Article

Development of an Object-Based Interpretive System Based on Weighted Scoring Method in a Multi-Scale Manner

Abbas Kiani [1], Hamid Ebadi [1] and Farshid Farnood Ahmadi [2],*

[1]   Geomatics Engineering Faculty, K.N. Toosi University of Technology, Tehran 1996715433, Iran;
     abbasekiani@yahoo.com (A.K.); Ebadi@kntu.ac.ir (H.E.)
[2]   Department of Geomatics Engineering, University of Tabriz, Tabriz 5166616471, Iran
*   Correspondence: farnood.tabrizu@iran.ir; Tel.: +98-413-339-2534

check for
updates

Abstract: For an accurate interpretation of high-resolution images, correct training samples are required, whose automatic production is an important step. However, the proper way to use them and the reduction of their defects should also be taken into consideration. To this end, in this study, the application of different combinations of training data in a layered structure provided different scores for each observation. For each observation (segment) in a layer, the scores corresponded to the obtained misclassification cost for all classes. Next, these scores were properly weighted by considering the stability of different layers, the adjacency analysis of each segment in a multi-scale manner and the main properties of the basic classes. Afterwards, by integrating the scores of all classes weighted in all layers, the final scores were produced. Finally, the labels were achieved in the form of collective wisdom, obtained from the weighted scores of all segments. In the present study, the aim was to develop a hybrid intelligent system that can exploit both expert knowledge and machine learning algorithms to improve the accuracy and efficiency of the object-based classification. To evaluate the efficiency of the proposed method, the results of this research were assessed and compared with those of other methods in the semi-urban domain. The experimental results indicated the reliability and efficiency of the proposed method.

Keywords: object-based analysis; remote-sensing classification; fusion decision; support vector machines (SVMs); high spatial-resolution images

1. Introduction

With the development of digital sensors, an increasing number of high spatial resolution (HSR) remote-sensing images have become available [1]. The availability and accessibility of vast amounts of high-resolution data have posed a challenge for remote-sensing image classification. As a result, object-based image analysis (OBIA) techniques have emerged to address these issues [2]. These techniques have now replaced the traditional pixel-based method as the new standard method [3] that will facilitate land-cover classification from HSR remote-sensing imagery.

Supervised classifications are faced with challenges related to, among others, the imbalance between high dimensions and the existing training samples or the presence of mixed samples in the data (noise in training samples occurring in HSR images) [4]. In terms of object-based image classification, the sampling method is undoubtedly a crucial step. Furthermore, it is common for some objects to be mixed in their class composition and thus violate the commonly made assumption of object purity that is implicit in a conventional object-based image analysis. Mixed objects can introduce a problem throughout the classification analysis, but are particularly challenging in the training stage

as they can result in degraded training statistics and reduce mapping accuracy [5]. In addition to segmentation and sampling, all features, classification, and accuracy evaluation can bring uncertainty for OBIA. In this regard, despite the fact that training samples or remote-sensing images may vary slightly, the accuracy assessment results of different commonly used classifiers are consistent with the expected conclusions from previous researches [6]. Moreover, compared to other factors, classifiers constitute a very important influential factor for supervised classification [2].

Supervised parametric classifiers such as the Maximum Likelihood Classification (MLC) deliver excellent results when dealing with unimodal data. However, they have limitations when dealing with multi-modal input datasets because these classifiers assume a normal data distribution. Non-parametric supervised classifiers such as Support Vector Machines (SVM) and Artificial Neural Network (ANN) classifiers do not make any assumptions regarding frequency distribution and have therefore become increasingly popular for classifying remotely sensed data, which rarely have normal distributions [7]. There are several ANN approaches that can be used to classify remotely sensed images which include: Multi-layer perceptron (MLP); Self-Organized Feature Map (SOM); and Fuzzy ArtMap. All these algorithms depend mainly on the operators experience in setting up their parameters in order to reach an optimal performance. SOM produced the lowest classification accuracy in the majority of articles. MLP requires a complete retraining of the whole network. This may lead to a long training time, even for small size test areas. The defect with Fuzzy ArtMap is that it is sensitive to noise and outliers that may decrease the classification accuracy. Unlike MLP and Fuzzy ArtMap, SOM allows for the discrimination of multimodal classes. On the other hand, SOM normally yields many unclassified pixels [8]. Huang et al. [9] suggest that SVMs typically outperform ANNs and MLC in terms of overall accuracy, even with smaller training sample sizes. Pal and Mather [10] used just such an argument when they cited the complexity of ANNs as a reason to use SVMs instead. Pal [11] and Adam et al. [12] found that SVM and Random Forest (RF) performed equally well in terms of accuracy, whereas Zhang and Xie [13] and Maxwell et al. [14–16] found that SVM outperformed RF.

The results of recent studies indicate that RF and SVM are suitable object-based classifications and confirm the expected general tendency, namely that the overall accuracies decline when increasing the segmentation scale [7,17]. Li et al. [6] stated that the RF and decision tree classifiers are the most robust with or without feature selection. The results of training sample analyses indicated that the RF and Adaptive Boosting (AdaBoost) processes offer a superior generalization capability, except when dealing with small training sample sizes. Ma et al. [2] mentioned that RF classification exhibited the best performance in object-based classification, and has thus, attracted considerable attention in recent years; moreover, RF is followed by SVM, with MLC performing the worst. Furthermore, the Nearest Neighbor (NN) appears unsuitable for more extensive use in object-based classification. Thus, the use of NN should be reduced, even though it used to be the most frequently employed classifier for object-based classification. Since around 2010, the AdaBoost technique, as another ensemble classifier, started to gain momentum in remote sensing classification, due to its high accuracy [18]. Chan and Paelinckx [18] also evaluated RF and AdaBoost tree-based ensemble classifications using airborne hyperspectral imagery and yielded almost the same accuracy results as those of established per-pixel classifiers. AdaBoost can process data with weights, and the mis-classification rate for each trial is used to update the distribution over the training samples. Finally, the voting for the labels is weighted by the accuracy of each classifier. In general, RF and SVMs outperform other classifiers in terms of classification accuracy [19,20].

Some sampling techniques have been introduced including heuristic or non-heuristic oversampling, under-sampling, and data cleaning rules, such as removing "noise" and "borderline" examples [2]. These works focus on data-level techniques. Other researchers concentrate on changing the classifier internally, to deal with class imbalance [21]. Liu and Huang [22] proposed an ensemble SVMs learning scheme based on AdaBoost to overcome the shortcomings of single SVMs. First, SVMs kernel function parameters and penalty parameters were optimized by the particle swarm optimization (PSO) algorithm. Then, the principal component analysis (PCA) method was adopted to eliminate noise

features for remote sensing image. At the same time, a fuzzy c-means (FCM) clustering technology was introduced to reduce the class noises. Finally, the improved SVMs were used as a base classifier to train different classifiers for the same training set. The results showed that it is superior to the single SVMs classification model. Huang and Zhang [23] proposed a new multi-feature model, aiming to construct a SVM ensemble combining multiple spectral and spatial features. It was found that the multi-feature model with semantic-based post-processing provides more accurate classification results (an accuracy improvement of 1–4% for the three experimental data sets) compared to the voting and probabilistic models.

Different classifiers resulted in different classes for the same test area. No single classifier can perform the best for all classes. In the hybrid classifiers-based approach, the classifiers should use an independent feature set and/or be trained on separate sets of training data. Two strategies exist for combining classifiers: (1) Classifier Ensembles (CE); and (2) Multiple Classifier Systems (MCS) [8]. A critical step is to develop suitable rules to combine the classification results from different classifiers. Previous research [24,25] has explored different techniques, such as a production rule, a sum rule, stacked regression methods, majority voting, and thresholds, to combine multiple classification results. In the present study, the aim was to develop such a hybrid intelligent system to improve the accuracy and efficiency of the object-based classification. It was focused on combining an ensemble learner and encapsulate expert knowledge within a rule-based system. In the proposed method, two methods of knowledge-based method (KBS) and SVM were combined; in other words, the MCS mode was used. Furthermore, this combination also applies the CE on a layered scoring manner. In this study, several scoring chances were created in different layers. In each layer, selections of samples were randomly controlled in the number of samples for each class relative to others in order to obtain a better distribution in the object-based sampling. On the other hand, processes per layer were performed in an object-based manner, thus, a few elements (relative to the pixel-based manner) were utilized. For this reason, the mentioned layered and randomized process proved to be feasible and cost-effective. Furthermore, due to the avoidance of the varied sizes of the segmentation objects' impact, the training data were produced and analyzed in a multi-scale manner and combined with different levels of scale. In addition to dealing with probable defects, as well as mixed segments in the samples, the KBS was used as an effective weight on the scores. Ultimately, for the final decision, the total score was obtained from the integration of misclassification cost among all classes for each layer in a weighted combination of all layers. Then, the system assigned the label of each segment, yielding the smallest average loss in a class.

This study aimed to improve the accuracy of object-based classification through the utilization of the advantages of ensemble learning methods, which are optimal in the process of object-based classification. Another goal was to use the advantage of the SVM approach, which is appropriate for dealing with a small number of training data (common in object-based classification). Accordingly, the present study proposes an object-oriented SVM within multi-layers by a weighted scoring mode in each iteration. In order to assess the proposed model, experiments were conducted on the internal evaluation of the proposed methods. Furthermore, for an external evaluation, an optimized SVM method and popular ensemble learning algorithms such as AdaBoost and RF were considered. Furthermore, by changing the input features, these were tested. Finally, the McNemar test was done. The experimental results indicate the efficiency of the proposed method.

2. Proposed Method

The general diagram of the proposed method is shown in Figure 1. In the following, each step (and used alphanumeric labels) is explained in more detail.
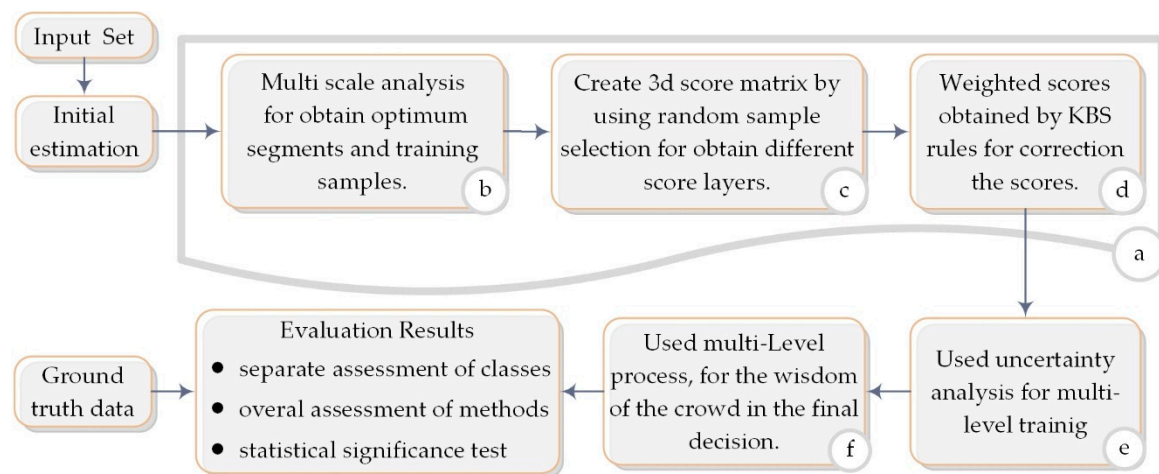
Figure 1. Workflow of the proposed method.

2.1. Initial Estimation

To obtain representative samples, several schemes are traditionally utilized in remote sensing, such as stratified random sampling. This strategy generally allows the reduction of the size of the training data required, but needs prior knowledge about the study site to construct the stratification [26]. In some ways, an unsupervised classification is used to provide data, e.g., [27], which proposed an approach that clustered the remote-sensing data by combining the fuzzy c-means clustering with SVM. This method does not have a specific view of target classes and works merely on the basis of clustering the number of classes; hence, it is not suitable for HSR images. This is due to the high level of details in HSR images. Similar studies have been conducted on images with medium and low resolution [28,29]. In this context, the uses of KBS are also useful [30]. Since the main classes (no subclasses) expected in the remote-sensing interpretation are limited, the required rules would remain stable, to a large extent, for a specified condition (images with the same resolution in specific scenes, such as high-resolution images in urban areas). In this research, the KBS was incorporated.

Image classes are associated with environmental concepts, for the identification of each class a series of indices and descriptors can be used. In the following, some of the features used in the initial estimation are presented. In order to identify vegetation cover, the SAVI index (I) [31], and for water, the NDWI (II), were used. In DSM data filtering, the method [32] was used and was automatically classified into ground and off-ground pixels [33] (III). Furthermore, the calculation of the surface normal (Normal Z) for DSM (IV) can be effective in the analysis of image classes. The difference band $(|R-G| + |R-NIR|)/(R + G + 2 \times NIR)$ was used for the road index (V). The lightness value (value or V) from the HSV color space was adjusted and used for the dark area (VI). It should be noted that the above-mentioned indices do not indicate the necessity of identifying target classes. In other words, the mentioned features can identify prone regions.

After preparing and storing the data, a preliminary labeling was used to determine the estimation of the initial interpretation process of the regions. The present research has been organized on the basis of the structure of high-resolution images in a semi-urban area. The target classes are the building, tree, low-elevation vegetation, water, and road class (including parking spots). Furthermore, the rules have been defined generally in order to maintain the comprehensiveness of the rules to a large extent. The features I-III-IV were used for the building class. The vegetation class used the same features, but were considered by specific conditions for each class. For example, the non-elevated vegetation classes can be valid in filtering by accepting I, denial III and accepting the IV features. The road class used I-II-III-IV-V features, and for the water class, I-II-III-V-VI features were also used. The initial estimation of the regions was based on a hierarchical system, so the binary mask of the mentioned features was used. The binary features were obtained in an automatic manner according to Otsu's method [34], in

which an adaptive image threshold chooses on the basis of the local first-order statistics. Since this training data was the same in all the comparative and proposed methods, their production was not the goal and only were discussed to represent general trends in the study. Input training samples were automatically generated and were not perfect, hence, some problems, such as a mixed sample in the data (noise in training samples occurring in HSR images), lack of comprehensive samples (training data defects), imbalanced samples, and varied sizes of segmentation objects may be found in the training data (various errors of training data are expressed in the introduction, Section 1).

2.2. Multi Scale

The difference in shape and size of the real-world objects will cause complexity in its detection using one scale parameter in addition to the signature and illuminance of their spectral changes. The degree of heterogeneity within an object is controlled by a subjective measure called the 'scale parameter'. The use of the hierarchical principle is one of the solutions to deal with this problem in creating the objects. The multi-resolution segmentation (MRS) of the image is based on the hierarchical theory, where multiple scales are used to create the image objects [35]. Dragut proposed the Estimation of Scale Parameters (ESP) method that builds on the idea of local variance (LV) of object heterogeneity within a scene [36]. The ESP iteratively generates image-objects at multiple scale levels in a bottom-up approach and calculates the LV for each scale. This tool is an automated methodology for the selection of scale parameters to extract three distinct scales using MRS. Its third levels produced segments much larger than our classes, so only level 1 and 2 were considered in this research (Figure 2).
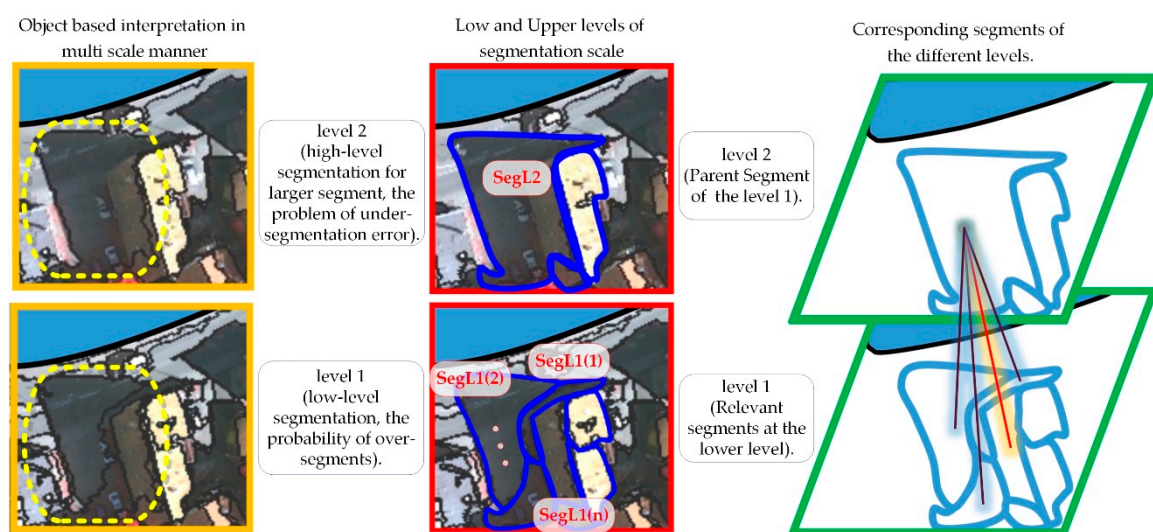

Figure 2. Multi-scale segmentation, identification of relevant parts at different scale levels.

The results of the segmentation process directly affected the final classification results. The under-segmentation error cannot be corrected in the following labeling processing, and the error will be inevitable in the next steps. However, as long as over-segmentation remains at an acceptable level, segmentation errors can be rejected so that a high level of classification accuracy can be achieved.

In terms of object-based image classification, the sampling method undoubtedly constitutes a crucial step. In this regard, the varied sizes of segmentation objects pose sampling difficulties and specificity in the process of object-based classification. In this regard, we used a multi-scale analysis in the proposed framework, which is advantageous in such an application. Accordingly, in this study, segments in discrete levels (different scale) were considered to generate the training samples for improving the training sample objects with different sizes problem. First, over-segments were introduced at the lowest level of the scale, then in upper scale level, the bigger segments were extracted (Figure 3). Training samples were extracted in the combined scale level. On the basis of which,

the levels of scale were entered into the process of scoring and interpretation of the decision (not just applied at the level of results). This brought the process of labeling closer to a structural and natural reality.
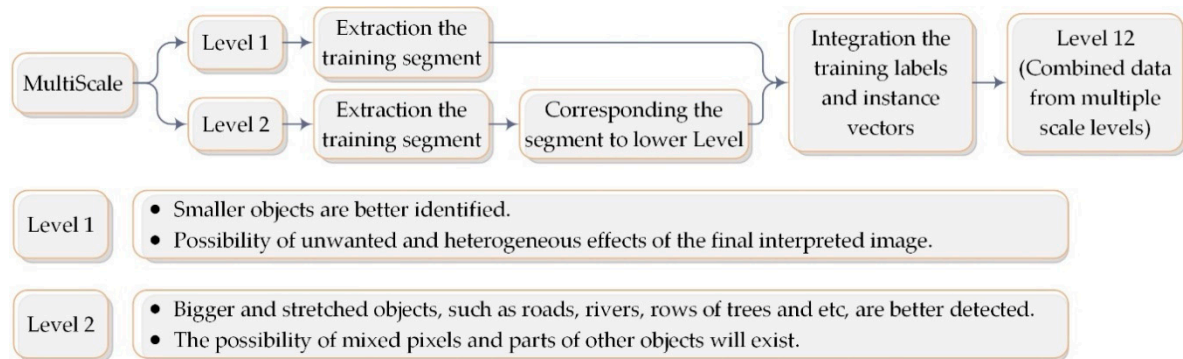

Figure 3. Training data sampling at different levels of scale.

2.3. 3D Score Matrix

For accurate supervised classification, precise and varied training samples are required. As mentioned in Section 2.1, the obtained training samples may have mixed and bias samples; furthermore, they suffer from a lack of comprehensive samples. Hence, in this section, using a random mode in the selection of training samples created various opportunities for system training, in order to finally be able to make the right decision, based on the comparison of layers and collective wisdom. As stated in the introduction, supervised classifications are faced with challenges, such as the imbalance between high dimensions and limited availability of training samples, that are very important in object-based classification. Based on this, in the proposed method, the number of samples in each layer was set with a minimum of samples in all classes. The number of samples in each training process (in one layer), as well as the total cycle (layers), was obtained from Equation (1).

$$\begin{cases} sampels & = \min(the\ number\ of\ train\ segments\ per\ class)/x & if\ sampels < x & = x \\ layers & = number(All\ training\ samples\ in\ all\ scale\ levels)/x & if\ layers < x^2 & = x^2 \end{cases}, \quad (1)$$

We obtained the constant value of "x" with different tests on various images and with different trial and error tests, equal to 10. Then, the scores derived from the image classification on each iteration (iterations number was equal to layers number) formed each layer of the three-dimensional score matrix. Each layer was a matrix of SVM scores per class, which was a numeric matrix with p-by-K array, where p is the number of observations in an image (rows), and K is the number of classes (columns), as shown in Figure 4. The score was the negated average of binary-losses. An array of the negated average binary learner loss per class for each segment determines how well a binary learner classifies an observation into the class [37,38]. In this research, its normalized form was used for each observation. Finally, it can be said that the scores indicated the likelihood that a label came from a particular class. Then, for each observation (segment) in an image, the label could assign to the class by the largest negated average binary loss (or, equivalently, the smallest average binary loss).

According to Equation (1), different modes of choosing the training data could be obtained randomly. The random mode in the selection of training samples was created through the various chances for classification, and consequently, different scores could be obtained. In a scene, due to the training data defects, or images with specific conditions, such as shadowy areas or classes with similar features, there was a potential for mistakes in the labeling. However, despite all these shortcomings, it can be said that the similarity of a segment was still high for components of the same class as compared to others (average rates in all classes) [39]. In these different layers, the balanced samples (in terms of the number, limited to the maximum number) were used in all classes. Furthermore, as mentioned, the

high similarity of each segment to its true class exists. Due to mixed samples etc., sometimes in some random combinations, a segment mistakenly obtained a higher score for a class other than its original class, but still at the same layer, and its original class score was pretty high. As a result, by combining different modes, it was finally expected that the average scores earned for that segment obtained the best score in its original class than others.
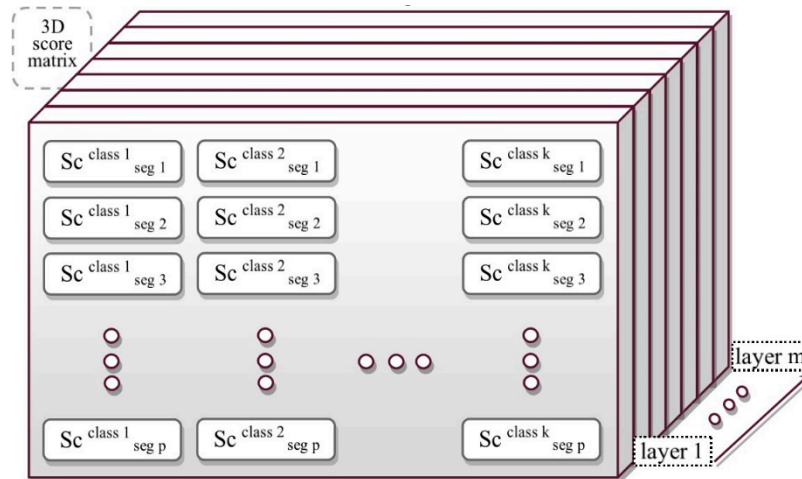

Figure 4. Three-dimensional score (Sc) matrix.

2.4. Weighted Scores

In order to improve the training process and compensate the obtained training sample defects in the expression of the target class properties, the correction step was done. By using the inherent characteristics of the target class that was largely independent of the scene, this could be done automatically. This control and correction were done by the knowledge-based rules. The KBS system prepares the possibility of weighting and can be used as an obligatory condition (such as elevation for buildings) or a positive effect for attainment of a class. In the proposed method, for a case with k classes, in the previous section, the k×m value was obtained for each image segment. Then, the weight of each value for a segment (such as the i'th segment) was obtained from the combination of the KBS rules, which was determined using Equations (2)–(6). The proposed rules are generally defined, in other words, not for a particular class, but for a group class, such as high-elevated objects, all vegetation covers, and so on. The weight of the classes associated with the vegetation cover and water area were obtained from Equations (2) and (3).

$$\text{Vegetation}(V)\begin{cases} w(c) = T_{\min}^{L-V} - T_{\min}^{L-V} \times e^{-3.Seg(SAVI)}, & c: L-V = \text{Low Vegetation} \\ w(c) = T_{\min}^{H-V} - T_{\min}^{H-V} \times e^{-3.Seg(SAVI)} - T_{\max}^{H-NonV}, & c: H-V = \text{High Vegetation} \\ \text{Non-Vegetation}(NonV) \quad --------- \end{cases}$$

(2)

$$\text{Water}(W) \quad \{ w(c) = Seg(NDWI) - T_{\min}^{W}, \quad \text{c:w} = \text{Water}$$
$$Seg(feture): \text{Segment value in the desired feature}$$

(3)

where $T_{\min}$ and $T_{\max}$ corresponded to the training samples of a considered class (Equation (4)). For example, $T_{\max}^{\text{H-NonV}}$ is obtained from the training samples of the non-vegetation elevated class (such as a building).

$$\text{T}_{\min} = \text{Average} - std, \quad \text{T}_{\max} = \text{Average} + std,$$

(4)

In order to formalize some concepts, quantitative rules were used. To implement the neighborhood rules, some cases were calculated, such as a near (small) buffer and far (large) buffer, around the segment (seg) to check the features (for example, DSM, labels, etc.) in its vicinity. For the small buffer, the neighborhood was about one meter. The large buffer was a distance of about half the oval radius

(the lengths of the semi-major and semi-minor axis) around the circumscribed ellipse in the desired segment. The weight of classes that are related to elevated objects were obtained according to the elevation data of the adjacent segments from Equation (5) and from the data obtained from the initial labeling (the label obtained from the classification of all training data) in Equation (6).

$$
E\begin{cases}
& c : E = Elevated\ class = tree, building, etc. \\
I : If\left(DSM_{seg} - DSM_{near}\right) > 0 & w(c) = +\log(1 + \left|DSM_{seg} - DSM_{near}\right|) \\
II : If\left(DSM_{seg} - DSM_{far}\right) > 0 & w(c) = +\log(1 + \left|DSM_{seg} - DSM_{far}\right|) \\
If\ (I\&II) < -1_{meter} & w(c) = -\log(average\{\left|DSM_{seg} - DSM_{near}\right|, \left|DSM_{seg} - DSM_{far}\right|\})
\end{cases}, \quad (5)
$$

$$
If\ Number\ of\left(Label\{E_{seg}\} - Label\{NonE_{seg}\}\right) > 0, \quad c : E = Elevated\ class = tree, building, etc.
$$

$$
\begin{cases}
if & Number\ of\ \left(Label\{E_{near}\} - Label\{NonE_{near}\}\right) > 0, \\
& w(c) = \frac{|a|}{a}(1 + \log|a|)/3 \quad , \quad a = 1 + \left(DSM_{seg} - DSM_{near}\right) \\
if & Number\ of\ \left(Label\{E_{far}\} - Label\{NonE_{far}\}\right) > 0, \\
& w(c) = \frac{|a|}{a}(1 + \log|a|)/3 \quad , \quad a = 1 + \left(DSM_{seg} - DSM_{far}\right)
\end{cases} \quad (6)
$$

In Equation (6), for each neighborhood (defined boundaries) the number of pixels belonging to different classes (in the initial labeling) was calculated and used. Finally, new scores (Equation (7)) for each segment per each class (individually) were obtained on the basis of the combination of earned scores for that segment in a gained weight vector (Figure 5).

$$
NewScore_i^c = \frac{\sum\limits_{l=1}^{m} Sc_i^c(l)}{m} \cdot \left(1 + w_i^c\right) \quad \begin{matrix} i = 1 : p \\ l = 1 : m \\ c = 1 : k \end{matrix}, \quad (7)
$$

where p was the total number of segments, k the number of given classes, and w the effective weights for the scores. Accordingly, the $NewScore^c_i(l)$ was the updated score of the ith segment. Then, by using the new scores, the new conditions were obtained for the labeling of that segment. In each segment, the label was equal to the highest score earned for that in all classes (Equation (8)).

$$
NewLabel_i = \left(\max\left\{ NewScore_i^1, \quad NewScore_i^2, \quad \ldots, \quad NewScore_i^k \right\} \equiv c\right), \quad i = 1 : p, \quad (8)
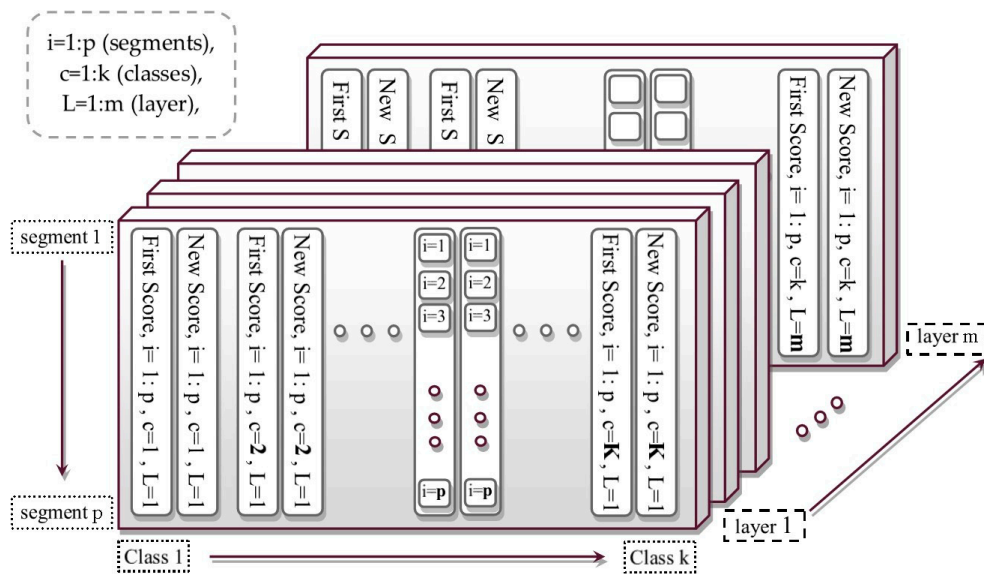$$

Figure 5. New scoring obtained from the weighted scores in three-dimensional score matrix.

2.5. Uncertainty Analysis

In the previous sections, based on the scores earned on each layer and their weighted combination, new scores were obtained. In this regard, in the scoring process described in Sections 2 and 3 (Figure 4) for different layers (l = 1:m), the stability process to predict the selected class in each segment (I = 1:p) could be investigated. In this section, new training data was obtained on the basis of the obtained labels from each layer and the degree of stability of the labels in each segment for different layers. For this purpose, the differences between the base and highest scores obtained for each segment, as well as the entropy of each segment was examined (Equations (9)–(13)).

$$v_{1 \times k} = \sum_{l=1}^{m} \begin{pmatrix} \max(Sc^c(l)) - Sc^1(l) \\ \vdots \\ \max(Sc^c(l)) - Sc^k(l) \end{pmatrix} , \quad c = 1:k , \tag{9}$$

$$u_{1 \times k} = Number\ of\ seleted\ class : \sum_{l=1}^{m} \left(\max\left\{ Sc^1(l), Sc^2(l), \ldots, Sc^k(l) \right\} \equiv c \right), \tag{10}$$

$$U_{1 \times k} = \log(u_{1 \times k}) \quad , \quad V_{1 \times k} = \left(1 - \frac{v_{1 \times k}}{\max(v_{1 \times k})} \right) , \tag{11}$$

$$R_{1 \times k} = sort\left(\frac{U_{1 \times k}}{sum(U_{1 \times k})} + \frac{V_{1 \times k}}{sum(V_{1 \times k})} \right) \quad sort : \max\ to \min, \tag{12}$$

$$R1 = R_{1 \times k}(1) - R_{1 \times k}(2) \quad , \quad R2 = ShannonEntropy(U_{1 \times k}) , \tag{13}$$

Finally, given that if a segment label in most of the layers belongs to a class, it can be said that it has a high stability, hence the outputs of Equation (14) can be used as the new training samples for the next step. Furthermore, the multi-scale weight matrix (Equation (15)) was obtained based on the multi-scale analysis of NewLabel (Equation (8)).

$$if \quad R1 > 1 - \xi \quad \& \quad R2 < \xi \quad then \quad NewTraining\ Data = \left(\max(AllScore_i^c) \equiv c \right) , \tag{14}$$

$$w_{1 \times k}^{Multiscale} = \sum_{j=1}^{n} (NewLabel(j) \equiv c), \quad n{:}Number\ of\ corresponding\ segments\ of\ SegL2\ in\ SegL1 , \tag{15}$$

2.6. Final Decision

According to the flowchart (Figure 1), Sections 2.2 and 2.3 (step b and c) give a score-matrix from the initial estimation (primary training data, Section 2.1). Then by applying the weighted vector (Figures 1 and 6, step d), the NewScores could be obtained in Section 2.4. Furthermore, by filtering the score-matrix (step e), a new estimation (new training data) was obtained and by applying step a in the flowchart, the UncertaintyScores were acquired. Additionally, in Sections 2.4 and 2.5, the weighted vector was also obtained for each segment (Equation (16)), which could be used for the final decision-making process.

Finally, by applying Equation (17), the final scores were acquired. Then, by finding the highest score among the classes of each segment, the final label of each was determined (Equation (18)).

$$W_i^c = 1 + w_{1 \times k}(c) + w_{1 \times k}^{Multiscale}, \tag{16}$$

$$Final\ Score_i^c = NewScore_i^c \bullet W_i^c \bullet UncertaintyScore_i^c, \tag{17}$$

$$Final\ Label_i = \left(\max\left\{ Final\ Score_i^1, \quad \ldots \quad , \quad Final\ Score_i^k \right\} \equiv c \right) , \quad i = 1:p, c = 1:k , \tag{18}$$
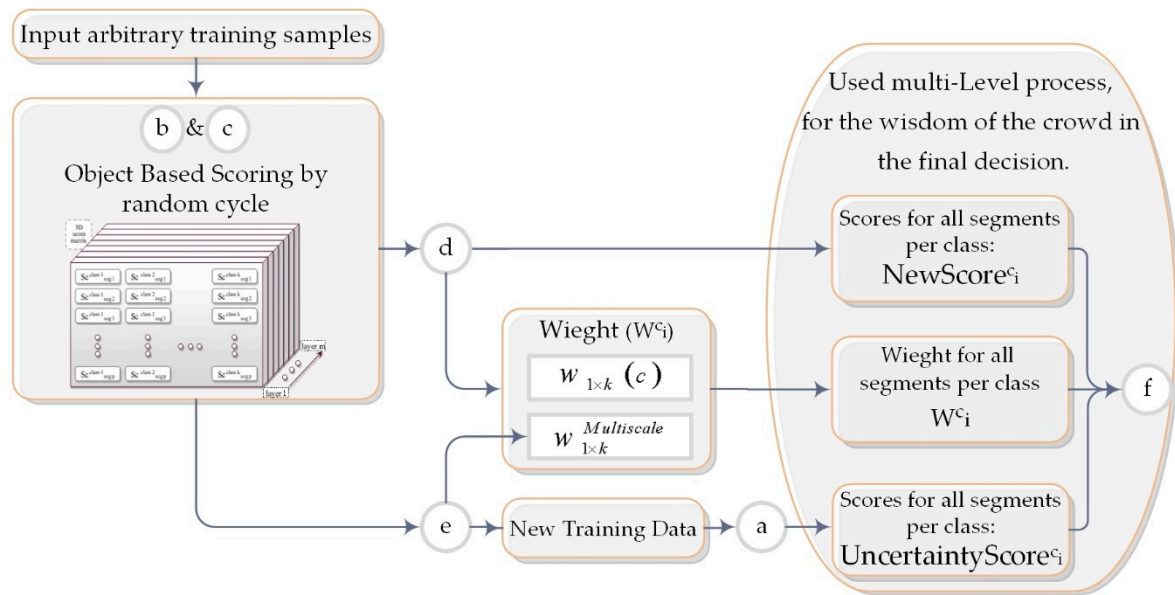
Figure 6. Workflow of the multi-level process in the final decision.

2.7. Accuracy Evaluation

In order to assess the classification accuracy, the classification results can be compared against the reference data (ground truth). Then, the F1_score, overall accuracy, and Kappa coefficient [40] can be calculated. However, not every difference is significant and, therefore, the statistical significance tests are required. A comparison with the McNemar test is perhaps the most recommended method for a thematic map accuracy comparison [41,42] that is based on a binary distinction between correct and incorrect class allocations (Equation (19)). This test is used to study the significance of the results' differences between the two methods. In this study, the McNemar test was run to analyze the proposed method results (Method 1) with each compared method (Method 2) with respect to the reference data. The difference inaccuracies were tested at 95% significance level. If Z_b was greater than 1.96, it meant that the two methods were significantly different from each other and that the two methods were not dependent on each other [43]. Z_b was computed as Equation (19).

$$z_b = \frac{\left| f_{12} - f_{21} \right|}{\sqrt{f_{12} + f_{21}}},$$

(19)

where f_{12} denotes the number of cases incorrectly classified by method 1 but correctly classified by method 2, and f_{21} denotes the number of cases correctly classified by method 1 and incorrectly classified by method 2. Furthermore, f_{11} was correctly classified and f_{22} was incorrectly classified in both methods.

3. Implementation and Results

3.1. Implementation

The data used in this research were aerial imagery with a resolution of 9 cm and three green, red, and NIR bands in Vaihingen city provided by ISPRS 2D semantic labeling [44] and nDSM data [33]. The implementation and evaluation of the proposed method was performed using MATLAB software. The basic computing unit in this method is the objects, hence, the MRS method was applied to the image; for this purpose, the ESP (estimation of scale parameter) was used for the automatic calculation of the optimal scale parameter [36]. The other parameters of this segmentation method, such as compactness and shape, in case of an accurate object extraction, required the precise adjustment of

the parameters. However, in this research, in order to investigate the effect of the proposed method and avoid the effect of the accuracy of selecting the parameters, the adjustment of such parameters was ignored and the default values (0.5 and 0.1 correspond to compactness and shape on default in eCognition software) were considered for all test images. In the proposed method, SVM was used with a linear kernel and the value of 1 for parameter C (default values). These were done to reduce the user dependency, and the impact of other factors was prevented. The features used in this research include the red, green, and NIR bands (the three-color bands) along with the SAVI index (With three different modes; L was used with values 0, 0.5 and 1). The lightness value (value or V) from the HSV color space was considered. For the elevation data, the nDSM, which was automatically generated by lastools-toolbox, was used [33]. Moreover, the slop, Normal-Z, and slop of DSM were extracted. Furthermore, the geometry property of segments, such as eccentricity, perimeter to area ratio etc., were used.

3.2. Results

In order to evaluate the proposed method, it was applied on the test images (the test images I to V correspond to Vaihingen 5, 7, 13, 26, and 28, respectively) in several ways. First, we evaluated the results of the proposed method on the test images and compared them with the base state in which all the training data was used (Table 1). Likewise, the layered form of the proposed method in majority voting mode was used for output labels (Table 1). Next, the results were reviewed at each class by F1_score criteria (Table 2). For more comparison, popular machine learning algorithms such as AdaBoost and RF were also studied (Table 3). In Table 4 we used an optimization algorithm in the parameter determination of the SVMs kernel. Furthermore, in order to evaluate the efficiency of the proposed method in different feature spaces as an input of the classification method, new features were extracted from the images band and DSM data. The results from the classification of images with these new features were examined by the proposed and compared methods (Tables 5 and 6). Finally, the McNemar test was done. In all the scenarios, the same training data was used. Furthermore, the ground truth including the whole image was used to evaluate all the tests (provided by [44]).

Table 1. Evaluation of the test images and internal investigation of the proposed method.

Test Image	Evaluation Parameters	Base	Majority Voting	Proposed Method
I	Kappa coefficient	65.52	68.82	77.65
	Overall accuracy	78.61	80.94	86.33
II	K.CO.	68.73	71.77	76.32
	O.A.	77.55	79.54	82.62
III	K.CO.	62.89	62.53	65.22
	O.A.	72.45	72.34	74.70
IV	K.CO.	56.70	66.73	75.81
	O.A.	68.63	75.02	81.93
V	K.CO.	66.59	66.87	71.36
	O.A.	75.05	75.28	78.74

Table 2. Results of F1_score evaluation parameters, per class in test images.

Test Image	Class	Base	Majority Voting	Proposed Method
IV	Building	62.53	78.84	86.83
	Grass land	49.60	53.07	63.14
	Tree	84.20	83.96	86.43
	Water body	41.52	72.52	76.00
	Road & Parking	70.09	73.29	81.76
V	Building	82.47	82.20	86.84
	Grass land	64.49	66.25	70.80
	Tree	72.51	72.31	73.73
	Road & Parking	76.76	76.85	79.89

Table 3. Comparisons with ensemble classification methods in different modes.

Test Image	Evaluation Parameters	RF #Tree	AdaBoost =100	RF #Tree	AdaBoost =150	RF #Tree	AdaBoost =200
I	K.CO.	68.08	65.53	67.1	63.56	67.56	63.4
	O.A.	80.37	78.52	79.81	77.28	80.04	77.16
II	K.CO.	70.91	67.02	71.28	66.04	70.08	64.94
	O.A.	78.85	76.31	79.17	75.57	78.24	74.78
III	K.CO.	61.84	62.99	61.64	62.35	61.57	62.44
	O.A.	71.59	72.61	71.44	72.09	71.39	72.18
IV	K.CO.	66.32	60.94	66.92	62.15	65.91	63.34
	O.A.	74.91	71.12	75.62	71.98	74.7	72.59
V	K.CO.	65.78	66.37	67.57	67.18	66.13	67.03
	O.A.	74.55	74.86	75.85	75.51	74.81	75.39

Table 4. Results of the implementation of comparative methods in terms of the classification method.

Test Image	Eva. Par. [1]	Pr. M. [2]	#Tree Pr. Layer [3]	Op. SVM [4]	RF	#Tree Best [5]	AdaBoost	#Tree Best
I	K.CO.	77.65	102	66.84	68.08	100	65.53	100
	O.A.	86.33		79.55	80.37		78.52	
II	K.CO.	76.32	100	71.53	71.28	150	67.02	100
	O.A.	82.62		79.51	79.17		76.31	
III	K.CO.	65.22	100	63.83	61.84	100	62.99	100
	O.A.	74.70		73.19	71.59		72.61	
IV	K.CO.	75.81	158	68.01	66.92	150	63.34	200
	O.A.	81.93		76.35	75.62		72.59	
V	K.CO.	71.36	152	67.61	67.57	150	67.18	150
	O.A.	78.74		75.79	75.85		75.51	

[1] Evaluation parameters, [2] proposed method, [3] #layers in the proposed method, [4] optimized, [5] best of Table 3.

In the comparative process, we tried to examine the proposed method in different aspects. Accordingly, in the base mode (Table 1, Column: *Base*), the used classifier was the same as in the proposed method, with the difference that all training data are used. Likewise, the majority voting mode used several layers and the training sampling was the same as in the proposed method, but all layer outputs were used as labels in the majority voting mode (Table 1, Column: *Majority vote*).

The results show that our approach increased the classification accuracy compared to the voting-based method. The accuracy depended on the scene conditions. The classification defects of test image IV were due to the presence of more shadowed areas, high interference between the water class and the road in the shadow, the existence of complex and diverse buildings with different roof slopes

(from flat to inclined), and also the dense buildings area. Hence, by maintaining the other conditions and only changing the classification method, we could obtain more improvement. Meanwhile, for some images (such as test image III), which has less class interference of the training data and separate buildings etc., a less improved accuracy was achieved. The shortage of its classification was more affected by the characteristics of the ground truth data and the scene conditions (in terms of the shrubs belonging to the high or low elevation vegetation class, as well as the separation of road and sparse vegetation cover, etc.). Furthermore, if the classification results were similar, the combination process could not improve the classification accuracy. Therefore, diversity is an important requirement for the success of hybrid systems [8,45]. In test image III, for all modes (Table 1) the results were the same (no significant changes), so a great improvement in the results was not expected. In order to compare the accuracy of the classes, the accuracy of each class was examined separately by F1_score. According to the number of test images and classes, only the two last test images in Table 1 are presented below (Table 2).

Table 5. Internal evaluation of the proposed method in terms of the type of features used.

Test Image	Eva. Par.	Without Multi-Layers [1]	Without Multi-Level [2]	Proposed Method
	K.CO.	63.14	64.86	65.31
III	O.A.	72.78	74.39	74.77
	Mc. Test [3]	180	50	
	K.CO.	74.37	75.49	78.23
IV	O.A.	80.76	81.41	83.50
	Mc. Test	210	210	

[1] Improved by using the weight without multi-layers, [2] without multi-level process, [3] McNemar's test (Z_b).

Table 6. Results of the implementation of comparative methods in terms of the type of features used.

#Layer	Evaluation Parameters	RF	Adaboost	Other Methods		
	K.CO.	60.44	61.36		K.CO.	61.87
100	O.A.	70.96	71.55	Op. SVM	O.A.	71.67
	Mc. Test	230	240		Mc. Test	190
	K.CO.	60.28	61.89		K.CO.	60.10
150	O.A.	70.84	71.96	All Maj. Vote	O.A.	70.71
	Mc. Test	250	190		Mc. Test	260
	K.CO.	60.05	61.54		K.CO.	60.36
200	O.A.	70.66	71.73	base	O.A.	70.56
	Mc. Test	260	200		Mc. Test	240

As seen in the class accuracy review, the proposed method had the highest accuracy in the building class, which is one of the most important urban indicators. In the test image IV, the precision of the water class was low in the base mode, since the asphalt pavement in the shaded areas (due to the coldness of the area) had a similar behavior to the water properties in the near-infrared band. Therefore, the precision of these classes was diminished (Table 2, Column: *Base*). In the majority voting, because we attempted to expand the distribution modes, the training segments in each layer were distributed regularly between classes and an improvement was achieved (Table 2, Column: *Majority vote*). Accordingly, the proposed method obtained appropriate results using the layered analysis system and weighted scores. In order to do a further evaluation, ensemble classification methods were considered.

In order to do a further evaluation, RF and AdaBoost methods were considered. The results of the RF method are presented on the test images in Table 3. Since the number of trees must be

defined in this method, this value was defined in three modes. The RF classification was implemented in the EnMAP-box software [46]. The EnMAP-box is a IDL-based tool for the classification and regression analysis of remote-sensing imagery. RF offers a cross-validation-like accuracy measure through the out-of-bag error estimate and gives an insight into the variable importance by assessing the accuracy loss when feature values are randomly permuted [47]. For greater comparison, the AdaBoost method [20,48], another popular ensemble classifier in the machine learning algorithms, was also studied (Table 3). The AdaBoost algorithm used classification trees as individual classifiers, and then a bootstrap sample of the training set was drawn using the weights for each trial on that iteration. The number of iterations and the number of trees were set to equal. For all tests, the inputs (used segments, training data, and features) were similar to the proposed method.

SVM uses kernel functions in order to map data into higher features space to obtain better results. For a comparison with the proposed method, in this step, the RBF kernel function was used. Hence, some parameters, such as the penalty term (C) and RBF kernel parameter were required to be optimally set. Ideal values for these parameters depended on the distribution of the classes in the feature space. In doing so, those parameters with the best performance were found by optimization methods (Table 4). Hence, we used the grid search to test ranges of parameters with an internal performance estimation for a new comparison method (Op. SVM). The accuracy of results during the grid search was monitored by three-fold cross-validation on the training data.

According to the results, RF generally exhibits a little better performance improvement over the AdaBoost. RF yields a generalization error rate that compares favorably to AdaBoost, yet is more robust to noise. For example, in image IV, which has a class interference and mixed sample, its accuracy improvement was greater. Furthermore, according to the results, it can be assumed that, on average, when the number of trees in the RF and AdaBoost methods was equal to the number of selected layers in the proposed method (which is automatically determined), it may be at is optimum.

In previous tests, the proposed method and all comparative methods used the same training data and similar features (Section 3.1). In order to evaluate the efficiency of the proposed method in different feature spaces, one of the most widely used methods was also used to produce useful features. The results from the classification of images with GLCM (Gray-Level Co-occurrence Matrix) features are presented in Tables 5 and 6. For this purpose, eight textual features (Contrast, Correlation, Dissimilarity, Entropy, Homogeneity, Mean, Second Moment, and Variance) of each image band and DSM, by the kernel dimensions of 3×3, 5×5 and 7×7, and in four directions (each 45 degrees) were extracted (384 feature bands). Then, it was averaged from different directions, so that the directional effect was eliminated in the production feature (96 feature bands). The original image and the nDSM were also used. Then, the classification was done by all available features (100 feature bands). For this, test images III and IV were considered as the lowest and highest difference in scores (Table 1).

In Table 5, two topics were examined. First, the effectiveness of the weights that were obtained by the automatic method in the proposed process (without multi-layers column) was checked. In doing so, after calculating the weights in the presented process (Section 2.4), they were applied to the scores (the same as in the proposed method but without the layered structure). Secondly, by evaluating the effectiveness of the multi-level process (without a multi-level column), the results of the internal evaluation of the proposed method were studied. In the analysis and comparison of various methods, any differences in results are not significant. Therefore, statistical tests are used to study the significance of the differences mentioned in the results (Mc Test row). According to the results, there was no dependency between the outputs of the mentioned methods and the proposed method. For more investigation, the process was repeated with competition methods and the results are listed for test image III (the image with the lowest improvement among all test images) in Table 6.

4. Discussion

The present study mainly aims to develop an object-based method in the learning process for classifying remote-sensing images. In OBIA, the basic unit of analysis is image objects; as a result, this

process is highly dependent on the initial segmentation, significantly affecting the improvement or weakening of the final results. This research, to some degree, attempts to untangle a few of the gaps in this field, especially by incorporating multi-scale analysis into the proposed framework, which is quite advantageous in such an application.

The classification methods, such as the SVM (due to their extensive adoptions and reliable performance for various remote sensing applications), RF, and AdaBoost methods (due to collective decision-making and good performance in dealing with the diversity of features), are the preferred methods for object-based image classification. ANN has the problem of overfitting, also, it is difficult to select the type of network architecture [49]. Fuzzy Classifiers is dependent on a priori knowledge, and without it, the output is not good. SVMs provide a good generalization and the problem of overfitting is controlled. Their computational efficiency is good and performs well with a minimum training set size (common in object-based classification) and high-dimensional data [8]. Compared to other methods, in the case of a limited number of training samples, SVMs have proven to be the best choice [49,50]. SVMs show a balance between the errors of the classes. Another property of SVM is the principle of margin maximization [19,23]. The special property of SVM is that it simultaneously minimizes the empirical classification error and maximizes the geometric margin. That is why SVM was employed in this research. However, to produce an appropriate model, it is dependent on training samples. An important point in producing training samples is that any inappropriate training samples are considered as the main source of mistakes in many classification processes [51,52].

In order to obtain desirable good results with a supervised algorithm, it is often necessary to collect large amounts of training data, particularly for the most heterogeneous classes [53], or representative training data which allow us to consider the entire diversity of the space or the classes studied [54]. We note that automatically generating training data for image labeling can be efficient only for the part of the image which includes the general properties of the target classes. These data have some problems. The first problem is that the process may be time-consuming. For example, in the knowledge-based method, the data may have a large volume, which increases the processing cost (unlike the supervised training data, which is small but almost covers the characteristics of the target class) while also having little information (most of the information is the same). The second problem is the lack of completeness; e.g., it does not cover the entire diversity of the classes studied and it may only cover the characteristics of one mode instead of different behaviors of the same class. Furthermore, the mixed-objects and errors in the training data lead to mistakes in the model training. According to the above, the production training data in addition to the mixed sample and mistake samples can also include defects.

In the face of these problems, the proposed method uses the object-based method in a multi-scale manner and various scoring modes in different layers to improve the mentioned problems (Section 2.3). Due to performing the analysis on pixel groups, the object-based methods reduce both the noise and computational cost (since they process an average of the existing data). Furthermore, because the analysis of the image is performed on a segment instead of the pixels, the process of grading the scores and performing the process in different layers can be implemented with a reasonable cost. According to the results (Table 1), merely applying the object-based classification is not effective in a repetitive mode. For this, the process of weighted scoring was added to achieve better accuracy. Furthermore, to improve the accuracy of classification, in regions that cannot be correctly predicted due to defects in training data, the successive scoring process has been used by creating different situations. In these areas, due to defects, there is a potential mistake in labeling. Nevertheless, despite all these shortcomings, it can be said that the similarity of a segment is still high for components of the same class as compared to others [39]. In this regard, the use of score weight (Section 2.4), different segmentation levels (Figure 2), neighboring information (Equations (5) and (6)), and a multi-level process (Section 2.5) can improve accuracy. Eventually, the final scores were calculated by the combination of different scores obtained in the proposed process. Finally, in addition to the advantage of SVM, collective decision-making methods were employed to improve the classification process (Section 2.6).

In the base state, all training data were used (Table 1). However, in the multi-layer analysis, in each layer (each cycle), some training samples were randomly selected. In this manner, in each layer, a variety of situations arose, causing a different scoring for each segment. However, since the majority vote did not deal with scores, the labels of each layer were examined and, as a result, the layers' answers were hard-coded (zero-one). As a result, the scores of classes were close to the original label (probable classes) and the variety of scores in different layers was lost, a conclusion which may not be optimal (Table 1, Column: *Majority vote*). In order to overcome this problem, in this study, the decision process became more flexible using the scores of not selected classes in each segment and the variety of scores in different layers. Finally, based on the final score resulting from the weighted combination of different layers, the label was selected for each segment. This weight was generated on the overall properties of target classes and used the spatial relationships with neighboring objects, and different scale level data were improved to obtain better classification results.

In this research, an image interpretation system was presented and tested for remote-sensing images based on the weighted scoring method in the object-based process by observing the class scale levels during the interpretation process. In the proposed process and comparison methods, input data, segments, and initial training data were the same. The proposed method worked automatically, and, as such, there was no need for the user to set any threshold. Finally, the numerical results of the proposed method on the test images showed its proper quality. Since the main classes (no subclasses) expected in the remote sensing interpretative procedure were limited, the required general rules remained stable to a large extent. However, it was difficult to develop an overall algorithm for interpreting all the classes because the image properties of different regions differed in environmental and scene features. In other words, the semantic rules held true in the case of maintaining the conditions, type of scene, and geographical area; nevertheless, with wide variations in the conditions of class and scene, the scene would need a proportionate updating of the knowledge base. The results demonstrated that the proposed technique is desirable as a semi-automatic method to interpret the high resolution of the semi-urban regions; still, this process can be completed in future studies.

5. Conclusions

In this research, it was attempted to develop an object-based scoring procedure to improve image interpretation in a multi-scale manner. As stated before, object-based classifications are faced with challenges, such as the imbalance samples, and discrepancies between mixed and homogeneous objects. Based on this, the present study also considered the labeling of image during supervised object-based classification.

In order to solve the problems, the simultaneous use of all training data may not produce an ideal response to the object-based classification process. By managing samples in different layers and observing the ratio of the number of training samples in various classes, one can create a variety of modes for each segment in different layers. Consequently, for each segment, different scores were assigned on the basis of the similarity of that segment to different classes. It is expected that the similarity of any segments under different conditions, such as shadows, remains high compared to the other classes. In addition, the individual weighting of each layer caused an improvement in the scoring process, and to some extent, improved on the deficiencies in the training data. In other words, scores obtained at each layer will be improved by weight correction. Finally, by integrating improvement scores in all layers, the chosen class in each segment will be a class which earns a higher score than others.

- In this process, the selection of training samples was performed in a layered form. This was done to create various opportunities for system training. In addition, we were able to make an optimum decision based on the comparison of the scores of layers and collective wisdom. The use of the SVM method in each layer can be well suited to the mentioned conditions in the object-based classification in order to make a maximum-margin separation between the classes.

- In addition to dealing with probable defects as well as mixed segments in the samples, the KBS was used as an effective weight on the scores. In the employed KBS system, by considering the neighboring effect in the spatial domain, the spatial effect was also entered in the analysis in addition to the spectral data.
- The use of weighted scores for all layers (instead of the labels) reduces the effect of similar scores in an observation and mixed classes, which improved the decision-making process.
- In terms of object-based image classification, the varied sizes of segmentation objects caused sampling difficulties in the process of object-based classification. Accordingly, in the proposed method, training samples were extracted in the combined scale level.
- Another relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide. In the used KBS system, by considering the neighboring effect in the spatial domain, in addition to the spectral data, the spatial effect as a weight was also entered into the analysis.
- Ultimately, for the final decision, the total score (instead of the resulting labels) obtained from the integration of different modes was incorporated, thus decreasing the effect of similar scores and mixed classes that weakened the decision-making process.

The proposed method combined the obtained misclassification cost for all classes in SVM classification and the ensemble learning idea; furthermore, in each cycle, by controlling the distribution of the training sample, an equitable distribution of the classes' priority in the classification of that layer was caused. In addition, due to the existing layered structure, it also had the collective decision-making property that was carried out in a weighted scoring process.

To evaluate the efficiency of the system, it was tested in the semi-urban area. Furthermore, the relative validity of the method was verified by the McNemar test. Overall, the results showed a proper performance. The results also demonstrated that the kappa coefficient of the proposed methods improved by 9.5% compared to the base method, and its accuracy was also improved by 8% and 6% compared to AdaBoost and RF (on average for all five test images), respectively. In this research, the aim was to determine the degree of accuracy improvement, so the parameters of the segmentation and the SVM method were used in the default mode, and the absolute accuracy of the proposed method was affected. Therefore, by considering the layered structure of the proposed method, future studies can use optimization methods for determining the segmentation and classification parameters (used in this study as a constant and default).

References

1. Ma, L.; Fu, T.; Blaschke, T.; Li, M.; Tiede, D.; Zhou, Z.; Ma, X.; Chen, D. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 51. [CrossRef]

2. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]

3. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F. Geographic object-based image analysis–towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [CrossRef] [PubMed]

4. Millard, K.; Richardson, M. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sens.* **2015**, *7*, 8489–8515. [CrossRef]

5. Costa, H.; Foody, G.M.; Boyd, D.S. Using mixed objects in the training of object-based image classifications. *Remote Sens. Environ.* **2017**, *190*, 188–197. [CrossRef]

6. Li, M.; Ma, L.; Blaschke, T.; Cheng, L.; Tiede, D. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 87–98. [CrossRef]

7. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

8. Salah, M. A survey of modern classification techniques in remote sensing for improved image classification. *J. Geomat.* **2017**, *11*, 1–21.

9. Huang, C.; Davis, L.; Townshend, J. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [CrossRef]

10. Pal, M.; Mather, P. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [CrossRef]

11. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]

12. Adam, E.; Mutanga, O.; Odindi, J.; Abdel-Rahman, E.M. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *Int. J. Remote Sens.* **2014**, *35*, 3440–3458. [CrossRef]

13. Zhang, C.; Xie, Z. Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques. *Wetlands* **2013**, *33*, 233–244. [CrossRef]

14. Maxwell, A.; Strager, M.; Warner, T.; Zegre, N.; Yuill, C. Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation. *GISci. Remote Sens.* **2014**, *51*, 301–320. [CrossRef]

15. Maxwell, A.E.; Warner, T.A.; Strager, M.P.; Pal, M. Combining RapidEye satellite imagery and Lidar for mapping of mining and mine reclamation. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 179–189. [CrossRef]

16. Maxwell, A.; Warner, T.; Strager, M.; Conley, J.; Sharp, A. Assessing machine-learning algorithms and image-and lidar-derived variables for GEOBIA classification of mining and mine reclamation. *Int. J. Remote Sens.* **2015**, *36*, 954–978. [CrossRef]

17. Ghosh, A.; Fassnacht, F.E.; Joshi, P.K.; Koch, B. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 49–63. [CrossRef]

18. Chan, J.C.W.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011. [CrossRef]

19. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [CrossRef]

20. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson Education Limited: Kuala Lumpur, Malaysia, 2016.

21. Bekkar, M.; Alitouche, T.A. Imbalanced data learning approaches review. *Int. J. Data Min. Knowl. Manag. Process* **2013**, *3*, 15–33. [CrossRef]

22. Liu, Y.; Huang, L. A novel ensemble support vector machine model for land cover classification. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1–9. [CrossRef]

23. Huang, X.; Zhang, L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 257–272. [CrossRef]

24. Liu, W.; Gopal, S.; Woodcock, C.E. Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 963–971. [CrossRef]

25. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]

26. Rougier, S.; Puissant, A.; Stumpf, A.; Lachiche, N. Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *51*, 60–73. [CrossRef]

27. Huang, Q.; Wu, G.; Chen, J.; Chu, H. Automated remote sensing image classification method based on FCM and SVM. In Proceedings of the 2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering, Nanjing, China, 1–3 June 2012; pp. 1–4.

28. Yu, X.; Zhou, W.; He, H. A method of remote sensing image auto classification based on interval type-2 fuzzy c-means. In Proceedings of the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing, China, 6–11 July 2014; pp. 223–228.

29. Lia, Q.; Baob, W.; Lic, X.; Lid, B. High Resolution Remote Sensing Image Classification based on SVM and FCM. In Proceedings of the 2015 2nd International Conference on Electrical, Computer Engineering and Electronics (ICECEE 2015), Jinan, China, 29–31 May 2015; pp. 1271–1278.

30. Kiani, A.; Ebadi, H.; Ahmadi, F.F.; Masoumi, S. Design and Implementation of an Expert Interpreter System for Intelligent Acquisition of Spatial Data from Aerial or Remotely Sensed Images. *Measurement* **2014**, *47*, 676–685. [CrossRef]

31. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

32. Axelsson, P. DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 111–118.

33. Gerke, M. Normalized DSM-Heights Encoded in Dm-See Report for Details. Available online: https://www.researchgate.net/profile/Markus_Gerke/publication (accessed on 1 December 2014).

34. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]

35. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258. [CrossRef]

36. Drăguţ, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127. [CrossRef] [PubMed]

37. Allwein, E.L.; Schapire, R.E.; Singer, Y. Reducing multiclass to binary: A unifying approach for margin classifiers. *J. Mach. Learn. Res.* **2000**, *1*, 113–141.

38. Escalera, S.; Pujol, O.; Radeva, P. On the decoding process in ternary error-correcting output codes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 120–134. [CrossRef] [PubMed]

39. Ning, J.; Zhang, L.; Zhang, D.; Wu, C. Interactive image segmentation by maximal similarity based region merging. *Pattern Recognit.* **2010**, *43*, 445–456. [CrossRef]

40. Carletta, J. Assessing agreement on classification tasks: The kappa statistic. *Comput. Linguist.* **1996**, *22*, 249–254.

41. de Leeuw, J.; Jia, H.; Yang, L.; Liu, X.; Schmidt, K.; Skidmore, A. Comparing accuracy assessments to infer superiority of image classification methods. *Int. J. Remote Sens.* **2006**, *27*, 223–232. [CrossRef]

42. Foody, G.M. Thematic map comparison. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633. [CrossRef]

43. Mushore, T.D.; Mutanga, O.; Odindi, J.; Dube, T. Assessing the potential of integrated Landsat 8 thermal bands, with the traditional reflective bands and derived vegetation indices in classifying urban landscapes. *Geocarto Int.* **2017**, *32*, 886–899. [CrossRef]

44. Rottensteiner, F.; Sohn, G.; Gerke, M.; Wegner, J.D. ISPRS Test Project on Urban Classification and 3D Building Reconstruction. Commission III-Photogrammetric Computer Vision and Image Analysis, Working Group III/4-3D Scene Analysis. 2013. Available online: http://www.cvlibs.net/projects/autonomous_vision_survey/literature/Rottensteiner2013.pdf (accessed on 1 January 2016).

45. Chandra, A.; Yao, X. Evolving hybrid ensembles of learning machines for better generalisation. *Neurocomputing* **2006**, *69*, 686–700. [CrossRef]

46. van der Linden, S.; Rabe, A.; Held, M.; Jakimow, B.; Leitão, P.J.; Okujeni, A.; Schwieder, M.; Suess, S.; Hostert, P. The EnMAP-Box—A toolbox and application programming interface for EnMAP data processing. *Remote Sens.* **2015**, *7*, 11249–11266. [CrossRef]

47. Waske, B.; van der Linden, S.; Oldenburg, C.; Jakimow, B.; Rabe, A.; Hostert, P. ImageRF—A user-oriented implementation for remote sensing image analysis with Random Forests. *Environ. Model. Softw.* **2012**, *35*, 192–193. [CrossRef]

48. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning (ICML'96), Bari, Italy, 3–6 July 1996; pp. 148–156.

49. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]

50. Rogan, J.; Franklin, J.; Stow, D.; Miller, J.; Woodcock, C.; Roberts, D. Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sens. Environ.* **2008**, *112*, 2272–2283. [CrossRef]

51. Pal, M.; Mather, P. Some issues in the classification of DAIS hyperspectral data. *Int. J. Remote Sens.* **2006**, *27*, 2895–2916. [CrossRef]

52. Radoux, J.; Lamarche, C.; Van Bogaert, E.; Bontemps, S.; Brockmann, C.; Defourny, P. Automated training sample extraction for global land cover mapping. *Remote Sens.* **2014**, *6*, 3965–3987. [CrossRef]

53. Chen, D.; Stow, D. The effect of training strategies on supervised classification at different spatial resolutions. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 1155–1162.

54. Foody, G.M.; Mathur, A. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sens. Environ.* **2004**, *93*, 107–117. [CrossRef]