

Review

Multiple Classifier System for Remote Sensing Image Classification: A Review

Peijun Du ^{1,2,*}, Junshi Xia ², Wei Zhang ³, Kun Tan ², Yi Liu ² and Sicong Liu ²

¹ Department of Geographical Information Science, Nanjing University, Nanjing 210093, China

² Key Laboratory for Land Environment and Disaster Monitoring of State Bureau of Surveying and Mapping of China, China University of Mining and Technology, Xuzhou 221116, China;
E-Mails: xiajunshi@126.com (J.X.); tankun@cumt.edu.cn (K.T.); rs.liuyi@163.com (Y.L.); allencong@163.com (S.L.)

³ Hebei Bureau of Surveying and Mapping, Shijiazhuang 050031, China;
E-Mail: cumtwzh@163.com

* Author to whom correspondence should be addressed; E-Mail: dupjrs@126.com;
Tel.: +86-139-5219-6497; Fax: +86-25-8359-2686.

Received: 22 February 2012; in revised form: 22 March 2012 / Accepted: 6 April 2012 /

Published: 12 April 2012

Abstract: Over the last two decades, multiple classifier system (MCS) or classifier ensemble has shown great potential to improve the accuracy and reliability of remote sensing image classification. Although there are lots of literatures covering the MCS approaches, there is a lack of a comprehensive literature review which presents an overall architecture of the basic principles and trends behind the design of remote sensing classifier ensemble. Therefore, in order to give a reference point for MCS approaches, this paper attempts to explicitly review the remote sensing implementations of MCS and proposes some modified approaches. The effectiveness of existing and improved algorithms are analyzed and evaluated by multi-source remotely sensed images, including high spatial resolution image (QuickBird), hyperspectral image (OMISII) and multi-spectral image (Landsat ETM+). Experimental results demonstrate that MCS can effectively improve the accuracy and stability of remote sensing image classification, and diversity measures play an active role for the combination of multiple classifiers. Furthermore, this survey provides a roadmap to guide future research, algorithm enhancement and facilitate knowledge accumulation of MCS in remote sensing community.

Keywords: multiple classifier system; classifier ensemble; remote sensing; classification

1. Introduction

Remote sensing images are widely used for land cover classification, target identification and thematic mapping from local to global scales owing to its technical advantages such as multi-resolution, wide coverage, repeatable observation and multi/hyperspectral-spectral records [1]. With the advancement of remote sensing data acquisition technology, remote sensing images can be acquired by various sensors, for example, hyperspectral imaging spectrometer, high resolution sensors, polarimetric synthetic aperture radar, *etc.* [2]. Under this situation, image classification techniques are exposed with new challenges to process the multi-source data and serve to multidisciplinary applications [2,3]. One effective solution is to generate a classifier ensemble by combining some individual classifiers, which is named as multiple classifier system (MCS) or classifier ensemble [4–10]. In the past twenty years, MCS has developed rapidly and been widely used in various fields such as pattern recognition, image processing and target identification. Besides, MCS has become a hot topic in the attractive series international workshops, mainly because of its capability to improve accuracy and efficiency [11–14].

MCS, which has been utilized by the remote sensing society recently, is viewed as one effective way to improve classification performance of remotely sensed imagery [3,4]. Many researchers investigated the scheme of combining several classifiers to generate a single result. Both theoretical and empirical studies indicate that a good MCS is one where individual classifiers are accurate and at the same time disagree on some different parts of the input space [15]. Two popular approaches for creating accurate ensemble are Bagging and Boosting [16,17]. Bagging uses bootstrap sampling to generate accurate ensemble [16]. Boosting is a general method of producing a very accurate prediction rule by combining rough and moderately inaccurate learner [17].

Some previous studies demonstrated that the simple majority vote of the classifier prediction is an effective strategy. Within this scheme, a pixel is classified as the class that was chosen by the majority of the individual classifiers. Foody *et al.* used simple majority vote rule to integrate multiple binary classifiers for the mapping of a specific class [7]. Benediktsson and Swain proposed a consensus theoretic classification approach and designed a hybrid consensus theoretic classification scheme [18,19]. The experimental results using hybrid consensus theoretic classification scheme outperform other statistical/neural network and consensus theoretic algorithms. Doan and Foody investigated the combination of soft classification methods for remote sensing image and found that classification combination could improve the accuracy [8]. Fauvel *et al.* fused the classification results derived from a neural network and a fuzzy classifier using fuzzy sets theory, and the proposed method improved the classification accuracy compared to the separate use of different classifiers [20]. Smits focused on the family of MCSs based on dynamic classifier selection (DCS) and proposed a modification to dynamic classifier selection by local accuracy (DCS-LA) [9]. Experiments had shown that DCS-LA is a valid alternative method to majority vote.

In addition, MCS was also used for multi-source and hyperspectral remote sensing images. Briem *et al.* applied multiple classifiers (minimum Euclidean distance classifier, maximum likelihood classifier, conjugate-gradient back propagation, decision table, J48, 1R) to multisource remote sensing data (multispectral, synthetic aperture radar and topographical data sets) and demonstrated that Bagging, Boosting and consensus theory obtained higher overall user and producer accuracies than traditional classifiers [4]. Debeir *et al.* integrated spectral features, spatial and contextual information and experimented Bagging and subset selection algorithms with high resolution remotely sensed images, and the accuracy was higher than neural network and decision tree classifier [21]. Kumar *et al.* designed a hierarchical fusion strategy for classifying hyperspectral remote sensing image, and the classifier fusion method outperformed other traditional approaches, including maximum likelihood classifier (MLC) *etc.* [11]. Waske *et al.* fused Support Vector Machines (SVM) for classifying SAR and multispectral imagery from agricultural areas [22]. The presented approaches are superior to all other techniques, such as a standard SVM. Ceamanos *et al.* designed a classifier ensemble for classifying hyperspectral data [15]. Firstly, hyperspectral image is divided into several parts according to the similarity of spectral bands. Then, each part is used as the input of SVM classifier. Finally, all outputs are applied to be combined together by an additional SVM classifier. The results clearly demonstrated that the proposed SVM fusion scheme outperforms a standard SVM classifier. Owing to its technical superiorities, MCS is also applied to change detection from multi-temporal remote sensing images [23,24].

As mentioned above, the classification results to be combined should be diverse. If the classification results are identical or similar, it would not improve the accuracy by combining them. Therefore, diversity is a vital requirement for the success of ensemble, demonstrated by theoretical and empirical studies [25,26]. Kuncheva and Whitaker summarized the diversity measures in classifier ensembles [27]. A special issue called “Diversity Measure in Multiple Classifier System” published in Information Fusion journal indicates that diversity measure is an important research direction in MCS [28]. Petrakos *et al.* applied agreement measure in decision fusion level combination [29]. Foody compared the different classification results from three aspects (similarity, non-inferiority, difference) using hypothesis tests and confidence interval algorithms [30]. Brown *et al.* indicated that the diversity for classification tasks is still an ill-defined concept, and defining an appropriate diversity measures for MCS is still an open question [28]. So far, there are many existing diversity measures, but in the remote sensing literatures, they are rarely used and compared for remote sensing image classification by MCS. In conclusion, a successful MCS system depends to a large extent on the proper selection of diverse classifiers for incorporation [31]. Kuncheva claims that the choice of an appropriate ensemble strategy can improve the performance of MCS [32]. More details about diversity measures and classifier ensemble approaches can be seen in Section 2.

The objective of this paper is to give a review on the uses of MCS in remote sensing and the ideas of improving MCS approaches, including classifier ensemble approaches and diversity measures, and some modified algorithms. In order to demonstrate the effectiveness of MCS, the mentioned MCS approaches are experimented with three typical optical remotely sensed images: high spatial resolution image (QuickBird image), hyperspectral image (OMISII image) and multi-spectral image (Landsat ETM+ image).

The remainder of this paper is organized as follows. In Section 2, we introduce some common approaches of MCS, summarize the advances of MCS to remote sensing image classification, and propose some improved algorithms. Experimental results are presented in Section 3. Section 4 is the conclusion of this paper.

2. Multiple Classifier System for Remote Sensing Image Classification

2.1. MCS: A Primer

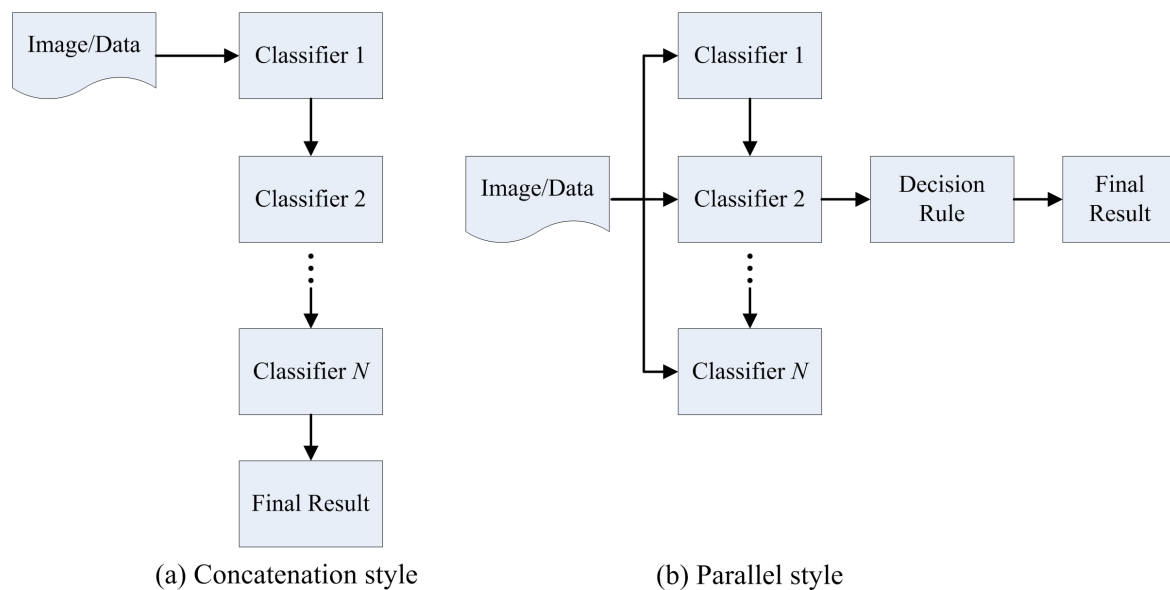
For remote sensing image classification, Giacinto *et al.* compared the performance of remote sensing classification approaches in different applications and found that no classifier could always perform well [33]. To alleviate this problem, MCS can provide the complementary information of the pattern classifiers and integrate the outputs of these pattern classifiers according to a certain combination approaches [5,34–36]. Therefore, it can outperform any individual classifier. In addition to MCS in the domain of classification, some concepts are similar to MCS, including multiple classifier combination [37,38], classifier ensemble [10,27,39], decision level combination [40,41], ensemble learning [42,43], *etc.*

We assume M classes exist on a remote sensing image, represented by $C_1 \cup C_2 \dots C_i \cup C_M, i \in \{1, 2, \dots, M\}$. For hard classification, each pixel is allocated a specific class label $J, J \in \{1, 2, \dots, M\}$. For soft classification, each pixel is depicted by a vector $\{P(1), P(2), \dots, P(M)\}$, where $P(i)$ is the probability or membership of the pixel belonging to the i^{th} class, or the abundance of the i^{th} endmember within the pixel. When multiple classifiers are combined, both hard and soft classification results can be integrated.

In order to introduce the MCS approaches clearly, three categories are adopted to summarize the classifier ensemble approaches:

- (1) Algorithms based on training sample manipulation. The most popular methods are Bagging and Boosting [16,17].
- (2) Concatenation combination. In a concatenation combination, the classification result generated by a classifier is used as the input into the next classifier. The results achieved through each classifier are similarly passed onto the next classifier until a result is obtained through the final classifier in the chain [31,44,45] (Figure 1(a)).
- (3) Parallel combination. In a parallel combination, multiple classifiers are designed independently without any mutual interaction and their outputs are combined according to certain strategies [44,45] (Figure 1(b)). The main design decision is the selection of a comprehensive ensemble approach. If the approach is well designed, the MCS can obtain the accurate result [31]. Some of the popular and successful ensemble approaches are majority voting, fuzzy integral, D-S evidence theory, *etc.*

Besides concatenation and parallel combination, hierarchical combination that combines both concatenation and parallel combination is also used [31]. But in this paper, we just focus on the concatenation and parallel combination.

Figure 1. The combination style in MCS.

As mentioned in Section 1, diversity of the classifier outputs is a vital requirement for the success of the ensemble. Diversity measures are often divided into parts: pairwise and non-pairwise [27]. Pairwise measures include kappa statistics, double fault, disagreement, *etc.*, and non-pairwise measures consist of entropy, weighted count of errors and correct results (WCEC), *etc.* [27,46]. When pairwise measures are used to calculate the diversity among more than two classification results, it adopts the average of the pairwise measures between each two classification results. The non-pairwise diversity measures consider all the classifiers together and calculate directly one diversity value for the ensemble [12]. In this paper, both pairwise and non-pairwise diversity measures, including kappa statistics, double fault, entropy, WCEC and a novel measure are reviewed.

The following sections describe the classifier ensemble approaches and diversity measures. Section 2.2 introduces Bagging and AdaBoost. Section 2.3 lists out a collection of parallel combination methods. Section 2.4 gives the basic idea of concatenation combination schemes. We proposed three modified concatenation combination approaches in Section 2.5. Section 2.6 presents five diversity measures, three of them are pairwise and the rest are non-pairwise.

2.2. Bagging and AdaBoost

Unlike statistical voting theory which is based on the assumption of independent data sources and uses all training samples only one time, Boosting and Bagging are exerted by manipulating training samples [16,17].

Bagging is the abbreviation of bootstrap aggregating. In this algorithm, n samples are selected at random from a training set with k samples, and instructive iteration is exerted to create some different bags, and each bag is classified by vote to predict its class [16]. Boosting can process data with weights, and the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples. Bagging has been shown to reduce the variance of the classification, while Boosting reduces both the variance and the bias of the classification. So in most cases, Boosting can produce more accurate

classification results than Bagging. However, the computation time of Boosting is more than Bagging, and Boosting is sensitive to noise [47]. AdaBoost is the popular used approach of Boosting algorithms. The pseudo-code of Bagging and AdaBoost can be seen in Figure 2 [16,17,47]. In addition, there is a great variety of approaches drawn upon the basic idea of Bagging and Boosting. Detailed descriptions can be found from Wagging [48], Random Forest [49], Random Subspace [50], Logistic Boosting [51], MultiBoost [52], Rotation Forest [39], Rotboost [53], *etc.*

Figure 2. The pseudo-code of Bagging and AdaBoost [16,17,47].

<hr/> <p>Input: training sample S, Classifier L, iterations I</p> <p>Output: result L_E</p> <p>Training:</p> <p style="padding-left: 20px;"><i>for</i> $i = 1$ <i>to</i> I</p> <p style="padding-left: 40px;">$S_i = \text{bootstrap sample from } S$</p> <p style="padding-left: 40px;">$L_i = \text{train a classifier on } S_i \text{ via } L$</p> <p style="padding-left: 20px;"><i>end for</i></p> <p style="padding-left: 20px;">$L_E = \arg \max_{y \in Y} \sum_{i: L_i(x)=y} 1$</p> <hr/>	<hr/> <p>Input: training sample S, Classifier L, iterations I</p> <p>Output: result L_E</p> <p>Training:</p> <p style="padding-left: 20px;"><i>normalize the weights and make the total weight is</i> m</p> <p style="padding-left: 20px;">$S_i = \text{sample from } S \text{ according to the distribution}$</p> <p style="padding-left: 20px;">$L_i = \text{train a classifier on } S_i \text{ via } L$</p> <p style="padding-left: 20px;">$e_i = \frac{1}{m} \sum_{x_i \in S_i: L_i(x_i) \neq y_i} \text{weight}(x_i)$</p> <p style="padding-left: 20px;">$\beta_i = \frac{e_i}{1 - e_i}$</p> <p style="padding-left: 20px;">$\text{weight}(x_i) = \text{weight}(x_i) \beta_i$, <i>for all</i> x_i <i>where</i> $L_i(x_i) = y_i$</p> <p style="padding-left: 20px;"><i>end for</i></p> <p style="padding-left: 20px;">$L_E = \arg \max_{y \in Y} \sum_{i: L_i(x)=y} \log(1 / \beta_i)$</p> <hr/>
---	--

(a) Bagging

(b) AdaBoost

2.3. Parallel Combination

Majority or weighted vote. The most popular ensemble approach is the majority vote by which each individual classifier “votes” for the specific class, and the class that collects the majority votes is predicted by the ensemble [12]. By generalizing this approach, Xu *et al.* proposed a threshold plurality vote, which imposed a threshold on the number of votes to select the class [44]. The major problem of majority vote is that all the classifiers have the same “authority” regardless of their respective abilities to classify properly [54]. Therefore, the weighting methods are best suited for problems where the individual classifiers perform the same task. The simple one is that the decision of each classifier is weighted according to the estimated accuracy on the training set. Furthermore, Moreno-Seco *et al.* proposed three variant of weighted voting approaches: re-scaled weighted vote (RSMV), best-worst weighted vote (BWWV) and quadratic best-worst weighted vote (QBWWV) [54]. Among all voting algorithm tested in [54], each method has shown the different performance in different datasets. In this paper, we adopted the specific class accuracy generated by the individual classifier as the weights.

Bayesian average method. Bayesian combiners are used to linearly combine the probabilistic predictions of multiple classifiers by weighting their posterior probabilities [55]. It can be defined as follows:

$$P(X \in C_i | X) = \frac{1}{k} \sum_{k=1}^N P_k(X \in C_i | X), i = 1, 2, \dots, M \quad (1)$$

Then, the class label is decided:

$$L(X) = \arg \max_{i=1, \dots, M} (P(X \in C_i | X)) \quad (2)$$

where $P_k(X \in C_i | X)$ represents the probability of class i of pixel X in k^{th} classifier, M is the total number of classes, N is the total number of classifiers. $L(X)$ represents the class label of pixel X .

Other simple operators such as *Maximum*, *Minimum* and *Product* have been applied to combine the posterior probabilities [56]. On the basis of Bayesian framework, Kittler *et al.* provided a theoretical underpinning of many existing combination schemes based on the product and sum rule, showing that the sum rule is less sensitive to the errors of subsets of base classifiers [57].

Fuzzy Integral. Fuzzy integral is an effective information fusion algorithm based on the fuzzy measure [58,59]. Sugeno integral is a widely used fuzzy integral algorithm, in which fuzzy measure g is defined as a function in measurable space: $g : \Theta \rightarrow [0, 1]$, meeting the conditions as follows:

- (1) $g(\Phi) = 0, g(S) = 1$,
- (2) If $A, B \in \Theta$, and $A \subseteq B \Rightarrow g(A) \leq g(B)$,
- (3) If $\{A_i\}_i^\infty$ is a incremental measurable set, then $g(\lim_{i \rightarrow \infty} A_i) = \lim_{i \rightarrow \infty} g(A_i)$.

Fuzzy measure g_λ was introduced in Sugeno integral.

For any $A, B \in \Theta$, $g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B)$, $\lambda > -1$.

Suppose that (S, R, g) is the fuzzy measurement space, and $f : S \rightarrow [0, 1]$ is a measurable function on R . When $S = s_1, s_2, \dots, s_N$ is a finite set and meets the condition of $1 \geq f(s_1) \geq f(s_2) \dots \geq f(s_N) \geq 0$, the fuzzy measurement can be calculated as the follows:

$$g(A_1) = g(\{s_1\}) \quad (3)$$

$$g(A_i) = g(\{s_1\}) + g(A_{i-1}) + \lambda g(\{s_1\})g(A_{i-1}), 1 < i \leq N \quad (4)$$

where $A_i = \{s_1, s_2, \dots, s_i\}$ is the classifier ensemble, and N is the amounts of classifiers.

Sugeno integral is calculated in the following:

$$E_j = \int_A f_j(s_i)g(A_i) = \sup_{i \in [1, m]} \min \{f_j(s_i), g(A_i)\} \quad (5)$$

When the member classifiers are more than two, the values of λ corresponding to each class is calculated as the following equation:

$$\lambda + 1 = \prod_{i=1}^N (1 + g(\{s_i\})) \quad (6)$$

In multiple classifier combination, $f_j(s_i)$ is defined as the probability of the i^{th} classifier to the j^{th} class in soft classification, or the accuracy of the i^{th} classifier to the j^{th} class in hard classification. We utilized the latter one, hard classification in this paper.

The final classification result is the class label corresponding to the maximum fuzzy integral value:

$$L = \arg \max_{j=1, \dots, M} (E_j) \quad (7)$$

Moreover, Ben Adballah *et al.* proposed a local fuzzy integral approach, called context extraction for local fusion with fuzzy integral (CELF-FI) for fusing different classifiers outputs [60]. This method based on a novel objective function that combines context identification and multi-algorithm fusion criteria into a joint objective function. This objective function consists of two terms: the first is designed to produce compact clusters, called contexts, and the second to produce Sugeno measures for fuzzy integral fusion for each context [60].

Consensus Theory. Consensus theory is suitable for the combination of classifiers with probability to specific classes as their outputs [18,19]. Usually logarithmic and linear consensus models are used in the following two ways:

$$T_j(X) = \prod_{i=1}^N (p_i(C_j|X))^{\lambda_{ij}} \quad (8)$$

$$T_j(X) = \sum_{i=1}^N (p_i(C_j|X))^{\lambda_{ij}} \quad (9)$$

Each classifier is viewed as an expert with the membership or probability of the pixel belonging to every class as its output. $T_j(X)$ is the probability the pixel X being assigned to the j^{th} class, and $p_i(C_j|X)$ is the probability of pixel X being assigned to the j^{th} class by the i^{th} classifier, and λ_{ij} is the weight. The simplest approach of the weighting scheme consists in giving all the individual outputs equal weights. Furthermore, the weights can be defined to not only weight the individual sources but also the individual classifiers [5]. Thus, we select the specific class accuracy as the weights. It means that λ_{ij} represents classification accuracy of the i^{th} classifier to the j^{th} class in this paper.

The final classification result is the class label corresponding to the maximum value of $T_j(X)$, $j = 1, \dots, M$:

$$L(X) = \arg \max_{j=1, \dots, M} (T_j(X)) \quad (10)$$

D-S Evidence Theory and its improvement. D-S evidence theory, which assigns probability to the sets, is able to handle the uncertainty caused by unknown factors [12]. It uses discrimination framework, confidence function, likelihood function and probability allocation function to represent and process knowledge. Suppose that $\Theta = C_1, C_2, \dots, C_i \dots C_M$ is the discrimination framework and M is the number of classes, therefore basic probability allocation function m is a function from 2^Θ to $[0, 1]$ and it meets the requirement of:

$$\begin{cases} m(\phi) = 0 \\ \sum_{A \subseteq \Theta} m(A) = 1 \end{cases} \quad (11)$$

If there are two or more different evidences, orthogonal sum can be used to combine those evidences. Assume that Z_1, Z_2, \dots, Z_n are the probability allocation functions corresponding to evidence F_1, F_2, \dots, F_n , and their orthogonal sum $Z = Z_1 \oplus Z_2 \dots \oplus Z_n$ is:

$$Z(\phi) = 0 \quad (12)$$

$$Z(A) = \frac{1}{K} \sum_{\cap A_i} \prod_{1 \leq i \leq n} Z_i(A_i) \quad (13)$$

$$K = \sum_{\cap A_i \neq \phi} \prod_{1 \leq i \leq n} Z_i(A_i) \quad (14)$$

When various evidences are inconsistent or contradictory from each other, the result may be unreasonable. In order to alleviate this problem, an improved evidence theory algorithm was proposed and experimented in [61], and it proved that the modified method was superior to traditional method while the evidences with high contradiction and inconsistency. For remote sensing image, different classifiers may generate different classified labels, resulting in the generation of evidence with high contradiction, so the improved evidence theory ensemble approach may improve the performance. The detailed equations are detailed as follows [61]:

$$Z(A) = \sum_{A_i \in F_i} Z_i(A_i) Z_2(A_2) \dots Z_n(A_n) + K \times \epsilon \times \frac{1}{n} \sum_{i=1}^n Z_i(A) \quad (15)$$

where:

$$\tilde{k} = \frac{2}{n(n-1)} \sum_{i < j} \sum_{A_i \cap A_j = \Phi, A_i \in F_i, A_j \in F_j} Z_i(A_i) Z_j(A_j) \quad (16)$$

$$\epsilon = \exp^{-\tilde{k}} \quad (17)$$

where ϵ is the confidence of evidence, \tilde{k} is the average of contradiction level between two evidences, and K is the total contradiction level of all evidences. This evidence combination method can reduce the limitations caused by evidence inconsistency. For multiple classifier combination of remote sensing, the result of each classifier can be viewed as a piece of evidence. Probability allocation function can be represented by the classification accuracy of specific class. For example, if a pixel belongs to the i^{th} class in a classifier, the basic probability is $m(C_i) = P_i$, $m(\Theta) = 1 - P_i$, where P_i is the accuracy of the i^{th} class by the specific classifier. After the completion of evidence combination, the class with maximum evidence is selected as the final result.

Dynamic Classifier Selection [9,62]. Smits applied dynamic classifier selection and improved models to multiple classifier combination for remote sensing images [9]. In this method, MCS are addressed that use estimates of each individual classifier's local accuracy in small regions of feature space surrounding an unknown test sample [9]. The local regions are defined by the k -nearest neighbors (KNN). The drawback of this approach is the selection of k , which is rather heuristic and subjective an exercise, and the classification result is sensitive to the choice of k [9,63,64].

The concept of local accuracy assumes that classes maintain a certain continuity in the feature space and neighboring pixels are expected to maintain a stronger relationship than pixels that are further away [9]. Dudani suggested that using distance-weighted k -NN can improve performance, especially with small training sample [65]. Suppose that d_k, d_t, d_m are the distances of the k^{th} , t^{th} and m^{th} samples to the pixel in the decreasing sort, the weight of the t^{th} sample is:

$$\omega_t = \begin{cases} \frac{d_k - d_t}{d_k - d_m} & d_k \neq d_m \\ 1 & \text{otherwise} \end{cases} \quad (18)$$

The local classification accuracy of the j^{th} classifier is:

$$L^j = \frac{1}{k} \cdot \sum_{\theta(X_t)=C_j(X_t)} \omega_t, \quad 1 \leq t \leq k \quad (19)$$

where $\theta(X_t)$ is the class label of the t^{th} neighboring sample, $C_j(X_t)$ is the assigned class label of the j^{th} classifier to the t^{th} sample. Finally, the label of the classifier with the biggest local accuracy is selected as the final classification result.

Differences between these methods: In addition, according to the individual output, classifier ensemble approach in a parallel combination can be divided into three levels: abstract level, rank level and measurement level [66]. The abstract level combination methods are applied when each classifiers outputs a unique label [66]. Rank level makes use of a ranked list of classes where ranking is based on decreasing likelihood. In the measurement level, probability values of the classes provided by the each classifier are used in the combination. Therefore, majority/weighted vote, fuzzy integral, evidence theory and dynamic classifier selection belong to the abstract level combination methods, while others are measurement level methods. Table 1 summarizes the classifier ensemble approaches in parallel combination. Weighted vote, fuzzy integral, Dempster–Shafer evidence theory and consensus theory require validation set to calculate the weights while other methods do not. Dynamic classifier selection calculates the distance between the samples so it requires the original image. And the computation time of dynamic classifier selection is more expensive than other approaches.

Table 1. Summary of the classifier ensemble approaches in parallel combination.

Name	hard labels	soft labels	validation set
Majority vote	Y	N	N
Weighted vote	Y	N	Y
Bayesian average	N	Y	N
Dempster-shafer evidence theory	Y	N	Y
Fuzzy integral	Y	N	Y
Consensus theory	Y	Y	Y
Dynamic classifier selection	Y	N	Y

* Note: “Y” and “N” mean whether or not the hard labels, soft labels or validation set are needed.

Dynamic classifier selection method needs the original image to calculate the distance.

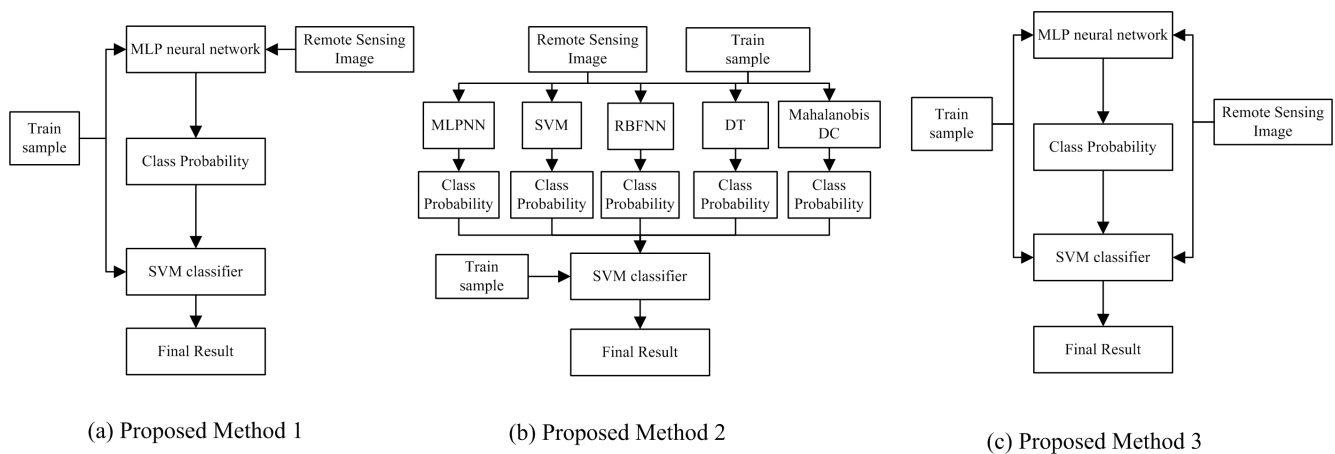
2.4. Concatenation Combination

In the original concatenation combination approach, the classification result generated by a classifier is used as an input into the next classifier [31]. The results obtained through each classifier are similarly passed onto the next classifier until a result is obtained through the final classifier in the chain, which can be clearly shown in Figure 1(a). In this structure, decisions could be first made by a subset of the committee classifiers and their conclusions are then passed to another subset of classifiers. The construction of this algorithm is usually required to improve the accuracy of the entire system and to incorporate error checks [31]. Two potential ways to improve concatenation combination are enhancing the input diversity of the final combiner, and selecting an advanced combination algorithm.

2.5. Modified Concatenation Combination Approaches

On the basis of the previous improvement ideas, we propose three modified concatenation combination approaches of MCS in this section. The overall architecture is given in Figure 3. All the proposed schemes contain two training stages. Among the three methods, SVM classifier is selected as the second classifier because of its satisfactory capability to deal with the classification problem [67,68].

Figure 3. The structure of concatenation combination.



Proposed method (1) is the original concatenation combination approach. Multi-layer perceptron neural network (MLPNN) is assigned as the first level classifier. The class probabilities generated by MLPNN are defined as the input of the second level. In order to enhance the diversity in concatenation combination approach, the class probabilities generated by the several classifiers are combined as the input in method (2). The class probabilities generated by MLPNN, Radial Basis Function neural network (RBFNN), C4.5 decision tree (DT), Minimum Mahalanobis distance classifier (MMDC) and SVM are treated as the input to train the new SVM classifier. Similar to method (2), method (3) adopts the input derived from the different source to enhance the diversity. Original remote sensing image and the class probabilities produced by MLPNN are defined as the input of the second level in method (3).

2.6. Diversity Measures

Kappa statistics, double fault, entropy and WCEC are used in this paper, and a novel measure is proposed for remote sensing based on the improvements in general MCS.

Kappa statistic. This measure is calculated by the following equation [29]:

$$k = \frac{2(N^{11}N^{00} - N^{10}N^{01})}{p_1q_2 + p_2q_1} \quad (20)$$

where N^{11} represents the samples correctly classified by both classifiers, N^{10} is the samples correctly classified by the first classifier and incorrectly by the second classifier, N^{01} represents the samples incorrectly classified by the first classifier and correctly by the second classifier, N^{00} is the samples incorrectly classified by both classifier. p_1 represents the samples correctly classified by the first classifier (it equals N^{11} plus N^{10}), p_2 represents the samples correctly classified by the second classifier (it equals

N^{11} plus N^{01}), q_1 represents the samples incorrectly classified by the first classifier (it equals N^{01} plus N^{00}), q_2 represents the samples incorrectly classified by the second classifier (it equals N^{10} plus N^{00}).

It not only gives a measure of the degree of agreement but also has a test associated with itself to check if the apparent agreement can be attributed to chance only. In addition, we also use a new kappa statistic k' and the final value is the average of the class' kappa statistic.

Double fault: Gacinto and Roli proposed the double-fault measure to select classifiers that are least related from a pool of classifiers. This diversity decreases when the value of the double-fault measure increases [55]. This index is calculated by:

$$DF = \frac{N^{00}}{N} \quad (21)$$

where N represents the number of sample.

Entropy: Entropy is calculated in the following [27]:

$$E = \frac{1}{N} \sum_{m=1}^N \frac{1}{L - \lceil L/2 \rceil} \min \{l(\mathbf{z}_m), L - l(\mathbf{z}_m)\} \quad (22)$$

where L is the number of classifiers, N is the number of sample, $\lceil \cdot \rceil$ rounds a number to the nearest integer. $l(\mathbf{z}_m)$ is the number of samples correctly classified by the classifiers \mathbf{z} .

This measure is based on the assumption: if the results of $\lceil L/2 \rceil$ is the same, the diversity has the biggest value. The range of this measure is 0 to 1, in which 1 represents the biggest diversity value and 0 represents the lowest diversity value.

Weighted count of errors and correct (WCEC). *WCEC* is also proposed by [46]. Compared to kappa statistics, WCEC not only gives a measure of the correctness of classifiers but also consider the diversity between two classifiers.

$$WCEC = \frac{N^{11} + \frac{1}{2}(N^{01} + N^{10}) - N_{diff}^{00} - 5N_{same}^{00}}{N} \quad (23)$$

where N_{diff}^{00} represents the different errors occurring in both classifiers, N_{same}^{00} represents the same errors occurring in both classifiers.

Disagreement-accuracy measure: Combining similar classifiers does not make too much sense as the results will not be improved [31]. If the sample is classified correctly or incorrectly in the classifiers, the final result is not improved. In order to consider both the classification accuracy and disagreement among the classification results comprehensively, we propose a new diversity measure, namely, disagreement and accuracy measure (*DA*). Firstly, we have made the classification results in descending order according to the overall accuracy. Then *DA* was calculated by the contribution of the classifiers to each sample. It can be defined as the following:

$$DA = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^L \frac{P_i(\mathbf{z}_j)}{2^{j-1}} \quad (24)$$

where N is total number of samples, L is the number of classifiers, $P_i(\mathbf{z}_j)$ is the specific class accuracy of correctly classified sample i by the classifier \mathbf{z}_j . This measure highlighted the contribution of classifier with high overall accuracy to the sample. That means if a sample is classified correctly by the classifier which has the highest overall accuracy, the contribution of the subsequent classifier to this sample should be decreased.

3. Experiments and Discussions

In order to evaluate the effectiveness and performance of MCS to remote sensing image classification, the aforementioned existing and modified methods are experimented in the three types of remote sensing images: high spatial resolution QuickBird image, airborne hyperspectral image, and medium-resolution multispectral Landsat ETM+ image.

The classifiers used in an ensemble should generally be accurate but different. Based on the survey to existing pixel-based classification algorithms by [3] and many suggestions on the selection of classifiers by [7], we have chosen nine classifiers to be used, including Maximum Likelihood Classifier (MLC), Minimum Euclidean Distance Classifier (MEDC), Minimum Mahalanobis distance classifier (MMDC), support vector machine (SVM), MLP neural network (MLPNN), RBF neural network (RBFNN), spectral angle mapper (SAM), J48(C4.5) decision tree classifier (DTC) and simplified fuzzy H-ARTMAP (SFH-ARTMAP). The descriptions of classifiers can be detailed in [68–75].

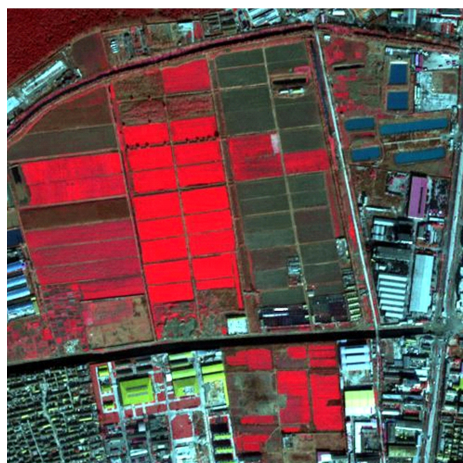
MLC, MDC, SAM, SVM in ENVI 4.5 and Bagging, AdaBoost, J48DT, MLPNN, RBFNN in Weka 3.6.3 software are used, while SFH-AREMAP is implemented in MATLAB 6.5 software [72]. The other classifier ensemble approaches and diversity measures are all implemented in ENVI+IDL software.

In this paper, we mainly focus on the final results of MCS, so we adopt the default parameters in the above classifiers. The iterations of AdaBoost and Bagging are 10. The different classifier ensemble algorithms are experimented in the three typical images, but the diversity measures are only experimented on QuickBird image.

3.1. Experiment 1: High Spatial Resolution Image Classification

QuickBird multispectral image (four bands of multi-spectral, spatial resolution: 2.44 m) is used as the example of high spatial resolution remote sensing image. The image size is 500×500 . The study area is located in the northern suburb of Xuzhou city along the urban and rural connecting areas. Through the image analysis and field works, the image is classified into five land cover types: water, built-up area, green areas, vegetation and barren soil. Figure 4 is the false color composite image (R: Band 4, G: Band 3, B: Band 2) of the study area.

Figure 4. False color composite of QuickBird remote sensing image.



In addition to original multispectral image, textural features extracted from the Band 4 which has the richest information are also added for classification. Mean and variance extracted by gray level co-occurrence matrix (GLCM) are used to describe the textural features. Two schemes are used for classification. The first scheme is only using the spectral features of four multispectral bands as the inputs, and the second scheme is using both four spectral features and two textural features as the inputs. Seven classifiers are used, including: MMDC, MLC, SVM, MLPNN, RBFNN, J48 DT, and SFH-ARTMAP. The accuracies of all classifiers using only spectral features and spectral together with textural features are listed in Tables 2 and 3. The classification result can be seen in Figures 5 and 6.

Table 2. Classification accuracy of the first Scheme (only spectral features).

Classifier	Water	Built-up Area	Green Space	Vegetation	Barren Soil	OA	Kappa
MMDC	96.81%	48.48%	92.18%	95.64%	97.68%	84.41%	0.81
MLC	95.17%	75.00%	96.54%	92.01%	98.01%	90.80%	0.88
SVM	87.56%	97.51%	95.73%	89.05%	92.77%	92.59%	0.91
MLPNN	92.12%	78.04%	96.98%	95.04%	97.80%	91.44%	0.89
RBFNN	94.23%	80.59%	95.25%	87.84%	96.10%	90.65%	0.88
J48 DTC	93.85%	89.19%	89.19%	86.75%	92.78%	91.92%	0.90
SFH-ARTMAP	92.77%	77.85%	97.25%	91.23%	96.39%	90.70%	0.88

Table 3. Classification accuracy of the second scheme (spectral and textural features).

Classifier	Water	Built-up Area	Green Space	Vegetation	Barren Soil	OA	Kappa
MMDC	88.08%	56.28%	93.61%	97.16%	96.76%	85.92%	0.82
MLC	95.59%	83.14%	97.11%	91.23%	98.34%	92.78%	0.91
SVM	90.56%	98.62%	97.39%	84.33%	94.51%	93.41%	0.92
MLPNN	97.60 %	86.76%	95.12%	88.87%	98.84%	93.49%	0.92
RBFNN	94.93%	88.66%	93.34%	89.35%	97.80%	92.83%	0.91
J48 DTC	92.82%	93.23%	95.69%	86.45%	96.51%	92.42%	0.90
SFH-ARTMAP	93.34%	87.29%	97.56 %	90.87%	98.84%	93.15%	0.91

Figure 5. Classification results of using only spectral features.

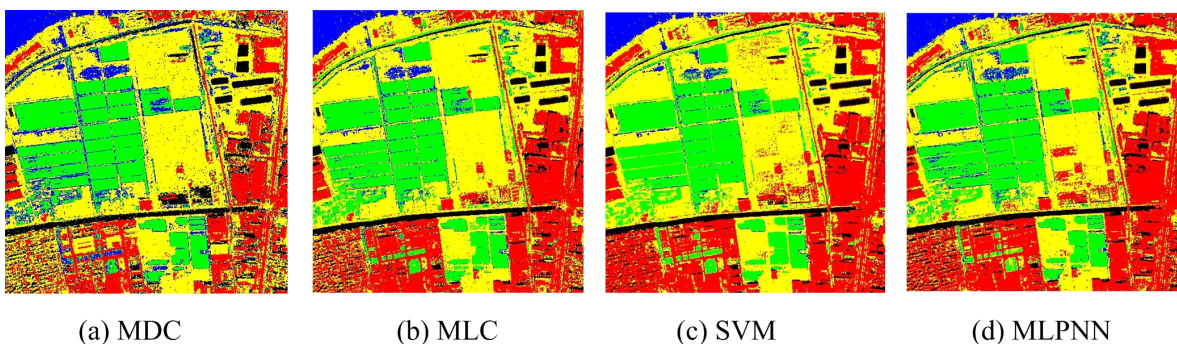


Figure 5. Cont.

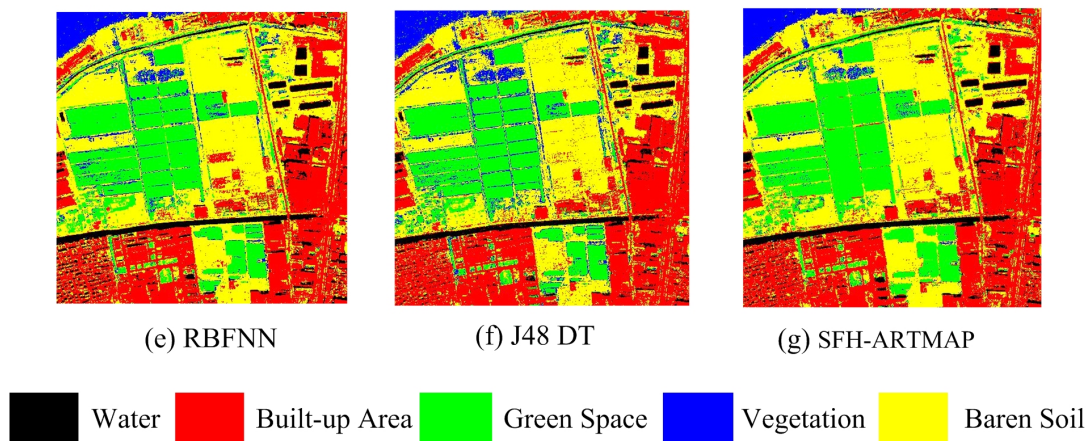
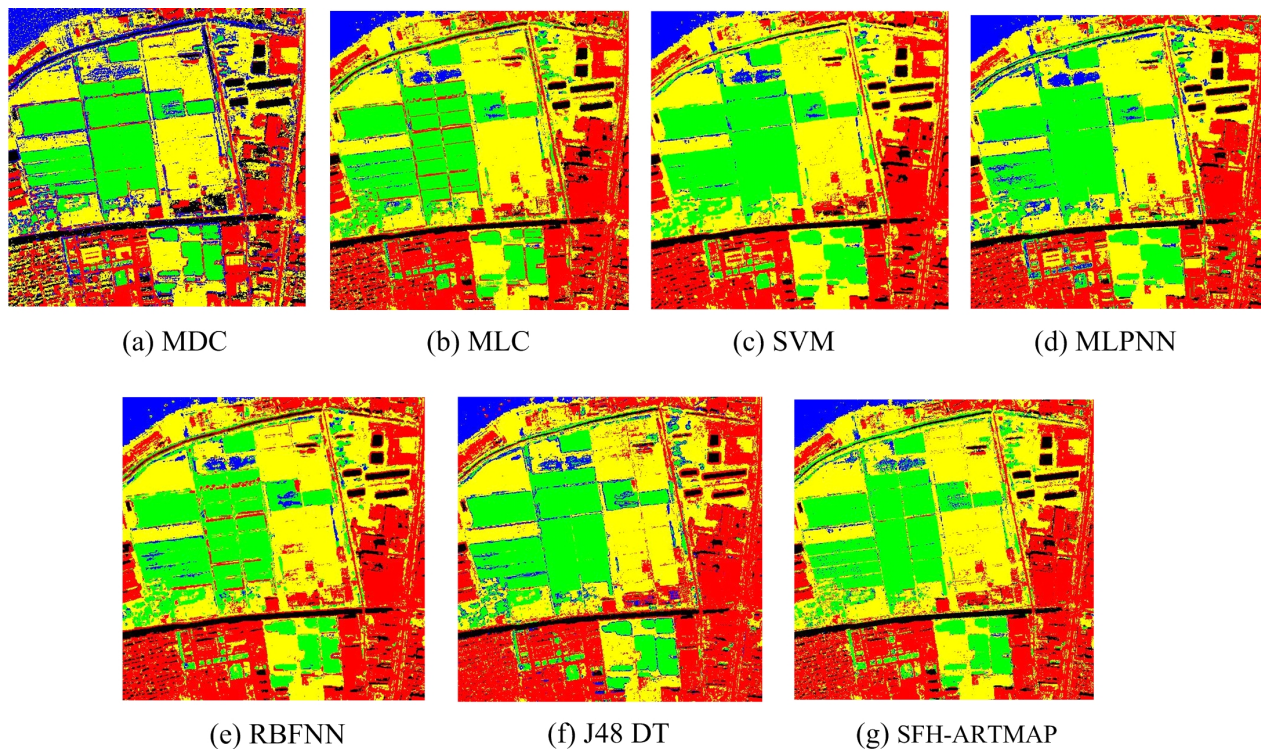


Figure 6. Classification results of using both spectral and textural features.



Comparing the test areas, the accuracy assessment demonstrates that all the classifiers perform very accurately and achieve overall accuracies of 85.92% (MMDC), 92.78% (MLC), 93.41% (SVM), 93.49% (MLP), 92.83% (RBFNN), 92.42% (J48 DTC) and 93.15% (SFH-ARTMAP) using spectral and texture features (seen Table 3). As expected, the kappa values of the classification results using spectral and texture features are slightly higher than the ones only using spectral features. The specific accuracies of built-up areas of MMDC, MLC, MLPNN and SFH-ARTMAP are significantly lower (less than 79%). Using spectral and texture features together, most obvious is that the accuracies of built-up areas are increased by up to 7.8%, 8.14%, 8.72%, 9.44% of MMDC, MLC, MLPNN and SFH-ARTMAP, respectively. Furthermore, some other findings can be summarized as follows:

- (1) When textural features are introduced as the input of the classifiers, both the overall and class accuracies are enhanced in the most cases, especially for built-up areas, demonstrating the superiority of incorporating textural features in the classification process.
- (2) The classifiers have shown the different performance on the specific classes, indicating that the classifier performing well for one class may be poor for other classes. For instance, when only using spectral feature, MMDC performs weaker on the land cover type of “built-up areas” (48.48%) but SVM achieve the accuracy of 97.51% (Table 2). The situation for spectral and texture features is similar to the one for spectral feature. The highlights in Tables 2 and 3 are the highest accuracy of class among the classifiers. From the above analysis, it is necessary to combine multiple classifiers to find a better result than any individual classifiers. In the following experiments on QuickBird image, the classification results derived from both spectral and texture features are adopted as the individual classification results.

For Bagging and AdaBoost, J48 DTC, RBFNN and MLPNN are used as the base classifiers, and the accuracies are summarized in Table 4. Both Bagging and AdaBoost obtain very accurately results and achieve very high overall accuracies (94.93% and 95.67% for J48 DTC, 95.09% and 95.43% for MLPNN, 93.36% and 95.35% for RBFNN). On the other hand, the accuracies of Bagging and AdaBoost increased by 0.89% to 1.96% and 1.94% to 2.25% for the three base classifiers. The classification maps can be seen in Figures 7–9. The theories of Bagging illustrate that it shows excellent performance when the base classifier is an unstable classifier such as decision tree, neural networks *et al.* [16]. In most cases, the performance of AdaBoost is superior to Bagging. Our experimental results are compatible with the theoretical analysis. In addition, we also studied the influence of bootstrapped sample size to classification accuracy. Here, we choose four sample sizes: 25%, 50%, 75% and 100%. The general trend is that the overall accuracies increase slightly (less than 1%) when the sampling rate rise, although the accuracy of RBFNN is a bit unsteady under different sampling rate. Perhaps the reason is that the iterations of RBFNN are not enough (only 10 in the experiment). The results indicate that Bagging and AdaBoost outperform the base classifier in terms of overall accuracy.

Table 4. Accuracy of AdaBoost and Bagging using different base classifier.

Base classifier	J48 DTC	MLPNN	RBFNN
Base classifier	92.42%	93.49%	92.83%
Bagging(25% samples)	94.22%	94.20%	93.76%
Bagging(50% samples)	94.55%	95.27%	93.39%
Bagging(75% samples)	94.30%	94.91%	93.40%
Bagging(100% samples)	94.93%	95.09%	93.36%
AdaBoost	95.67%	95.43%	95.35%

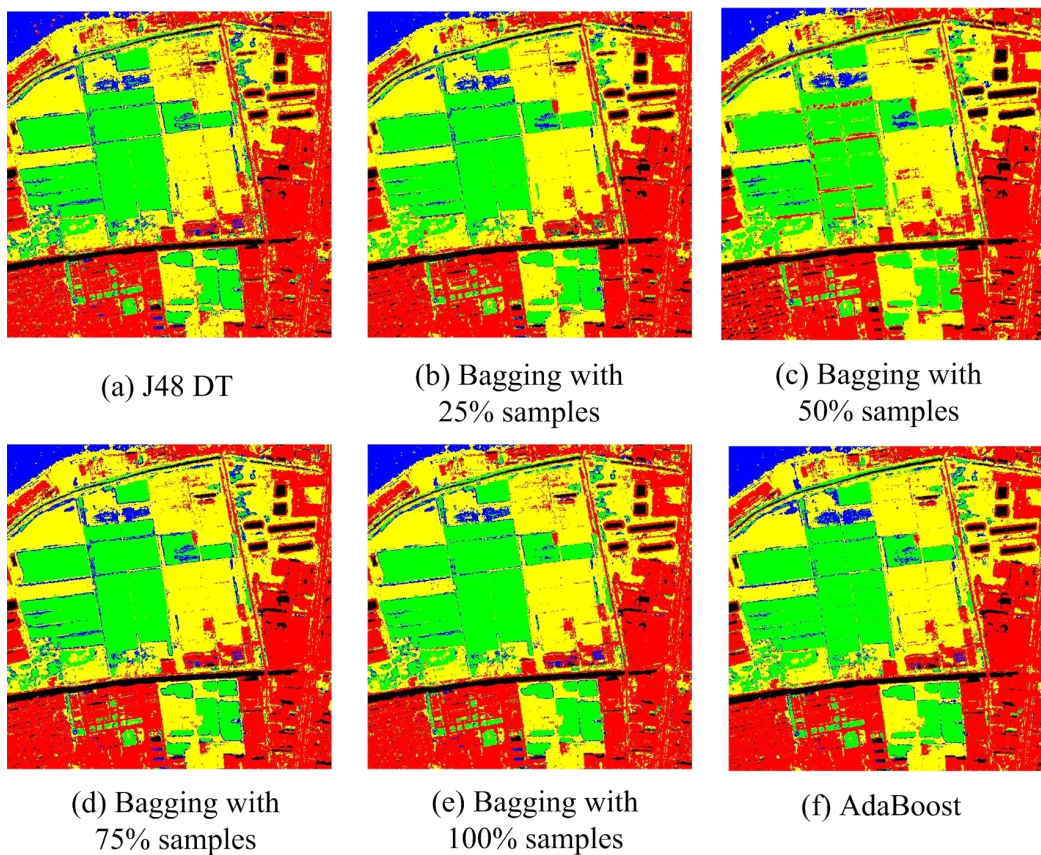
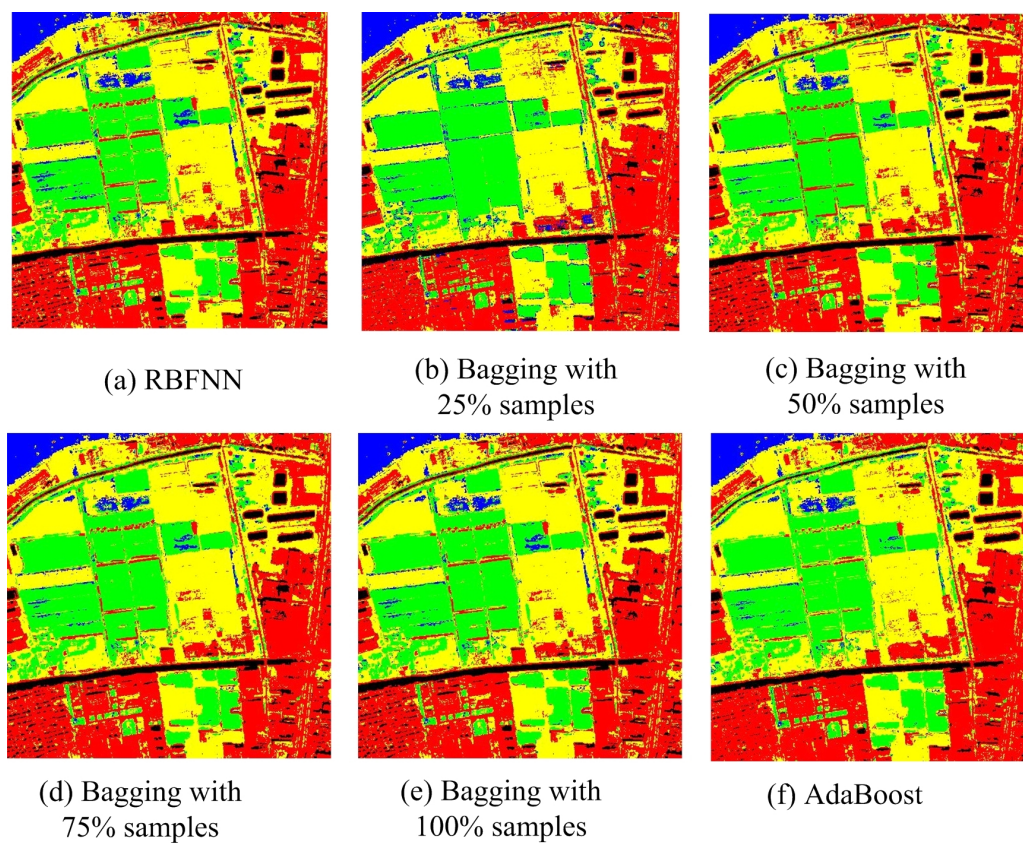
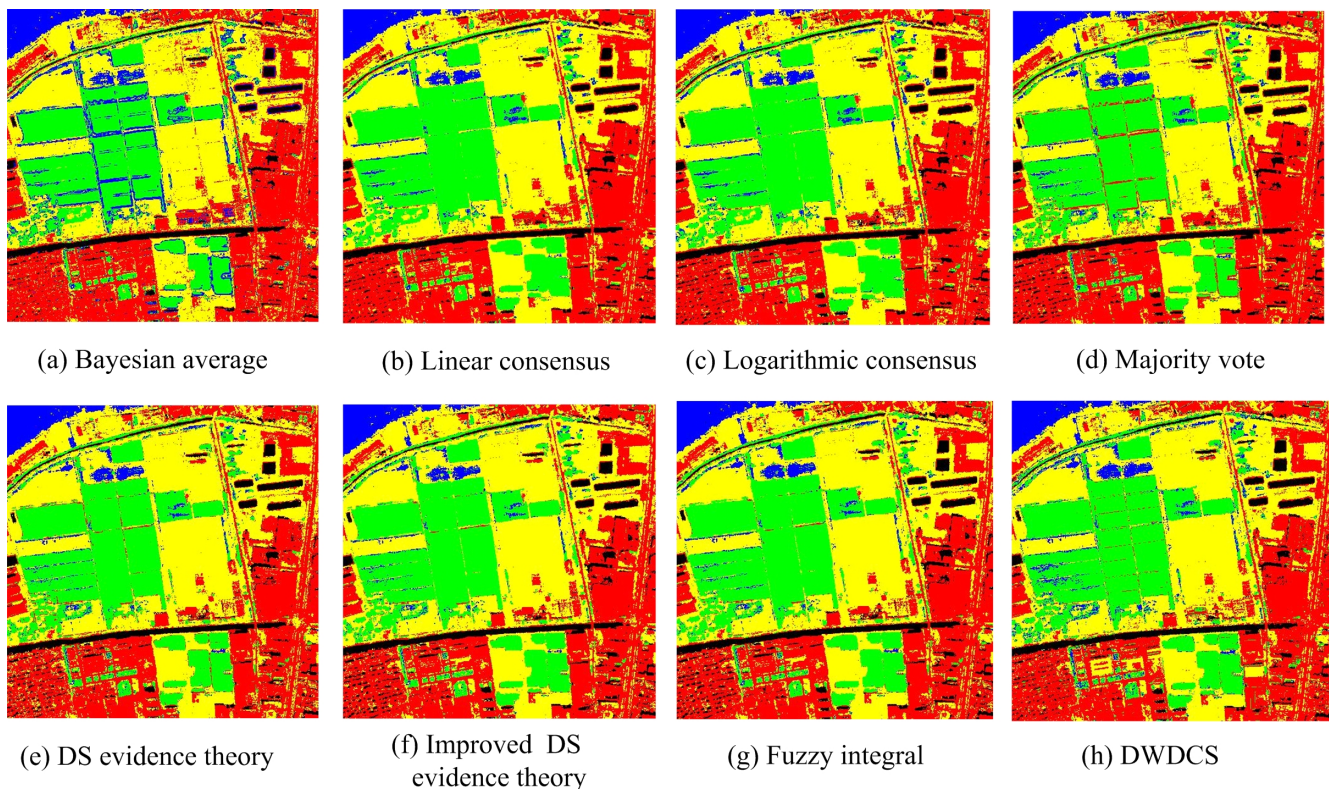
Figure 7. Classifications of J48 DTC Based on Bagging and AdaBoost.**Figure 8.** Classifications of RBFNN Based on Bagging and AdaBoost.

Figure 9. Classification results based on parallel combination approaches.

In order to illustrate the effectiveness and universality of MCS, we choose different member classifiers applied to different parallel combination rules. For the parallel combination methods such as Bayesian average, logarithmic consensus and linear consensus, the member classifiers are MLC, SVM, MLPNN, RBFNN and SFH-ARTMAP. The member classifiers for others are SVM, MLPNN, RBFNN, MEDC and J48DT. Figure 9 presents the classification results of MCS on parallel combination, and their accuracy statistics are shown in Table 5. Among the individual classifiers, SVM classifier achieves the highest accuracy of 93.49%. Compared to SVM classifier, Bayesian average achieves the accuracy of 94.99%, with the improvement of 1.5%; Majority vote and fuzzy integral improve the accuracy from 93.49% to 94.74% and 93.49% to 94.37%, respectively. Other parallel combination approaches improve the accuracy slightly with the improvement of 0.28% to 1%. The “local” ensemble method (DWDCS) results in 93.98% accuracy with the improvement of 0.49%. Compared with the overall accuracies, the kappa coefficients show similar characteristics. Bayesian average obtains the highest accuracy for all parallel combination schemes.

Table 5. Classification accuracy of parallel classifiers combination.

Combination strategy	OA	Kappa
Bayesian average	94.99%	0.94
Logarithmic consensus	93.77%	0.92
Linear consensus	94.37%	0.93
Majority vote	94.74%	0.93

Table 5. *Cont.*

Combination strategy	OA	Kappa
Evidence theory	93.76%	0.92
Improved evidence theory	94.18%	0.93
Fuzzy integral	94.37%	0.93
DWDCS	93.98%	0.92

Unlike other combination algorithms, DWDCS pays attention to the local optimum while other combination algorithms achieve the global optimum. The computation time of DWDCS is proportional to the test sample size and the nearest neighbors. In our experiments, the test sample size is more than 500 and the nearest neighbor is selected to be 15. The computation time of DWDCS is more than 10 min and other parallel combination strategies are less than 20 s.

Figure 10 presents the classification result of MCS based on concatenation combination and Table 6 shows the accuracy statistics. The accuracy assessment indicated that all the proposed concatenation combination algorithms perform very accurately and obtain overall accuracies of 95.21% (method 1), 97.06% (method 2) and 95.05% (method 3). Among all the classifier ensemble approaches in MCS, the proposed method 2 obtains the highest accuracy (97.06%) due to the following reasons:

- (1) This proposed method adopts the probabilities as the input of the second level and there is obvious diversity among these probabilities;
- (2) The output distributions of different classifier are different, for instance, the probability distribution of SVM and MLP are focused on 0 or 1, while the probability distribution of MEDC is ranged from 0.45 to 0.55. The linear combination rules such as linear consensus are not suited, and SVM can tackle non-linear problems to a good extent. Therefore, it can achieve the highest accuracy compared with others.

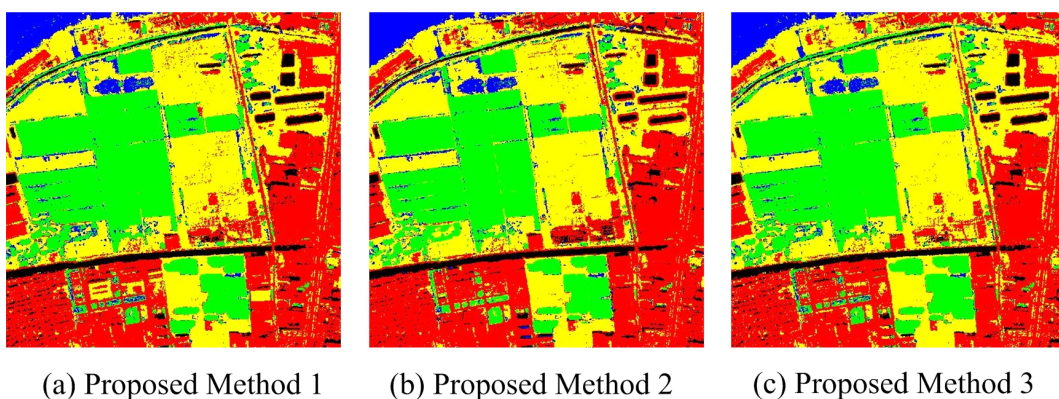
Figure 10. Classification results based on proposed concatenation combination algorithms.

Table 6. Classification accuracies of concatenation combination algorithms.

Concatenation combination	OA	Kappa
Proposed method 1	95.21%	0.94
Proposed method 2	97.06%	0.96
Proposed method 3	95.05%	0.94

In the experiment of diversity measures, we choose five member classifiers from the classifier pool to be combined based on diversity measure. The classifiers in the pool are MLPNN (1), SVM (2), SFH-ARTMAP (3), RBFNN (4), MLC (5), J48DT (6), MMDC (7). Thus, there are twenty-one selection schemes of five classifiers. We have calculated the diversity of the all combination. The optimum combination corresponding to the diversity measure can be seen in Table 7. Then, we adopted the different parallel combination approaches (majority voting, weighted voting, fuzzy integral, DS evidence theory) to evaluate the performance of the combination of classifiers selected by diversity measures.

Table 7. The optimum combination selected by diversity measures.

Diversity		Classifiers combination	Majority vote	Weighted majority vote	Fuzzy integral	DS Evidence theory
k		1,2,3,6,7	94.83%	93.54%	93.77%	93.55%
k'		1,2,3,6,7	94.83%	93.54%	93.77%	93.55%
DF		1,3,4,6,7	95.05%	94.79%	94.76%	94.86%
Entropy		1,2,3,6,7	94.83%	93.54%	93.77%	93.55%
WCEC		1,2,3,6,7	94.83%	93.54%	93.77%	93.55%
DA	1	1,3,4,5,6	95.52%	95.39%	95.40%	95.36%
	2	1,3,4,6,7	95.05%	94.79%	94.76%	94.86%
	3	1,3,4,5,6	95.32%	95.15%	95.22%	95.20%

From Table 7, k , k' , *entropy* and *WCEC* diversity measures selected the same combination, which contains the classifiers: MLPNN, SVM, SFH-ARTMAP, J48DT, MMDC. *DF* selected the combination of MLPNN, SFH-ARTMAP, RBFNN, J48 DTC and MMDC. MLPNN, SFH-ARTMAP, RBFNN, MLC, J48DTC and MMDC were chosen by *DA*. The overall accuracies of the combination selected by *DA* are 95.52% (majority vote), 95.39% (weighted vote), 95.40% (fuzzy integral) and 95.36% (DS evidence theory). In contrast, the combination selected by *DF* results in the accuracies of 95.05% (majority vote), 94.79% (weighted vote), 94.76% (fuzzy integral) and 94.86% (DS evidence theory). The combination selected by other four measures achieves the accuracies of 94.83% (majority vote), 93.54% (weighted vote), 93.77% (fuzzy integral) and 93.55% (DS evidence theory). Comparison of these results with the accuracy assessment indicates that the combination selected by *DA* outperform the ones selected by other diversity measures in terms of overall accuracy. It demonstrates that the proposed diversity measure is a positive guidance for classifier combination. Furthermore, we ranked all the combination according

to the *DA* measure in Table 8 and the highlights are the first three ones, whose accuracy produced by the parallel combination approaches can be seen in Table 7.

Table 8. Results of DP measure of all five classifiers' combination.

Classifiers combination	DP	Rank	Classifiers combination	DP	Rank
1,2,3,4,5	1.68520	10	1,3,4,5,7	1.74182	6
1,2,3,5,6	1.69234	8	1,3,5,6,7	1.75065	3
1,2,3,5,7	1.67308	16	1,3,4,6,7	1.75104	2
1,2,3,4,6	1.69413	7	1,4,5,6,7	1.74351	5
1,2,3,4,7	1.67177	18	2,3,4,5,6	1.68378	11
1,2,3,6,7	1.68283	12	2,3,4,5,7	1.66168	21
1,2,4,5,6	1.68565	9	2,3,5,6,7	1.67331	15
1,2,4,5,7	1.66347	19	2,3,4,6,7	1.67285	17
1,2,5,6,7	1.67524	13	2,3,5,6,7	1.66311	20
1,2,4,6,7	1.67524	13	3,4,5,6,7	1.74569	4
1,3,4,5,6	1.75277	1			

3.2. Experiment 2: OMISII Hyperspectral Remote Sensing Image

Original airborne OMISII hyperspectral remote sensing image has 64 bands with spectral range: 0.45–1.09 μm . The spectral resolution is 10 nm . In this paper, Zhongguangcun, a high-tech zone of Beijing City is chosen as the study area. The five noisy bands are removed so it is remaining 59 bands for classification. The image size is 400×400 . Figure 11(a) is the false color composite of the image by using Band 27, 25 and 2 as R, G and B components. Training samples and test samples are selected independently from the image. The land cover is classified into five classes: water, building, vegetation, forest and bare soil. From Table 9, SVM classifier has the highest classification accuracy of forest (92.69%) and bare soil (92.97%); MLPNN classifier produces the highest classification accuracy of vegetation (98.10%) and building (98.34%); RBFNN generates the highest classification accuracy of water (95.72%). These results indicate the diversity among classifiers, which is similar to the experiment on QuickBird image. MLPNN achieves the highest overall accuracy of 92.53%. Figure 11(b–e) shows the classification results of SVM, J48 DTC, RBFNN and MLPNN respectively. Based on diversity measure among all classifier combination schemes, MMDC, MLPNN and SVM classifier are selected to generate the final classification.

From Table 10, both AdaBoost and Bagging can enhance the classification accuracy of the base classifier. Bagging can improve the accuracy from 87.86% to 90.44% for J48 DTC, 89.46% to 89.88% for RBFNN and 92.53% to 93.09% for MLPNN. In contrast, AdaBoost performs a bit better (91% for J48 DTC, 90.65% for RBFNN and 93.27 for MLPNN) than bagging when the same base classifier is used. Figure 12 are the classification results based on AdaBoost and Bagging using RBFNN, MLPNN, J48 DTC as the base classifiers. In classification ensemble based on parallel/concatenation combination, both improved D-S evidence theory and concatenation combination (proposed method 1) can obtain

very accurately classification result with the accuracies of 92.65% and 93.58% than the best individual classifier (92.53%). This demonstrates that these methods are effectively to improve the accuracy. Figure 13(a) is the result of improved D-S evidence theory using MMDC, MLPNN and SVM as member classifiers. Figure 13(b) is the result of concatenation combination (proposed method 1).

Figure 11. False color composite of OMIS image and classification results of individual classifier.

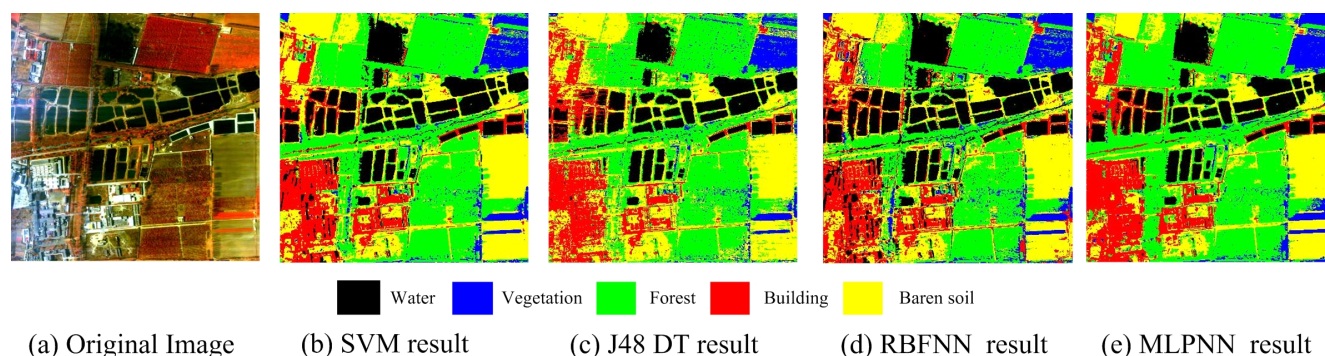
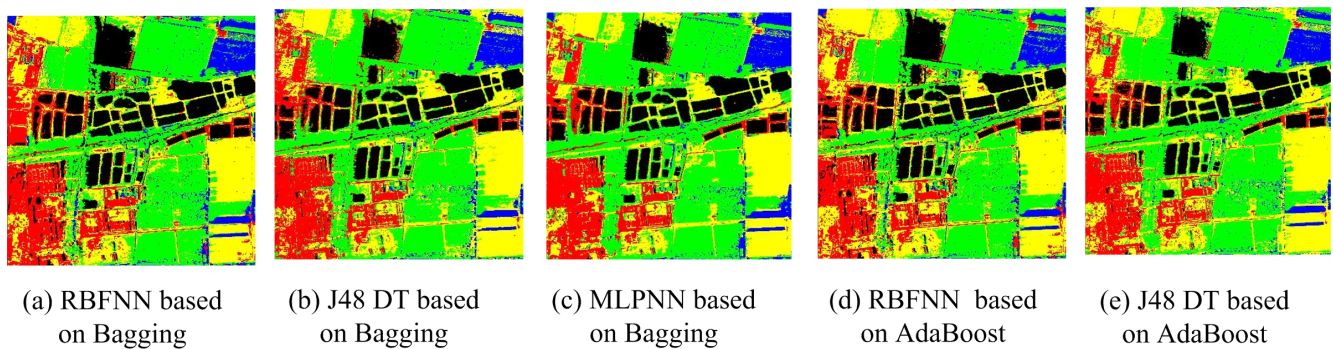
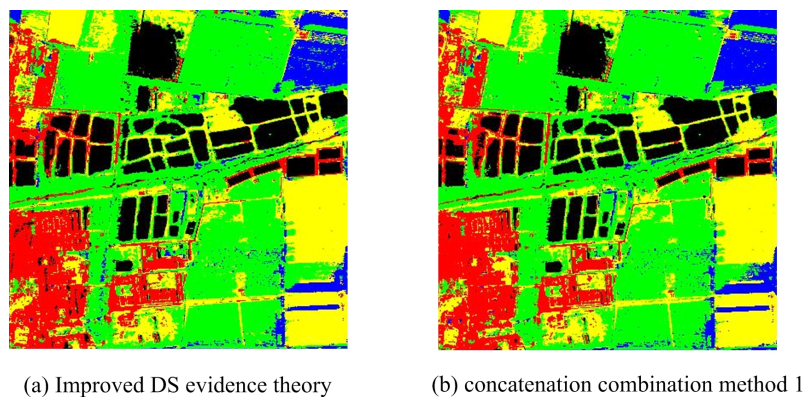


Table 9. Classification accuracy of single classifier.

Classifier	Water	Building	Forest	Vegetation	Bared soil
SVM	80.57%	97.34%	92.69%	89.96%	92.97%
J48DT	88.07%	86.38%	88.80%	95.82%	80.92%
RBFNN	95.72%	88.70%	89.58%	90.87%	81.63%
MMDC	94.80%	96.35%	88.42%	95.82%	80.92%
MLPNN	87.77%	98.34%	90.35%	98.10%	88.69%

Table 10. Classification accuracy statistics.

Method	OA	Kappa
SVM	90.72%	0.88
J48 DTC	87.86%	0.85
RBFNN	89.46%	0.87
MMDC	91.42%	0.89
MLPNN	92.53%	0.91
RBFNN based on bagging	89.88%	0.87
MLPNN based on bagging	93.09%	0.91
J48 DT based on bagging	90.44%	0.88
RBFNN based on AdaBoost	90.65%	0.88
J48 DT based on AdaBoost	91.00%	0.89
MLPNN based on AdaBoost	93.27%	0.91
Improved DS evidence theory	92.95%	0.91
Concatenation combination	93.58%	0.92

Figure 12. Classification result based on AdaBoost and Bagging.**Figure 13.** Classification result using parallel and concatenation combination style.

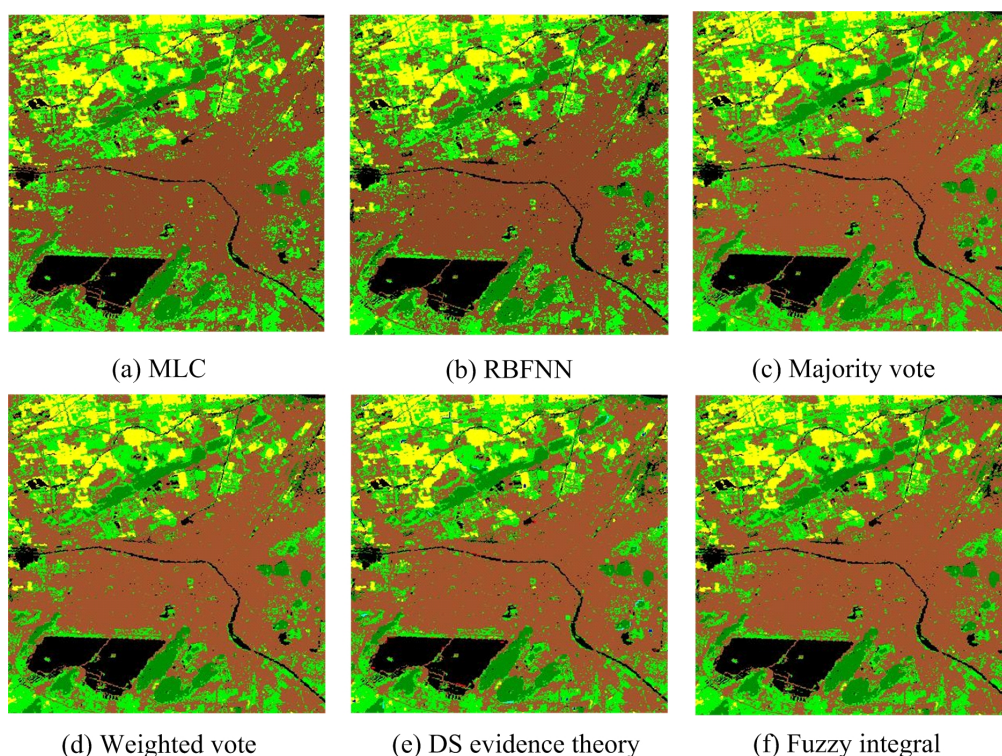
3.3. Experiment 3: Medium Resolution Multi-Spectral Landsat ETM+ Image

A Landsat ETM+ image acquired on September 17th 2004 is used to validate the above methods. The used bands of Landsat ETM+ images are as follows: band 1: Blue band; band 2: Green band; band 3: Red band; band 4: NIR band; band 5: MIR band, band 7: LIR band. The image size is 500×500 . The study area is the central part of Xuzhou City, China. Through the image analysis and field works, the land cover is classified into five classes: water, farmland, woodland, built-up land, and public green space. We use five member classifiers: RBFNN, J48DT, MMDC, MLC and MEDC. The adopted classifier combination methods are majority voting, weighted voting, D-S evidence theory, and fuzzy integral.

Table 11 summarizes the accuracies of those classification results. MLC achieves the highest accuracy (91.19%). All the four classifier ensemble methods obtain higher overall accuracies than MLC classifier. The accuracy of majority vote increases from 91.19% to 93.55%. Weighted vote achieves the accuracy of 93.55%, with the improvement 2.36%. The accuracy is increased by up to 2.23% and 1.8% by DS evidence theory and fuzzy integral approaches. This indicates that these methods effectively improve the accuracy. Figure 14 presents the classification results of different classifiers and classifier combination strategies.

Table 11. Classification accuracies of Landsat ETM+ image.

Method	OA	Kappa
RBFNN	88.09%	0.85
MMDC	86.35%	0.83
MLC	91.19%	0.89
MEDC	80.40%	0.76
J48 DTC	87.84%	0.85
Majority vote	93.55%	0.92
Weighted vote	92.56%	0.91
DS evidence theory	93.42%	0.92
Fuzzy integral	92.99%	0.91

Figure 14. Classification results of Landsat ETM+ image using different schemes.

3.4. Discussion

Based on the above analysis, we identified that the incorporation of diversity among the classifiers is of great importance in order to obtain better classification results via a MCS [31]. Combining the similar classification results would not further improve the overall accuracy. Bagging uses bootstrap sampling to generate diversity [16]. In order to enhance the diversity, concatenation combination constructs a proper dataflow so that diverse decisions could be first made by a subset of the committee classifiers and their conclusions are then set to another subset of classifiers. It is worthy to point out that the most important issue to combine the classifiers in parallel combination method is to examine the

difference among them using diversity measures. In addition, the selection of an appropriate ensemble strategy is also an important issue of MCS. This paper presents a comprehensive list of the most common parallel combination approaches and also analyzes the situation in which each of them are most useful. We also experimentally prove the effectiveness of the three modified concatenation combination algorithms within certain images. Furthermore, we studied the diversity measures to select the optimum combination. Compared to other diversity measures, the proposed diversity measure considers both the difference and accuracy among the classifiers. Thus, it can select the optimum combination whose accuracy is higher than the selected ones by other measures.

4. Conclusions

This review discusses the intensive contributions of MCS-based work in remote sensing. We present a comprehensive review of the most popular approaches and analyze the situations in which each of them is most useful. Those algorithms are experimented with three typical remotely sensed images, high resolution QuickBird image, airborne hyperspectral OMISII image, and medium Landsat ETM+ image. Some modified algorithms are proposed based on the ideas of improving MCS performance, and we experimentally proved the usefulness of these proposed methods.

Most of the findings from this paper show that both concatenation and parallel combination can enhance classification accuracy, but their performances are affected by different factors such as selected member classifiers, classifier combination criterion, *etc.* Furthermore, according to our experimental results, diversity measures can play active guidance for the selection of multiple classifiers combination.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments. The authors thank the support to this research from the National Natural Science Foundation of China (40871195, 41171323), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Natural Science Foundation of Jiangsu Province, China (BK2010182).

References

1. Navalgund, R.R.; Jayaraman, V.; Roy, P.S. Remote sensing applications: An overview. *Current* **2007**, *93*, 1747–1766.
2. Zhang, Y. Ten years of technology advancement in remote sensing and the research in the CRC-AGIP lab in GCE. *Geomatica* **2010**, *64*, 173–189.
3. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870.
4. Briem, G.; Benediktsson, J.; Sveinsson, J. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2291–2299.
5. Benediktsson, J.A.; Chanussot, J.; Fauvel, M. Multiple classifier systems in remote sensing: From basics to recent developments. *Mult. Classif. Syst.* **2007**, *4472*, 501–512.
6. Steele, B.M. Combining multiple classifiers: An application using spatial and remotely sensed information for land cover type mapping. *Remote Sens. Environ.* **2000**, *74*, 545–556.

7. Foody, G.M.; Boyd, D.S.; Sanchez-Hernandez, C. Mapping a specific class with an ensemble of classifiers. *Int. J. Remote Sens.* **2007**, *28*, 1733–1746.
8. Doan, H.T.X.; Foody, G.M. Increasing soft classification accuracy through the use of an ensemble of classifiers. *Int. J. Remote Sens.* **2007**, *28*, 4609–4623.
9. Smits, P.C. Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 801–813.
10. Waske, B.; van der Linden, S.; Benediktsson, J.A.; Rabe, A.; Hostert, P. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2880–2889.
11. Kumar, S.; Ghosh, J.; Crawford, M.M. Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *Pattern Anal. Appl.* **2002**, *5*, 210–220.
12. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*; Wiley-Interscience: Hoboken, NJ, USA, 2004.
13. Rokach, L. *Pattern Classification Using Ensemble Methods*; World Scientific: Singapore, 2010.
14. Biggio, B.; Fumera, G.; Roli, F. Multiple classifier systems for robust classifier design in adversarial environments. *J. Mach. Learn. Cybern.* **2010**, *1*, 27–41.
15. Ceamanos, X.; Waske, B.; Benediktsson, J.A.; Chanussot, J.; Fauvel, M.; Sveinsson, J.R. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *Int. J. Image Data Fusion* **2010**, *1*, 293–307.
16. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.
17. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. In *Proceedings of the International Conference on Machine Learning*, Bari, Italy, 3–6 July 1996; pp. 148–156.
18. Benediktsson, J.A.; Swain, P.H. Consensus theoretic classification methods. *IEEE Trans. Syst. Man Cybern.* **1992**, *22*, 688–704.
19. Benediktsson, J.A.; Sveinsson, J.R. Hybrid consensus theoretic classification. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 833–843.
20. Fauvel, M.; Chanussot, J.; Benediktsson, J.A. Decision fusion for the classification of urban remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2828–2838.
21. Ebeir, L.D.; Atinne, P.A.L.; Teen, I.S.V. Remote sensing classification of spectral, spatial and contextual data using multiple classifier systems. *Image Rochester N. Y.* **2001**, *20*, 584–589.
22. Waske, B.; Benediktsson, J.A. Fusion of support vector machines for classification of multisensor data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3858–3866.
23. Nemmour, H.; Chibani, Y. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS J. Photogramm. Remote Sens.* **2006**, *61*, 125–133.
24. Waske, B.; Braun, M. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 450–457.
25. Krogh, A.; Vedelsby, J. Neural network ensembles, cross validation, and active learning. *Adv. Neural Inf. Process. Syst.* **1995**, *7*, 231–238.
26. Chandra, A.; Yao, X. Evolving hybrid ensembles of learning machines for better generalisation. *Neurocomputing* **2006**, *69*, 686–700.

27. Kuncheva, L.I.; Whitaker, C.J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.* **2003**, *51*, 181–207.
28. Brown, G.; Wyatt, J.; Harris, R.; Yao, X. Diversity creation methods: A survey and categorisation. *Inf. Fusion* **2005**, *6*, 5–20.
29. Michail, P.; Benediktsson, J.A.; Ioannis, K. The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. *IEEE Trans. Geosci. Remote Sens.* **2002**, *39*, 2539–2546.
30. Foody, G.M. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sens. Environ.* **2009**, *113*, 1658–1663.
31. Ranawana, R.; Palade, V. Multi-classifier systems: Review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.* **2006**, *3*, 1–41.
32. Kuncheva, L.I. A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 281–286.
33. Giacinto, G.; Roli, F.; Vernazza, G. Comparison and combination of statistical and neural network algorithms for remote-sensing image classification. In *Neurocomputation in Remote Sensing Data Analysis*; Springer-Verlag: Berlin, Germany, 1997; pp. 117–124.
34. Ho, T.K.; Hull, J.J.; Srihari, S.N. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 66–75.
35. Giacinto, G. An approach to the automatic design of multiple classifier systems. *Pattern. Recogn. Lett.* **2001**, *22*, 25–33.
36. Ghosh, A.; Shankar, B.U.; Bruzzone, L.; Meher, S.K. Neuro-fuzzy-combiner: An effective multiple classifier system. *Int. J. Knowl. Eng. Soft Data Paradig.* **2010**, *2*, 107–129.
37. Woods, K.; Kegelmeyer, W.P.; Bowyer, K. Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 405–410.
38. Czyz, J.; Kittler, J.; Vandendorpe, L. Multiple classifier combination for face-based identity verification. *Pattern. Recogn.* **2003**, *37*, 1459–1469.
39. Rodriguez, J.J.; Kuncheva, L.I. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *28*, 1619–1630.
40. Benediktsson, J.A.; Kanellopoulos, I. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1367–1377.
41. Fauvel, M.; Chanussot, J.; Benediktsson, J.A. Decision fusion for the classification of urban remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2828–2838.
42. Zhou, Z.H. When Semi-Supervised Learning Meets Ensemble Learning. In *Proceedings of the 8th International Workshop on Multiple Classifier Systems (MCS '09)*, Reykjavik, Iceland, 10–12 June 2009; Volume 5, pp. 529–538.
43. Brown, G. Ensemble Learning. In *Encyclopedia of Machine Learning*; Webb, G.I., Sammut, C., Eds.; Springer: New York, NY, USA, 2010.
44. Xu, L.; Krzyzak, A.; Suen, C.Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Syst. Man Cybern.* **1992**, *22*, 418–435.

45. Rahman, A.F.R.; Fairhurst, M.C. Serial combination of multiple experts: A unified evaluation. *Pattern Anal. Appl.* **1999**, *2*, 292–311.
46. Aksela, M.; Laaksonen, J. Using diversity of errors for selecting members of a committee classifier. *Pattern. Recogn.* **2006**, *39*, 608–623.
47. Gislason, P.; Benediktsson, J.A.; Sveinsson, J.M. Random forests for land cover classification. *Pattern. Recogn. Lett.* **2006**, *27*, 294–300.
48. Bauer, E.; Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.* **1999**, *36*, 105–139.
49. Breiman, L. Random forest. *Mach. Learn.* **2001**, *45*, 5–32.
50. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
51. Collins, M.; Schapire, R.E.; Singer, Y. Logistic regression, adaboost and bregman distances. *Mach. Learn.* **2002**, *48*, 31–44.
52. Webb, G.I. MultiBoosting: A technique for combining boosting and wagging. *Mach. Learn.* **2000**, *39*, 159–196.
53. Zhang, C.; Zhang, J. RotBoost: A technique for combining Rotation Forest and AdaBoost. *Pattern. Recogn. Lett.* **2008**, *29*, 1524–1536.
54. Moreno-Seco, F.; Inesta, J.; de Leon, P.; Mico, L. Comparison of classifier fusion methods for classification in pattern recognition tasks. *Struct. Syntactic Stat. Pattern Recogn.* **2006**, *4109*, 705–713.
55. Giacinto, G.; Roli, F. Design of effective neural network ensembles for image classification. *Image Vis. Comput. J.* **2001**, *19*, 697–705.
56. Shipp, C.A.; Kuncheva, L.I. Relationships between combination methods and measures of diversity in combining classifiers. *Inf. Fusion* **2001**, *3*, 135–148.
57. Kittler, J.; Hatef, M.; Duin, R.P.W.; Matas, J. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 226–239.
58. Nemmour, H.; Chibani, Y. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *ISPRS J. Photogramm. Remote Sens.* **2006**, *61*, 125–133.
59. Kong, Z.; Cai, Z. Advances of Research in Fuzzy Integral for Classifiers' fusion. In *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing—Volume 02*, Washington, DC, USA, 30 July–1 August 2007; pp. 809–814.
60. Ben Abdallah, A.C.; Frigui, H.; Gader, P. Adaptive local fusion with fuzzy integrals. *IEEE Trans. Fuzzy Syst.* **2012**, in press.
61. Sun, Q.; Ye, X.Q.; Gu, W.K. A new combination rules of evidence theory (in Chinese). *Acta Electron. Sin.* **2000**, *8*, 117–119.
62. Ko, A.; Sabourin, R.; Brittojr, A. From dynamic classifier selection to dynamic ensemble selection. *Pattern. Recogn.* **2008**, *41*, 1718–1731.
63. Woods, K.; Kegelmeyer, W.P.; Bowyer, K. Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 405–410.

64. Giacinto, G.; Roli, F.; Fumera, G. Selection of image classifiers. *Electron. Lett.* **2000**, *36*, 420–422.
65. Dudani, S.A. The distance-weighted k-nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *SMC-6*, 325–327.
66. Ruta, D.; Gabrys, B. An overview of classifier fusion methods. *Comput. Inf. Syst.* **2000**, *7*, 1–10.
67. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259.
68. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790.
69. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565.
70. 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.
71. Hu, X.; Weng, Q. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sens. Environ.* **2009**, *113*, 2089–2102.
72. Witten, I.H.; Frank, W. *Data Mining: Practical Machine Learning Tools and Techniques*; China Machine Press: Beijing, China, 2006.
73. Dennison, P.E.; Halligan, K.Q.; Roberts, D.A. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sens. Environ.* **2004**, *93*, 359–367.
74. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993.
75. Zhang, W.; Du, P.; Yin, Z.X. Application of simplified fuzzy H-ARTMAP network to hyperspectral remote sensing classification (in Chinese). *Sci. Surv. Mapp.* **2009**, *33*, 113–115.