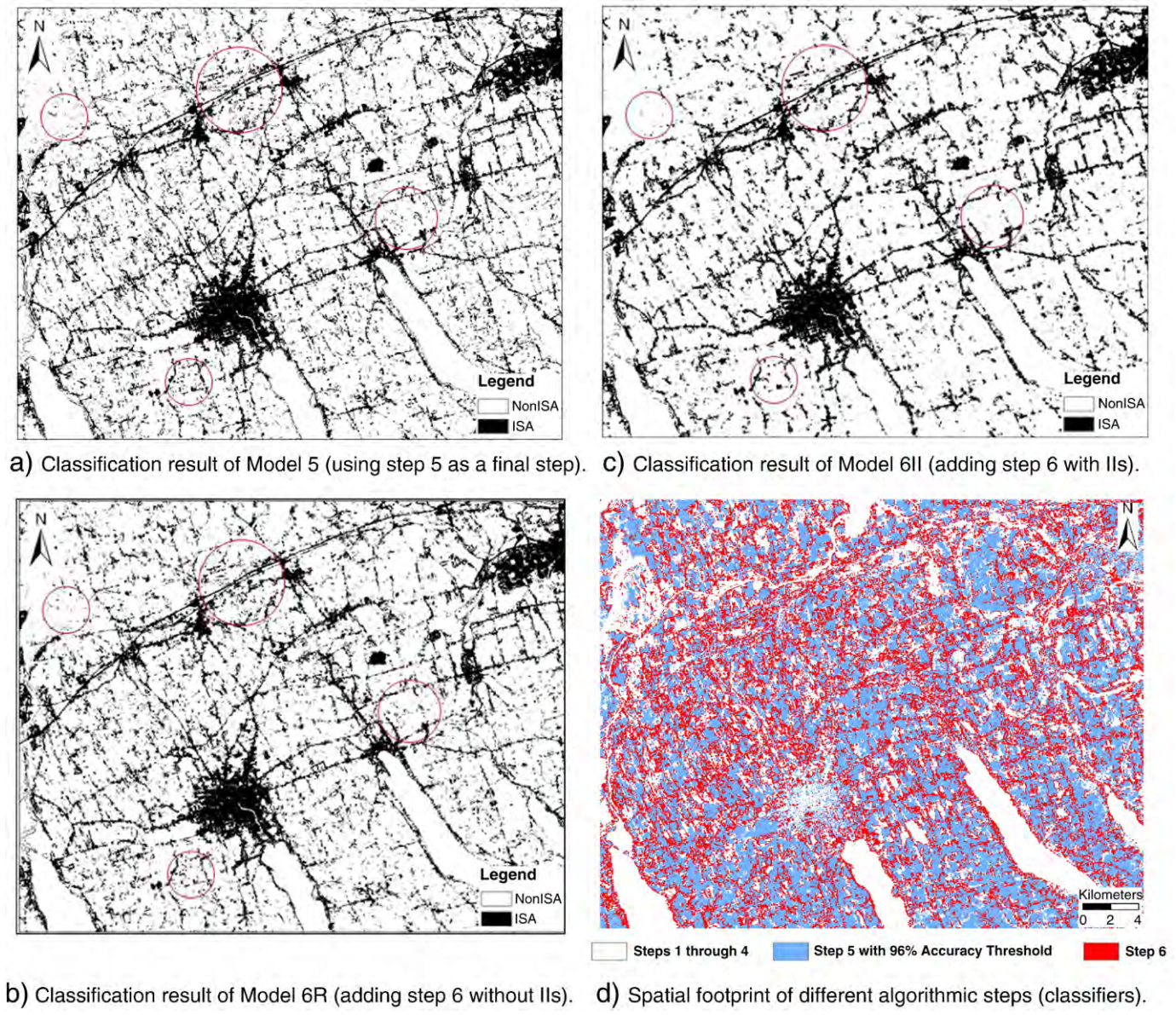

Fig. 9. Visual comparison of different classification models on a 39×31 km subarea.

extended to numerous spatial resolutions, classification tasks and classification methodologies.

Acknowledgements

This research was partially supported by the National Science Foundation (award GRS-0648393), by the National Aeronautics and Space Administration (awards NNX08AR11G, NNX09AK16G) and by the Syracuse Center of Excellence CARTI Program. We would like to thank Dr. Dave Nowak and Dr. Jeffrey Walton from the USDA Forest Service for providing training data.

References

Arnold, C. L., & Gibbons, C. J. (1996). Impervious surface: The emergence of a key urban environmental indicator. *Journal of the American Planning Association*, *62*(2), 243—258.

Bauer, M., Burk, T. E., EK, A. R., Coppin, P. R., Lime, S. D., Walsh, T. A., et al. (1994). Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering & Remote Sensing*, *60*(3), 287—298.

Bauer, M., Loeffelholz, B., & Wilson, B. (2005). Estimation, mapping and change analysis of impervious surface area by Landsat remote sensing. *Proceedings of Pecora 16 conference, ASPRS annual conference, Oct. 23–27, Sioux Falls, SD.*

Benediktsson, J. A., Palmason, J. A., & Sveinsson, J. R. (2005). Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), 480—491.

Binaghi, E., Gallo, I., & Pepe, M. (2003). A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(12), 2906—2922.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123—140.

Chen, C. H., Pau, L. F., & Wang, P. S. P. (1998). *The handbook of pattern recognition and computer vision* (pp. 207—248). : World Scientific Publishing Company.

Coe, S. E., Hepinstall, J. A., & Coburn, R. (2005). A hybrid approach to detecting impervious surface at multiple-scales. *Proceedings of ISPRS joint conference,* 3rd international symposium remote sensing and data fusion over urban areas and international symposium remote sensing of urban areas, Mar. 14–16, Tempe, AZ.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37—46.

Congalton, R. G. (1991). A review of assessing the accuracy of classification of remotely sensed data. *Remote sensing of Environment*, *37*, 35—46.

Crane, M., Xian, G., & McMahon, C. (2005). Estimation of sub-pixel impervious surface using Landsat and Aster imagery for assessing urban growth. *Proceedings of ISPRS joint conference,* 3rd international symposium remote sensing and data fusion over urban areas and international symposium remote sensing of urban areas, Mar. 14–16, Tempe, AZ.

Dougherty, M., Dymond, R. L., Goetz, S. J., Jantz, C. A., & Goulet, N. (2004). Evaluation of impervious surface estimates in a rapidly urbanizing watershed. *Photogrammetric Engineering & Remote Sensing*, *70*(11), 1275—1284.

Forman, R. T. T., & Alexander, L. E. (1998). Roads and their major ecological effects. *Annual Review of Ecology and Systematics*, *29*, 207—231.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008

Forster, B. C. (1980). Urban residential ground cover using Landsat digital data. *Photogrammetric Engineering & Remote Sensing*, *46*(4), 547−558.
Franke, J., Roberts, D. A., Halligan, K., & Menz, G. (2009). Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments. *Remote Sensing of Environment*, *113*, 1712−1723.
Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *12*, 993−1001.
Herold, N. (2003). Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. *Proceedings of 2003 ASPRS annual convention, May 5–9, Anchorage, Alaska*.
Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering & Remote Sensing*, *70*(7), 829−836.
Hu, X., & Weng, Q. (2009). Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sensing of Environment*, *113*, 2089−2102.
Iyer, S. V., & Mohan, B. K. (2002). Urban landuse monitoring using neural network classification. *Proceedings of 2002 international geoscience and remote sensing symposium, June 24–28, Toronto, Canada* (pp. 3124−3126).
Klein, R. (1979). Urbanization and stream quality impairment. *Water Resources Bulletin*, *15*(4), 948−963.
Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation and active learning. *Advances in Neural Information Processing Systems*, *7*, 231−238.
Lee, S., & Lathrop, R. G. (2005). Sub-pixel estimation of urban land cover components with linear mixture model analysis and Landsat Thematic Mapper imagery. *International Journal of Remote Sensing*, *26*(22), 4885−4905.
Lee, S., & Lathrop, R. G. (2006). Subpixel analysis of Landsat ETM+ using self-organizing map (SOM) neural networks for urban land cover characterization. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1642−1654.
Liu, W., Gopal, S., & Woodcock, C. E. (2004). Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogrammetric Engineering & Remote Sensing*, *70*(8), 963−971.
Lu, D., & Weng, Q. (2006). Spectral mixture analysis of ASTER images for examining the relationship between urban thermal features and biophysical descriptors in Indianapolis, Indiana, USA. *Remote Sensing of Environment*, *104*(2), 157−167.
Melgani, F., & Serpico, S. B. (2003). A Markov random field approach to spatio-temporal contextual image classification. *IEEE transactions on Geoscience and Remote Sensing*, *41*(11), 2478−2487.
Mountrakis, G. (2008). Next generation classifiers: Focusing on integration frameworks. *Photogrammetric Engineering & Remote Sensing*, *74*(10), 1178−1180.
Mountrakis, G., Watts, R., Luo, L., & Wang, J. (2009). Developing collaborative classifiers using an expert-based model. *Photogrammetric Engineering & Remote Sensing*, *75*(7), 831−843.
Myeong, S., Nowak, D. J., Hopkins, P. F., & Brock, R. H. (2001). Urban cover mapping using digital, high-spatial resolution aerial imagery. *Urban Ecosystems*, *5*, 243−256.
Paul, M. J., & Meyer, J. L. (2001). Streams in the urban landscape. *Annual Review of Ecology and Systematics*, *32*, 333−365.
Phinn, S., Stanford, M., Scarth, P., Murray, A. T., & Shyy, T. (2002). Monitoring the composition and form of urban environments based on the vegetation-impervious surface-soil (VIS) model by sub-pixel analysis techniques. *International Journal of Remote Sensing*, *23*(20), 4131−4153.
Powell, R. L., Roberts, D. A., Dennison, P. E., & Hess, L. L. (2007). Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, *106*, 253−267.
Richards, J. A., & Jia, X. (1999). *Remote sensing digital image analysis.* Springer.
Schueler, T. R. (1994). The importance of imperviousness. *Watershed Protection Techniques*, *1*, 100−111.
Steele, B. M. (2000). Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, *74*, 545−556.
Story, M., & Congalton, R. G. (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering & Remote Sensing*, *52*(3), 397−399.
Tilton, J. C., Vardeman, S. B., & Swain, P. H. (1982). Estimation of context for statistical classification of multispectral image data. *IEEE transactions on Geoscience and Remote Sensing*, *20*(4), 445−452.
Weng, Q. (2007). *Remote sensing of impervious surfaces.* Boca Raton, Florida: CRC Press.
Wu, C., & Murray, A. T. (2003). Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment*, *84*, 493−505.
Yang, L., Huang, C., Homer, C. G., Wylie, B. K., & Coan, M. J. (2003). An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing*, *29*(2), 230−240.

Please cite this article as: Luo, L., & Mountrakis, G., Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example, *Remote Sensing of Environment* (2010), doi:10.1016/j.rse.2010.01.008