Performance enhancement of standard fuzzy majority voting-based fusion of probabilistic classifiers

Mahmoud Salah
Department of Surveying Engineering, University of Benha, 108 Shoubra Street, Cairo, Egypt
Email: engmod2000@yahoo.com

(Received: Dec 16, 2018; in final form: May 10, 2019)

Abstract: Combining classifiers is essential for feature extraction and mapping applications. This paper proposes an approach to improve the performance of one of the most frequently used Multiple Classifier Systems (MCSs), namely the Fuzzy Majority Voting (FMV). First, a set of texture attributes has been generated from a 0.82mpan-sharpened IKONOS image covers the test area. The generated attributes along with the original image have been applied as input for three-member classifiers: Artificial Neural Networks (ANN); Support Vector Machines (SVM); and Classification Trees (CT). Before combination, a weighting criterion has been determined, based on the performances of each member classifier, and assigned to the output of that classifier. After that, The FMV has been applied for combining the weighted results from the three-member classifiers to extract buildings (B), roads (R) and vegetation (G). The proposed method has been tested and compared with the three-member classifiers as well as the standard FMV. The results have been analyzed considering four different aspects: (1) overall accuracy; (2) class accuracy; (3) sensitivity to training sample size; and (4) computational complexities. The proposed method resulted in an overall classification accuracy of about 95.60%, which is 3.88, 6, 8.51 and 1.24% better than ANN, SVM, CT, and standard FMV respectively. On the other hand, most of the class-accuracies are much better and less variable than those obtained by any member classifier as well as the standard FMV. While the proposed method is stable and always outperforms individual classifiers even in the cases of small size training samples, its computational cost is still comparable with that of standard FMV.

Keywords: MCSs, FMV, High Resolution (HR) satellite imagery, feature extraction.

1. Introduction

Feature extraction from HR satellite imagery is an important task in remote sensing (RS) and geographic information system (GIS) applications. However, Classification of remotely sensed imagery is still a challenge because of the low illumination and low spatial resolution of satellite imagery, noise, ambiguity and uncertainty in the decision-making process. There is no single classifier that can be optimal for all classification problems since single classifiers sometimes lead to inefficient classification accuracy. Combining classifiers can improve the classification accuracy by integrating the outputs of multiple individual classifiers through some linear or non-linear MCSs (Moustakidis et al., 2012). Over the last decade, MCSs can be considered as one of the most important advancements in the field of pattern recognition (International workshop on multiple classifier systems, 2007).

The appropriate combination of individual classifiers usually results in better performance in terms of classification accuracy and/or CPU time. MCSs can outperform any individual classifier in cases of small training dataset, local optima problem and a huge amount of data (Woz´niak et al., 2014). As well, it can improve the performance of weak classifiers and stabilize the decision of random ones such as ANN and CT.

On the other hand, the diversity of member classifiers of a given MCS can perfectly handle the problem of noisy data (Ponti-Jr. and Papa, 2011). It is worth mentioning that classifiers with correlated results may provide lower accuracy than the worst classifier. In the case of correlated classifiers, the MCSs error will be almost the same as the average error of the member classifiers. On the other hand, it will be *n* times smaller than the average error of the members in the case of statistically independent classifiers (Tumer and Ghosh, 1996). In this regard, the successfulness of a MCS is based on the degree of diversity between individual classifiers. Diversity can be achieved by using: different input features, different training samples, or different classifiers. A detailed review of the most common diversity measures is given by Ranawana and Palade (2006).Recently, MCSs have been introduced to remote sensing applications in: satellite image classification (Maulik and Chakraborty, 2010); land cover mapping (Han et al., 2012); and change detection (Du et al., 2013).

In general, MCSs can be grouped into three categorize: parallel; serial; and hierarchical (Lv et al., 2000). In the parallel form, the same data are used as input for each individual classifier independently, and the final decision is made by combining their independent results. In the serial form, individual classifiers are applied in sequence. This form starts with a primary classifier, the classifier with the cheapest computational cost, followed by the secondary classifiers, the ones with higher computational cost (Fumera et al., 2004). The hierarchical method combines both parallel and serial techniques in order to obtain optimal combination results (Ranawana and Palade, 2006). The majority of publications are focused on the parallel combination approach since it is simple to implement, easy to analyze and can take advantages of all member classifiers (Woz´niak et al., 2014).

Segrera and Moreno (2005) categorized the methods for building MCSs into two groups: ensemble, and hybrid. Ensemble MCSs combine classifiers with the same learning models while modifying the input training data set for each classifier. Ensemble-based combination typically has lower generalization errors than any of its individual models (Ahn et al., 2007). Bagging and boosting techniqueshave been considered in most of the previous studies as a typical example of ensemble-based approaches. Both bagging and boosting introduce diversity by considering different training samples and only one base classifier. Dietterich (2000) concluded that with low noise data, boosting outperforms bagging technique. On the other hand, bagging outperforms boosting in the case of high noise data (Khoshgoftaar et al., 2011). The main problem with Bagging and boosting is the large number of classifiers in the final MCS. As well, measuring diversity is still an open research area (Cavalcanti et al., 2016). Hybrid MCSs, on the other hand, combine classifiers with different learning algorithms. In this regard, many combination rules have been considered in the literature: the intersection of decision regions (Haralick, 1976); voting methods (Mazurov et al., 1987); prediction by top choice combinations (Wemecke, 1992); Dempster–Shafer theory (Xu et al., 1992); and ranking methods (Ho et al., 1994).

Dai and Liu (2010) proposed a MCS to combine the results from six base classifiers: maximum likelihood (ML); SVM; ANN; spectral angle mapper (SAM); minimum distance (MD); and CT. A voting strategy was applied for the combination. The results confirmed that the MCS performed much better than any base classifier. Ghimireet al. (2012) have compared three combination techniques based on decision trees (DT): bagging; AdaBoost; and random forest (RF). For the three cases, the MCS has outperformed the DT-base classifier. Kumar and Meher (2013) proposed an effective MCS based on multiple rules of granular neural networks (GNN) with improved performance classification accuracy. Khosravi and Beigi (2014) applied bagging and boosting to construct a MCS based on SVM to classify a hyperspectral dataset. The results showed a perfect performance of the MCS for classifying high dimensional data. Chen et al. (2017) constructed a MCS to combine SVM, DT and ANN using the AdaBoost technique. The results showed that the MCS has effectively improved the classification accuracy as compared with individual classifiers.

In order to select the most suitable combiner for a given problem, some guidelines have been given in the literature: majority voting (Kimura and Shridhar, 1991) for combining class labels; averaging techniques (Hashem and Schmeiser, 1995) for combining posterior probabilities; fuzzy logic (Zadeh, 1965), Dempster-Shafer theory of evidence (DS) (Rogova, 1994) or ANN (El-Melegy and Ahmed, 2007) for combining fuzzy membership values. Detailed reviews of MCSs are given by Woz´niak et al. (2014). The most recent techniques are usually presented in the International Workshop on MCSs and in Machine Learning and Pattern Recognition Journals.

It is worth mentioning that the fuzzy set theory is more suitable for pattern recognition in the case of remotely sensed data where classes are normally ill-defined and overlapped (Kuncheva, 2000). On the other hand, it is not depending on the input data which is the main drawback of most of the existing MCSs. Many attempts have been made for RS image analysis and classification using fuzzy sets (Chen, 2000; Tso and Mather, 2001; Ghosh et al., 2008). Salah et al. (2010) applied the standard FMV to combine non-weighted ANN, CT and SVM classifiers using aerial images and LiDAR data. The results demonstrate that the standard FMV has improved the classification accuracy as compared with the best single classifier. FMV is the most commonly used fuzzy sets technique since it is easy to apply and able to manage imperfect data (Ponti-Jr., 2011).

To the best of the author knowledge, the effect of weighting classifiers and training sample size on the accuracy and robustness of FMV-based fusion have not been considered in the literature. The objective of this paper is to define clear guidelines to explain under which conditions the FMV are able to improve the performance of individual classifiers. To meet the objective, ANN, SVM and CT have been adopted as base classifiers. The three classifiers have different modelling and learning criteria which lead to different errors and then complementary information. In this regard, the parallel combination technique has been applied. The three-member classifiers have been applied to classify the test area using IDRISI Taiga software (Clark labs, 2012). The proposed fusion method in this research has been implemented through a set of codes generated by the author in Matlab environment. For the rest of the paper, the term WFMV will be used to refer to the FMV-based fusion of the weighted classifiers, while the term SFMV will be used to refer to the FMV-based fusion of the standard or non-weighted classifiers.

2. Study area and data sources

2.1 Pan-sharpened satellite image

An IKONOS image covers the test area was collected on April 17, 2010, and supplied in a digital TIFF format. The IKONOS image has been created using a pan-sharpening process that combines the 0.82m panchromatic band with the 3.2m multispectral bands to create 0.82m colour image. Table 1 summarizes IKONOS satellite pacifications. The test area is a dense urban area with medium size residential buildings, a large network of main and minor roads, as well as open vegetation areas as shown in figure 1.

Table 1: IKONOS satellite specifications

Imaging Mode	Panchromatic	Multispectral
Pixel Size	0.82 meter	3.2 meter
Spectral Range	450-900 nm	450-520 nm (blue) 520-600 nm (green) 625-695 nm (red) 760-900 nm (NIR)
Dynamic Range	11 bit/pixel	11 bit/pixel

Figure 1: The one-meter pan-sharpened IKONOS image covers the study area

2.2 Reference data

In order to evaluate the performance of the proposed combination method, B, R and G were digitized in the image and applied as reference data as shown in figure 2, Class “G" mainly corresponds to grass and trees. All recognizable objects were digitized independently of their size. Joined buildings that were obviously separated were digitized as separate buildings; otherwise, they were digitized as one building.

Figure 2: Reference data used for evaluating the performance of the proposed method. Red: B, green: G, black: R

2.3 Feature attributes

In order to describe classes effectively, a wide variety of spectral attributes have been generated and only the most useful ones, as shown in figure 3, have been statistically selected based on Yang (2007). The selected attributes and the original multispectral image have been used simultaneously as input data for the classification process. The objective is to solve for two common problems associated with HR digital imagery which are: 1) shadows caused by buildings and trees; 2) and spectral variability within the same land use/cover class (Lu and Weng, 2007). On the other hand, it provides useful information for improved land use/cover classification (Hirose et al., 2004). For more details about the formulas used for calculating attributes, one can refer to Russ (2002).

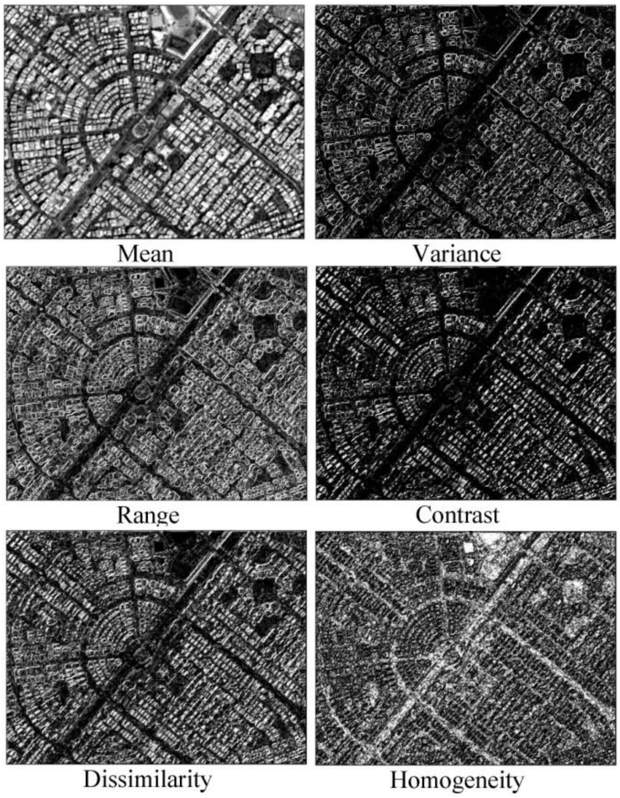


Figure 3: The set of attributes that have been applied as the input for individual classifiers

2.4 Training Datasets

Training datasets assemble a set of statistics to describe the spectral pattern for each land use/cover class in the image. A minimum of (n+1) pixels is required for a signature with n is the number of bands (Lillesand and Kiefer, 2004). The training data used are sets of manually digitized samples from the image for each land use/cover class. Polygons of approximately the same areas were digitized for B, R and G classes. As recommended by Kuncheva (2004), the same training samples have been applied to train all the member classifiers. Digital numbers (DNs) in a range between 0 and 255, corresponding to reflectance values, have been applied to generate the training samples. The selected signatures are compared in a graph representing DNs for each signature from the red band as shown in figure 4. The clear separation for most DNs values indicates that the selected signatures represent a completely distinct set of pixels, which is essential for good classification results.

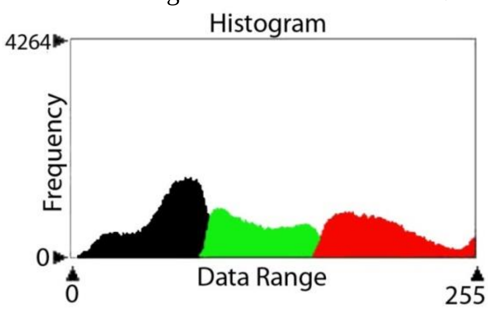


Figure 4: Minimum and maximum DNs for signatures, from the red band: black for R; green for G; and red for B

3. Methodology

The proposed FMV-based fusion of ANN, SVM and CT classifiers has been implemented in several steps as shown in figure 5. The proposed MCS has four phases. First, training data are fed into the three-member classifiers to obtain individual decisions. After that, the obtained probabilities were weighted according to the relative importance of each classifier. The 10-fold cross-validation technique was then applied to tune the shape and position of the fuzzy membership function. At the end, FMV is applied to combine the weighted probabilities from the three-member classifiers and form the final decision.

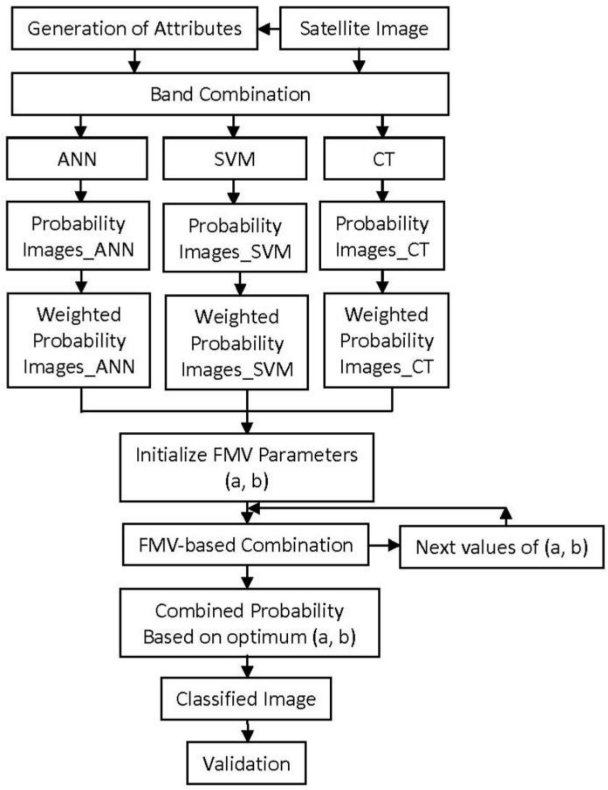


Figure 5: The proposed FMV-based fusion workflow

3.1 Base Classifiers

In order to improve the performance of the MCS, individual classifiers should have different mathematical concepts and offer complementary information. As well, in the case of two base classifiers, a limited improvement in classification accuracy can be obtained by the MCS (Chen et al., 2017). In this regard, three different algorithms have been applied as member classifiers. These classifiers represent different learning criteria and include SVM as a machine learning classifier; ANN as an artificial neural networks classifier; and CT as a statistical classifier. The output of each classifier is a degree of membership of every pixel for each class.

3.1.1 ANN

ANN is a self-learning algorithm that can compensate for uncertainty in information and can perfectly handle the problem of high spectrum confusion in remotely sensed data. This can be done by setting the number of nodes in the hidden layers (Coppin et al., 2010). The ANN is trained based on randomly chosen initial weights (Hu, 2000). The most common and widely used feed-forward back propagation neural algorithm, multi-layer perceptron (MLP), has been applied. The network consists of three layers: input; hidden; and output. The number of input nodes of the MLP is the number of the input features; the number of output nodes is the number of classes; and the number of hidden nodes is between *2N* to *3N* where *N* is the number of classes (Ghosh and Uma Shankar, 2010). In this regard, the MLP was a seven hidden layers MLP with nine input neurons, one for each input variable, and three output neurons, one for each class. Except for the input nodes, a weight is calculated for each node as the sum of the output at the nodes to which it is connected in the preceding layer. In order to derive the final output and fed it to the nodes in the next layer, the weighted sum is passed through a transfer function as follow:

$$net_v = \sum_u W_{uv}\, O_u + bias_v and O_v = S(net_v) \quad (1)$$

Where:
w_{uv}: the weight for the connection between nodes u and v
$bias_v$: the bias for node v
O_u: the output at node u
S: the sigmoid activation function. This function can perfectly handle nonlinear problems (Cybenko, 1989).

For weights updating, MLP uses a back-propagation learning algorithm to reduce the sum of square error between the obtained and desired output in a descending manner as follows (Haykin, 1998):

$$\Delta W_{uv}(n+1) = \alpha \Delta W_{uv}(n) + \eta \delta_v O_u \qquad (2)$$

Where n, α, η and δ are the iteration number, momentum parameter, learning rate and node error respectively. In this regard, each input pattern is assigned to the class that corresponds to the highest node value obtained at the output of the MLP. In order to improve the performance of the MLP with reasonable processing time, a set of parameter values suggested by Kavzoglu and Mather (2003) have been applied as shown in table 2.

Table 2: The basic architecture to start the MLP classifier

Parameter	Value
n	10,000
α	0.5
η	0.05
δ	0.0001

3.1.2 CT

CT was introduced by Breiman et al. (1984). It is a non-parametric technique that uses an iterative procedure known as binary recursive partitioning. In this regard, a heterogeneous sample of training data with multiple classes is hierarchically and progressively subdivided into more homogeneous classes based on a binary splitting rule to form the tree. The tree is then used to classify the whole datasets. Classification trees have proved to be strong, simple to implement, ideal for noise minimization, highly automatic and perfect for complex data such as multi-source and/or multi-scale data. In the classification process, only the most useful attributes are selected and used (Chen et al., 2017).

Three models can be used with CT as splitting criteria which include: Entropy, Gain Ratio, and Gini. The Entropy algorithm (after Shannon, 1949) has proved to be preferable for classification problems from HR digital imagery (Salah et al., 2011) and has been applied as the splitting criterion in this study. The method decreases the entropy until a terminal node that has zero entropy (contains pixels from one class) is reached. In order to identify class x_i of a training dataset in node N, the entropy can be described as in equation 3 where $P(x_i)$ is the probability of class x_i.

$$Entropy(N) = -\sum_{i=1}^{l} P(x_i)\log_2 P(x_i) \qquad (3)$$

A 10-fold cross-validation process has been applied for pruning the trees. This technique has proved to be highly accurate and requires no independent dataset for assessing the performance of the model.

3.1.3 SVM

Recently, SVM has become a common tool to classify linear and nonlinear problems. It is based on statistical learning theory and has excellent learning performance especially when applied to remotely sensed data. First, the input feature space is transformed into a high dimensional one, and then an optimal hyperplane is fitted into data to separate 0 and 1 classes by maximizing the margin between them. The closest data points to the hyperplane are referred to as support vectors (Vapnik, 1995).

Since the One-Against-One (1A1) technique normally results in a huge number of binary SVM as well as intensive computations, the One-Against-All (1AA) technique has been applied to solve for the 0/1 classification problem. The radial basis function (RBF)kernel can nonlinearly map more complex data into a higher dimensional space with reasonable processing times and has proved to be effective for remote sensing applications. In this regard, it has been applied for fitting the hyperplane into data (Van der Linden et al., 2009). The general mathematical representation of the RBF kernel is shown in equation 4. The gamma term, γ, is a user-controlled parameter and its correct definition can significantly improve the performance of the SVM classification.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \qquad (4)$$

The 10-fold cross-validation technique has been applied to determine the optimal γ value, 0.03. This technique has proved to be effective to prevent over-fitting problems and usually results in better performance (Hsu et al., 2009). The sequential minimal optimization (SMO) algorithm has been applied for training the SVM through breaking the large size optimization problems into a series of smallest ones. This can speed up the computations and minimize memory requirements (Platt, 1999).

3.2 WFMV-Based fusion

FMV has proved to be a powerful technique to handle the uncertainties and imprecision in remotely sensed data by defining a fuzzy membership function. A membership function is a relation that shows how a certain point in the input space will be mapped as a membership value in the output space. Before incorporating the probabilities, pp_i, from ANN, SVM, and CT into the FMV, the probabilities are weighted by the accuracies estimated for the corresponding classifiers as an importance weight. This weighting process minimizes errors for the FMV in the case of independent outputs. Let the classification accuracies obtained from the accuracy assessment process are: α_{c1}; α_{c2}; and α_{c3} for ANN, SVM, and CT respectively. The weighted probabilities, pp_i`, for a classifier $c_i$may be given as follows:

$$w_i = \log\left(\frac{\alpha_{ci}}{1-\alpha_{ci}}\right) \qquad (5)$$

$$pp_i` = w_i * pp_i \qquad (6)$$

The idea behind the FMV is to give some semantics, meanings, to the weights. Based on this semantics, the weights can be determined directly (Zadeh, 1983). First, the membership function of relative quantifiers are defined as in equation 7 (Herrera and Verdegay, 1996), With parameters a, $b \in [0, 1]$ and pp_i`is the weighted class membership of pixel i.

$$Q_{P_i} = \begin{cases} 0 & if\ pp`_i < a \\ \dfrac{pp`_i - a}{b-a} & if\ a \le pp`_i \le b \\ 1 & if\ pp`_i > b \end{cases} \qquad (7)$$

The value 1 of Q_{P_i} indicates that the quantifier is completely satisfied. On the other hand, the quantifier is not fulfilled at all if Q_{P_i} = 0 and any intermediate value Q_{P_i} indicates an intermediate fulfilment degree. The optimal selection of fuzzy parameters (a, b) has a direct impact on the FMV performance since they control the shape and position of the membership function. Unfortunately, there is no precise mathematical method to define these parameters (Saheb et al., 2013). A grid-search on a and b using a 10-fold cross-validation was used for this purpose. Basically, pairs of (a, b) were tested and the one with the best cross-validation accuracy was selected. In this regard, a grid with an interval of 0.1 for both a and b has been applied. Then, the weights based on the linguistic quantifier can be determined as in equation 8, with i is the order of a given classifier after ranking Q_{P_i} in a descending order and N is the total number of classifiers (Yager 1998):

$$w_{P_i} = Q_{P_i}\left(\frac{i}{N}\right) - Q_{P_i}\left(\frac{i-1}{N}\right),\ for\ i = 1\ ,\ \ldots..,N \qquad (8)$$

The final combined probability can be determined as in equation 9, with k is the number of classes.

$$P_{WFMV} = \underset{k}{\mathrm{argmax}}\left[\sum_{i=1}^{N} w_{p_i}\; pp`_i\right] \quad (9)$$

3.3 Accuracy assessment

In order to evaluate the performance of the proposed method, the results have been compared with the reference data. The overall classification accuracy *OA* has been determined as in equation 10 with *NCP* is the total number of correctly classified pixels and *NRP* is the total number of reference pixels. Since the overall classification accuracy is just a global measure for the performance of the combination process, the users and producers accuracies (*UA* and *PA*) have been used. Unlike overall classification accuracy, *UA* and *PA* clearly indicate how the proposed methods improve or deteriorate the results for individual classes as shown in equations 11 and 12. *CP* is the correct class predictions, *TP* is the total predictions and *TCP* is the total class pixels.

$$OA = \frac{NCP}{NRP} \quad (10)$$

$$UA = \frac{CP}{TP} \quad (11)$$

$$PA = \frac{CP}{TCP} \quad (12)$$

4. Results and discussion

Initially, classification of the satellite image has been performed and the parameters for each classifier have been estimated using the labeled training samples. Once the optimum parameters were selected, each classifier has been applied to classify the whole image. The obtained results are nine probability images, three for each classifier, representing the membership of each pixel to each class. The probabilities were then modified by assigning weights derived from the classification accuracy of the corresponding classifier. The membership values are true probabilities in the range of 0 to 1 as shown in figure 6.

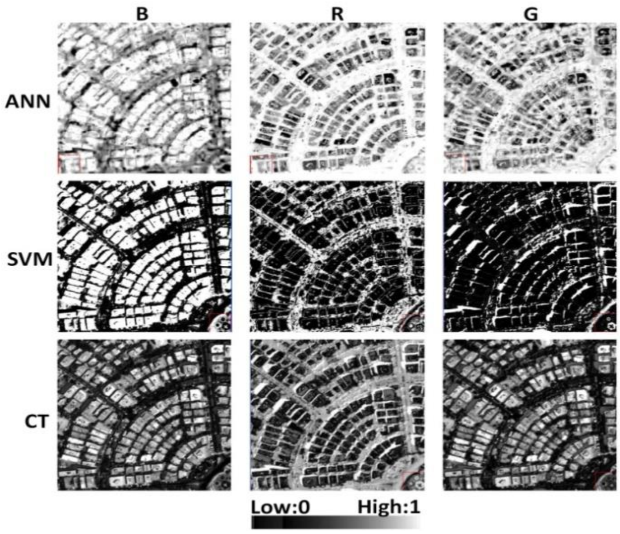


Figure 6: A typical example showing the weighted membership images of individual classifiers

Before applying the WFMV algorithm to combine these probabilities, a grid-search on *a* and *b* using a 10-fold cross-validation and grid interval of 0.1 for both *a* and *b* has been applied. As a result, a relative quantifier with parameters (0.1, 0.5) has performed the best for the membership function QP_iin equation 7 as graphically depicted in figure 7. Once *a* and *b* have been determined, they were applied with the nine weighted probability images to perform the WFMV-based combination. Figure 8 is a typical example of the WFMV output which is three probability images representing the membership values of every pixel for each class.

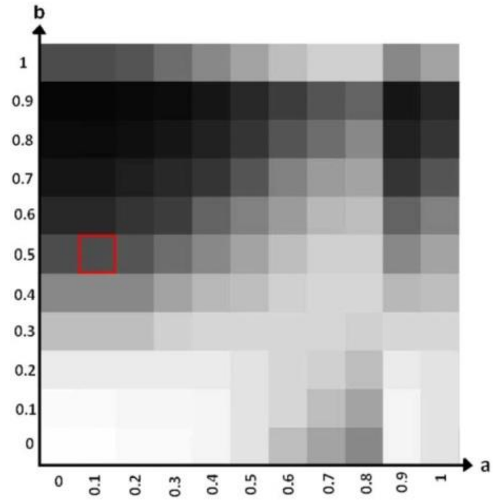


Figure 7: The grid search results using the input data

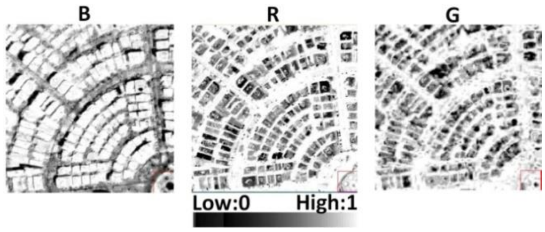


Figure 8: A typical example showing the membership images of the WFMV system

For each pixel, the membership values for all classes were compared and the class with the highest value was assigned to that pixel to create a WFMV-based classification image. Figure 9 is a typical example illustrates the original image, the classification results obtained for individual classifiers, SFMV and WFMV. By focusing on the buildings inside the white squares, one can find that the WFMV has detected complete buildings much better than SVM and CT, and as good as ANN and SFMV. On the other hand, the white circled regions indicate that WFMV has detected separate buildings more accurate than any individual classifier, as well as SFMV. However, many vegetation was classified as roads by the three-member classifiers and hence by SFMV and WFMV as graphically depicted in the white rectangles. An expected reason for that can be the high degree of similarity between the spectral reflectance of roads and vegetation in the used pan-sharpened IKONOS image. One possible solution is to use a Normalized Difference Vegetation Index (NDVI) which may increase the classification accuracy due to its ability to detect vegetation accurately.

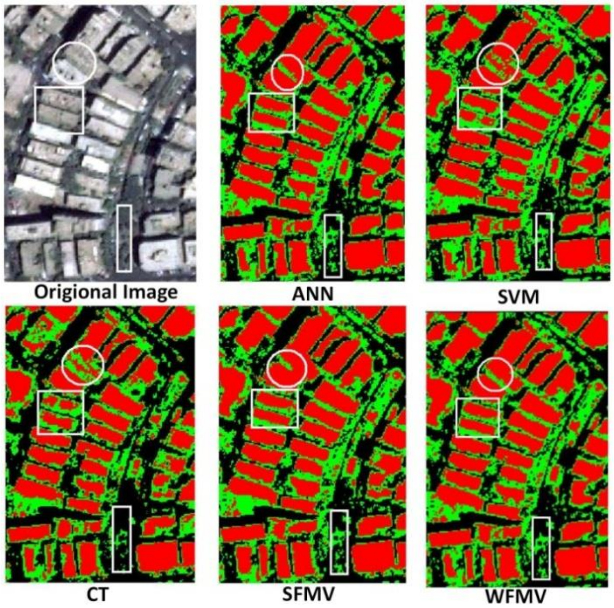


Figure 9: Results of individual classifiers, SFMV and WFMV

4.1 Overall accuracy
In order to evaluate the performance of the WFMV, it has been compared with single classifiers and with SFMV. Table 3 summarizes the obtained accuracies for the three classifiers and their combinations by SFMV and WFMV. The WFMV has statistically performed better than any individual classifier as well as SFMV. For individual classifiers, ANN performed the best with 91.72% overall accuracy, followed by SVM with 89.60% overall accuracy and CT with an overall accuracy of 87.09%. The WFMV resulted in overall classification accuracy of about 95.60%, which is 3.88, 6, 8.51 and 1.24% better than ANN, SVM, CT and WFMV respectively.

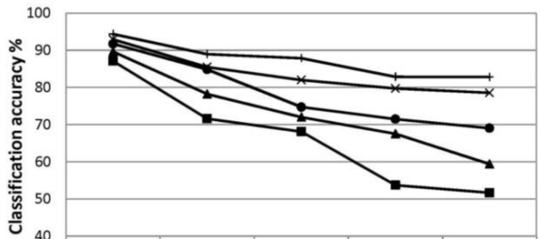
Table 3: Overall accuracy of different classification methods

classifier	ANN	SVM	CT	SFMV	WFMV
Overall Accuracy	91.72	89.60	87.09	94.36	95.60

4.2 Class accuracy
In terms of class accuracy, the three-member classifiers resulted in different class accuracies for the same test area as shown in table 4. No single classifier has performed the best for all classes. A typical example is that CT resulted in lower UA for vegetation, 58.75%. On the other hand, it outperformed the ANN and SVM in classifying buildings with UA of about 99.99%. These results confirm that classifiers with different algorithms are complementary and result in different classification accuracies for different classes.

Assessments of class accuracies confirmed that the WFMV-based fusion performed the best in most cases as shown in table 4. Most of the class-accuracies are improved by the WFMV. Whereas the application of ANN, SVM, CT and SFMV resulted in average class accuracies (average of UA and PA) of 90.68, 88.91,86.73 and 91.91% respectively, the application of WFMV fusion displayed a significant improvement and resulted in average class accuracy of 95.38%. Another advantage of WFMV-based fusion is that the obtained errors are less variable. Whereas the application of ANN, SVM, CT and SFMV resulted in standard deviations (SD) of 10.59, 10.75, 14.89 and 9.79 respectively; the WFMV-based fusion resulted in SD of 4.02. Thus it conforms to the requirement of Anderson et al. (1976) that the class accuracies of different classes should be about equal.

4.3 Sensitivity to training sample size
In order to obtain a robust decision about the performance of the WFMV system, five different training samples (100, 200, 300, 400 and 500 pixels) evenly distributed through the test area were selected and tested. As can be observed from figure 10, WFMV always improves the performance of individual classifiers and outperforms the SFMV even in the cases of small size training samples. This behaviour can be clearly observed in the cases of training samples of size less than 300 pixels. On the other hand, WFMV is the most stable classifier followed by SFMV, ANN, SVM and CT respectively. Decreasing the sample size from 500 to 100 pixels has decreased the obtained classification accuracies by 12.83, 14.38, 22.7, 30.23 and 35.42% for WFMV, SFMV, ANN, SVM and CT respectively.



	500	400	300	200	100
CT	87.09	71.6	68.09	53.7	51.67
SVM	89.6	78.2	72.01	67.52	59.37
ANN	91.72	84.84	74.7	71.49	69.02
SFMV	92.92	85.49	82	79.74	78.54
WFMV	94.36	88.94	87.84	82.86	82.77

Figure 10: Performance evaluation of individual classifiers along with their combination using SFMV and WFMV on different data samples

Table 4: Classification accuracies obtained for different classifiers

classifiers	B		R		G		Average	SD
	UA	PA	UA	PA	UA	PA		
ANN	99.13	91.09	98.02	91.12	70.19	94.52	90.68	10.59
SVM	91.34	70.53	81.94	99.03	94.55	96.06	88.91	10.75
CT	99.99	85.28	96.15	85.85	58.75	94.38	86.73	14.89
SFMV	99.23	94.02	98.79	89.33	73.37	96.72	91.91	9.79
WFMV	98.15	95.38	87.99	96.75	99.41	94.62	95.38	4.02

4.4 Computational complexities

Another problem to be investigated is the computational cost associated with each model. In this regard, it is worth mentioning that the computer system used is of: Genuine Intel (R) CPU T2130, 1.86GHz and 783MHz, and 896 MB of RAM along with a test area of approximately one Km^2. In order to evaluate the performance of the proposed model, it has been compared with the three-member classifiers and SFMV. Table 5 shows the elapsed time in seconds during the combination process. In terms of individual classifiers, CT is the cheapest classifier with almost 12 second/km^2 processing time, followed by SVM and ANN with almost 15 and 17 second/km^2 respectively. The time required by the SFMV technique is almost 24 seconds/km^2 which is almost two times more than that required by the cheapest classifier. WFMV, on the other hand, is the most complex classifier with 27 second/km^2. However, it is still comparable with SFMV.

Table 5: Computational complexity comparison of WFMV-based fusion with individual classifiers and SFMV

classifier	CPU time (second/km^2)
ANN	17.082077
SVM	15.333318
CT	12.165660
SFMV	24.294089
WFMV	27.922584

5. Conclusions

In this paper, a MCS using ANN, SVM and CT has been proposed to produce land use/cover maps from 0.82m pan-sharpened IKONOS satellite imagery. ANN was the most accurate individual classifier (91.72%), followed by SVM (89.60%) and CT (87.09%). A modified weighted majority voting scheme, WFMV, has been applied for combining the results obtained for individual classifiers. The architecture of the WFMV has been obtained through two steps: 1) weighting the decisions of the member classifiers; and 2) setting the parameters of the WFMV model using a 10-fold cross-validation. The proposed model has been tested and evaluated considering four different aspects: 1) overall accuracy; 2) class accuracy; 3) sensitivity to training sample size; and 4) computational complexity. The results showed an improvement of about 3.88% in the classification accuracy over the best individual classifier. The application of WFMV resulted in an average class accuracy of 95.38% which is 4.7 and 3.5% better than the best individual classifier and SFMV respectively. Another advantage of WFMV fusion is that the obtained errors are less variable and class accuracies of the different classes are almost equal. In terms of computational complexities, WFMV is still comparable with SFMV. Although ANN can perform as well as WFMV when large training samples are available, WFMV performs much better than ANN in the case of small training samples. The proposed WFMV approach can be practically extended to integrate any number of member classifiers. As future work, Deep Learning Based (DLA) approach should be incorporated into the classification process to further improve the accuracy of the obtained results.

Acknowledgements

The author would like to acknowledge Giza Utility Data Centre, Giza Governorate, Egypt for the provision of the 0.82mpan-sharpened IKONOS image.

References

Ahn, H., M.J. Fazzari, N. Lim, J.J. Chen and R.L.Kodell (2007). Classification by ensembles from random partitions of high dimensional data. Computational Statistics and Data Analysis 51(12), 6166–6179.

Anderson, J.R., E.E. Hardy, J. T. Roach and R. E. Witmer (1976). A land use and land cover classification system for use with remote sensor data. Geological Survey Professional Paper No. 964, U.S. Government Printing Office, Washington DC, 28.

Breiman, L., J. Friedman, R. Olshenand and C. J. Stone (eds) (1984). Classification and regression trees. Chapman & Hall, New York, 358 p.

Cavalcanti, G., L. Oliveira, T. Moura and G. Carvalho (2016). Combining diversity measures for ensemble pruning. Pattern Recognition Letters, 74(2016), 38–45.

Chen, C.F. (2000) Fuzzy training data for fuzzy supervised classification of remotely sensed images, In: Proceedings of the 20th Asian Conference on Remote Sensing (ACRS 1999), 22-25 November 1999, Hong Kong Convention and Exhibition Centre, Venue, Hong Kong, China, pp. 460–465.

Chen, Y., P. Peng Dou and X. Xiaojun Yang (2017). Improving Land Use/Cover classification with a multiple classifier system using Ada Boost integration technique. Remote Sensing, 9(10), 1055; doi: 10.3390/rs9101055.

Clark labs (2012). Guide to IDRISI software [online]. Clark University. Available from: http://www.clarklabs.org/applications/upload/land change-modeler-IDRISI-focuspaper. pdf [Accessed 23 October 2018].

Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys and E. Lambin (2010). Digital change detection methods in ecosystem monitoring: A review, International Journal of Remote Sensing 2010(25), 1565–1596.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems, 2(4), 303–314.

Dai, L. and C. Liu (2010). Multiple classifier combination for land cover classification of remote sensing image. In: Proceedings of the 2010 2nd International Conference on Information Science and Engineering (ICISE), 4–6 December 2010 Hangzhou, China, pp. 3835–3839.

Dietterich, T. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, Machine Learning, 40, 139–157.

Du, P., S. Liu, J. Xia and Y. Zhao (2013). Information fusion techniques for change detection from multi-temporal remote sensing images, Information Fusion, 14(1), 19–27.

El-Melegy, M.T. and S.M. Ahmed (2007). Neural networks in multiple classifier systems for remote-sensing image classification. In: Nachtegael, M., Vin der Weken, D., Kerre, E.E. and Philips, W., ed. Soft Computing in Image Processing: Recent Advances, Volume 210 of Studies in Fuzziness and Soft Computing, pp. 65–96, Springer, Germany.

Fumera G., I. Pillai and F. Roli (2004). A two-stage classifier with reject option for text categorisation. In: Fred A., Caelli T.M., Duin R.P.W., Campilho A.C., de Ridder D., ed. Structural, Syntactic, and Statistical Pattern Recognition (SSPR/SPR 2004). Lecture Notes in Computer Science, 3138. Springer, Berlin, Heidelberg, pp. 771–779.

Ghimire, B., J. Rogan, V.F. Rodriguez-Galiano, P. Panday and N. Neeti (2012). An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. GIScience and Remote Sensing, 49(5), 623-643.

Ghosh, A. and B. Uma Shankar (2010). Neuro-fuzzy-combiner: an effective multiple classifier System. International Journal of Knowledge Engineering and Soft Data Paradigms, 2(2), pp. 107- 129.

Ghosh, A., S. K. Meher and B.U. Shankar (2008). A novel fuzzy classifier based on product aggregation operator, Pattern Recognition, 41(3), 961–971.

Han, M., X. Zhu and W. Yao (2012). Remote sensing image classification based on neural network ensemble algorithm, Neurocomputing, 78(1), 133–138.

Haralick, R.M. (1976). The table look-up rule, Communications in Statistics-Theory and Methods A5, pp. 1163–1191.

Hashem, S. and B. Schmeiser (1995). Improving model accuracy using optimal linear combinations of trained neural networks, IEEE Transactions on Neural Networks, 6(3), 792–794.

Haykin, S. (1998). Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice Hall.

Herrera, F. and J. L. Verdegay (1996). A Linguistic decision process in group decision making. Group Decision Negotiation, 5, pp. 165-176.

Hirose, Y., M. Mori, Y. Akamatsu and Y. Li (2004). Vegetation cover mapping using hybrid analysis of IKONOS data. ISPRS Archives, XXXV (B7), 12-23 July 2004 Istanbul, Turkey.

Ho, T.K., J.J. Hull and S.N. Srihari (1994). Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(1), 66–75.

Hsu, C.W., C.C. Chang and C.J. Lin (2009). A practical guide to support vector classification [online]. Department of Computer Science, National Taiwan University, Available from: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [Accessed 23 October 2018].

Hu, Y. (2000). Handbook of neural network signal processing, 1st ed., CRC Press, Inc., Boca Raton, FL, USA, 2000.

Kavzoglu, T. and P.M. Mather (2003). The use of back propagating artificial neural networks in land cover classification. International Journal of Remote Sensing, 24(3), 4907- 4938.

Khoshgoftaar, T., J. Van Hulse and A. Napolitano (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans (41), 552–568.

Khosravi, I. and M. Mohammad-Beigi (2014). Multiple classifier systems for hyperspectral remote sensing data classification. Journal of the Indian Society of Remote Sensing, 42(2), 423–428.

Kimura, F. and M. Shridhar (1991). Handwritten numerical recognition based on multiple algorithms, Pattern Recognition, 24(10), 969–983.

Kumar, D.A. and S.K. Meher (2013). Multiple classifiers systems with granular neural networks. In: Proceedings of the 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), 26–28 September 2013 Solan, India, pp. 1–5.

Kuncheva, L.I. (2000). Fuzzy Classifier Design, Springer-Verlag.

Kuncheva, L.I. (2004). Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience.

Lillesand, T. and R. Kiefer (2004). Remote sensing and image interpretation. Fourth Edition, John Willey & Sons, Inc., New York.

Lu, D. and Q. Weng (2007). A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28(5), 823 - 870.

Lv, Y., P.F. Shi and Y.M. Zhao (2000). Voting principle for combination of multiple classifiers. Journal of Shanghai Jiao Tong University, 34(5), 680-684.

Maulik, U. and D. Chakraborty (2010). A robust multiple classifier system for pixel classification of remote sensing images, Fundamental Informaticae, 101, 286–304.

Mazurov, V.D., A.I. Krivonogov and V.S. Kazantsev (1987). Solving of optimization and identification problems by the committee methods. Pattern Recognition, 20, 371–378.

Moustakidis, S., G. Mallinis, N. Koutsias and J.B. Theocharis (2012). SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images. IEEE Transaction on Geoscience and Remote Sensing, 50, 149–169.

Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization. In: B. Schölkopf,

C. J. C. Burges, and A. J. Smola, ed. Advances in Kernel Methods - Support Vector Learning, Cambridge, MA.: MIT Press, pp. 185-208.

Ponti-Jr, M. (2011). Combining Classifiers: from the creation of ensembles to the decision fusion. In: Proceedings of the 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, 28-30 August 2011, Alagoas, Brazil, pp. 1-10.

Ponti-Jr. M. and J. P. Papa (2011). Improving accuracy and speed of optimum-path forest classifier using combination of disjoint training subsets. In: Proceedings of the 10th International Workshop on Multiple Classifier Systems (MCS 2011) LNCS 6713, 15-17 June 2011, Naples, Italy, pp. 237–248.

Ranawana, R. and V. Palade (2006). Multi-classifier systems: review and a roadmap for developers. International journal of hybrid intelligent systems, 3(1), 35-61·

Rogova, G. (1994). Combining the results of several neural network classifiers. Neural Networks, 7(5), 777–781.

Russ, J.C. (2002). The image processing handbook, fourth edition. Boca Raton, FL: CRC Press.

Saheb, B., k. Subba rao and S. Phani kumar (2013). A survey on voting algorithms used in safety critical systems. International Journal of Engineering and Computer Science ISSN: 2319-7242, 2(7), 2272-2275.

Salah, M., J. Trinder and A. Shaker (2011). Performance evaluation of classification trees for building detection from aerial images and Lidar data: A Comparison of Classification Trees Models. International Journal of Remote Sensing, 32(20), 5757-5783.

Salah, M., J. Trinder, A. Shaker, M. Hamed and A. Elsagheer (2010). Integrating multiple classifiers with fuzzy majority voting for improved land cover classification. In: Paparoditis N., Pierrot-Deseilligny M., Mallet C., Tournaire O., ed. IAPRS, XXXVIII (3A), 1-3 September 2010, Saint-Mandé, France, pp. 7-12.

Segrera, S. and M. Moreno (2005). Multi classifiers: applications, methods and architectures. In: Proceedings of the International Workshop on Practical Applications of Agents and Multiagents Systems, IWPAAMS05, pp. 263–271.

Shannon, C. E. (editor) (1949). Reprinted 1998, The mathematical theory of communication, (Urbana, IL: University of Illinois Press), 324.

Tso, B. and P.M. Mather (2001). Classification methods for remotely sensed data, Taylor and Francis, London.

Tumer, K. and J. Ghosh (1996). Error correlation and error reduction in ensemble classifiers, Connection Science, 8(3), 385–204.

Van der Linden, S., A. Rabe, A. Okujeni and P. Hostert (2009). Image SVM classification, application manual: imageSVM version 2.0.

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer-Verlag, New York.

Wemecke, K.D. (1992). A coupling procedure for the discrimination of mixed data, Biometrics, 48, 497–506.

Woz´niak, M., M. Graña and E. Corchado (2014). A survey of multiple classifier systems as hybrid systems. Information Fusion, 16 (2014), 3–17.

Xu, L., A. Krzyzak and C.Y. Suen (1992). Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man, and Cybernetics, 22, 418–435.

Yager, R.R. (1998). On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Transactions on Systems, Man, and Cybernetics, 18, 183-190.

Yang, Z. (2007). An interval based attribute ranking technique. Unpublished report, ITT Visual Information Solutions.

Zadeh, L.A. (1965). Fuzzy sets. Information Control, 8, 338–353.

Zadeh, L.A. (1983). A computational approach to fuzzy quantifiers in natural languages, Computers and Mathematics with Applications, 9, 149-184.