

# Classification of major construction materials in construction environments using ensemble classifiers

Hyojoo Son<sup>a</sup>, Changmin Kim<sup>a</sup>, Nahyae Hwang<sup>a</sup>, Changwan Kim<sup>a,\*</sup>, Youngcheol Kang<sup>b</sup>

<sup>a</sup> Department of Architectural Engineering, Chung-Ang University, Seoul 156-756, Republic of Korea

<sup>b</sup> OHL School of Construction, Florida International University, Miami, FL 33174, USA

## ARTICLE INFO

### Article history:

Received 11 March 2013

Received in revised form 12 September 2013

Accepted 6 October 2013

Available online 8 November 2013

### Keywords:

Ensemble classifier

Color

Construction material detection

Data mining techniques

Image processing

## ABSTRACT

*Importancia*

The automatic detection of construction materials in images acquired on a construction site has been regarded as a critical topic. Recently, several data mining techniques have been used as a way to solve the problem of detecting construction materials. These studies have applied single classifiers to detect construction materials—and distinguish them from the background—by using color as a feature. Recent studies suggest that combining multiple classifiers (into what is called a heterogeneous ensemble classifier) would show better performance than using a single classifier. However, the performance of ensemble classifiers in construction material detection is not fully understood. In this study, we investigated the performance of six single classifiers and potential ensemble classifiers on three data sets: one each for concrete, steel, and wood. A heterogeneous voting-based ensemble classifier was created by selecting base classifiers which are diverse and accurate; their prediction probabilities for each target class were averaged to yield a final decision for that class. In comparison with the single classifiers, the ensemble classifiers performed better in the three data sets overall. This suggests that it is better to use an ensemble classifier to enhance the detection of construction materials in images acquired on a construction site.

© 2013 Elsevier Ltd. All rights reserved.

*Tema*

*Vacio de conocimiento*

*¿Cómo? Metodología*

*Conclusión*

*Resultado*

## 1. Introduction

The automatic detection of construction materials in images acquired on a construction site is essential for a wide range of construction applications, from the generation of a 3D as-built model to progress monitoring (see, for example, [1–5]). With the rapid deployment of image sensors on construction sites, images containing valuable project information are readily available. However, material detection in construction images is non-trivial and difficult. Construction materials in construction images may appear cluttered, occluded, or articulated, and the shapes and positions of construction materials are unpredictable.

Color has been recognized as an efficient feature for distinction of a material of interest from the background. Color has obvious advantages over other features such as texture and shape, especially in complex environments, as it is independent of the shapes and positions of the objects [6–8]. In addition, it is simple and computationally efficient to implement because it requires only the color values of each pixel in an image. It is therefore expected that material detection using color would be more robust and accurate and would overcome problems associated with construction environments compared to other features such as texture and shape.

In past research, color distribution has been used in efforts to detect construction materials in images acquired on a construction site. Neto et al. [9] proposed a method that employs edge detection and color to identify construction materials in images. In their approach, an edge detector algorithm detects edge pixels that belong to construction materials by comparing the RGB values in such pixels with the predetermined RGB values of the construction materials. After the edges have been detected, it groups the interior pixels to each set of edge pixels as an object. At the end of the operation, both of the resulting linked lists (the one for edges and the one for internal pixels) are stored under an object name. Zou and Kim [10] suggested a color-based method for identification of hydraulic excavators on a construction site. They use the hue feature to separate hydraulic excavators of different colors, and the saturation feature to differentiate each excavator of interest from its background, which consists of dark-colored soil and white snow. Simple thresholding methods using the hue and saturation features in conjunction with a method of calculating object centroid coordinates enable their system to produce accurate determinations of excavator idle time and working rate. The changing centroid coordinates of an excavator in successive images taken at constant time intervals are used as indicators of movement. Son and Kim [11] proposed an automated structural component recognition method that employs color and 3D data acquired from a stereo vision system for use in construction progress monitoring. The data

\* Corresponding author. Tel.: +82 2 820 5726; fax: +82 2 812 4150.

E-mail address: [changwan@cau.ac.kr](mailto:changwan@cau.ac.kr) (C. Kim).

processing first relies on color features to effectively extract information on structural components by employing color invariance, 2D object segmentation, and two-stage post-processing of removing unnecessary noise unrelated to the structure of interest and supplementing the image with information that may have been unintentionally eliminated. That information is then utilized to extract 3D coordinates for each color feature. The color image is used to guide the detection of features, while the 3D data are used to compensate for the pose of the feature.

In recent years, data mining methods such as artificial neural networks (ANNs), Gaussian mixture models (GMMs), and support vector machines (SVMs) have been investigated as a way to detect construction materials in images by use of a color model. Zhu and Brilakis [12] applied ANNs to classify regions of concrete in images acquired on a construction site. The images were first divided into regions through image segmentation using color. Then the color and texture features of each region were calculated, and the regions were classified using a pre-trained ANNs classifier. Son et al. [5] performed a comparative analysis of three data mining algorithms (GMMs, ANNs, and SVMs) for detection of concrete regions in images acquired on construction sites. The results show that the accuracy of the SVM they employed is better than that of the GMM or the ANN in dealing with concrete detection.

In previous studies, single classifiers have been employed to detect construction materials—and distinguish them from the background—by using color as a feature. However, a single classifier might not produce the optimal result in construction environments in which the color of one construction material is similar to that of others around it, or the inherent color property of a construction material is altered because of the effects of variation in illumination. For these reasons, the detection of construction materials still remains a challenging problem and there is still room for further improvement of detection performance. In order to solve such complex classification problems, heterogeneous ensemble classifiers (combined into a so-called multiple classifier system) have been proposed and they have been shown to be considerably successful in highly complex domains compared to ones with individual classifiers [13–17].

An ensemble classifier is comprised of a set of individual classifiers whose predictions are combined to obtain a highly accurate classification. Systems of this type have been proposed as a way to achieve better classification performance than with a single classifier [15,17] and are expected to reduce the variance in the estimation errors made by the individual classifiers [18,19]. The effectiveness of ensemble classifiers for detection of materials in complex environments has been demonstrated in various fields (see, for example, [20–24]). To the knowledge of the authors, the applicability of ensemble classifiers to the detection of construction materials has not been explored thus far.

### Objetiva

The aim of this study was to improve the accuracy of detection of major construction materials such as concrete, steel, and wood by using ensemble classifiers. It was hypothesized that ensemble classifiers achieve higher accuracy than single classifiers in detecting construction materials in construction environments. This belief is based on the general expectation that ensemble classifiers can outperform individual classifiers [17]. The rest of the paper is organized as follows. Section 2 describes data collection and pre-processing. Section 3 describes the methods employed by the single classifiers and the scheme employed in the proposed ensemble classifier. In Section 4, the results of experiments on the performance of the proposed ensemble classifier are compared with that of single classifiers in terms of average prediction accuracy. Finally, the conclusions are presented in Section 5.

### Hipótesis

## 2. Data collection and data pre-processing

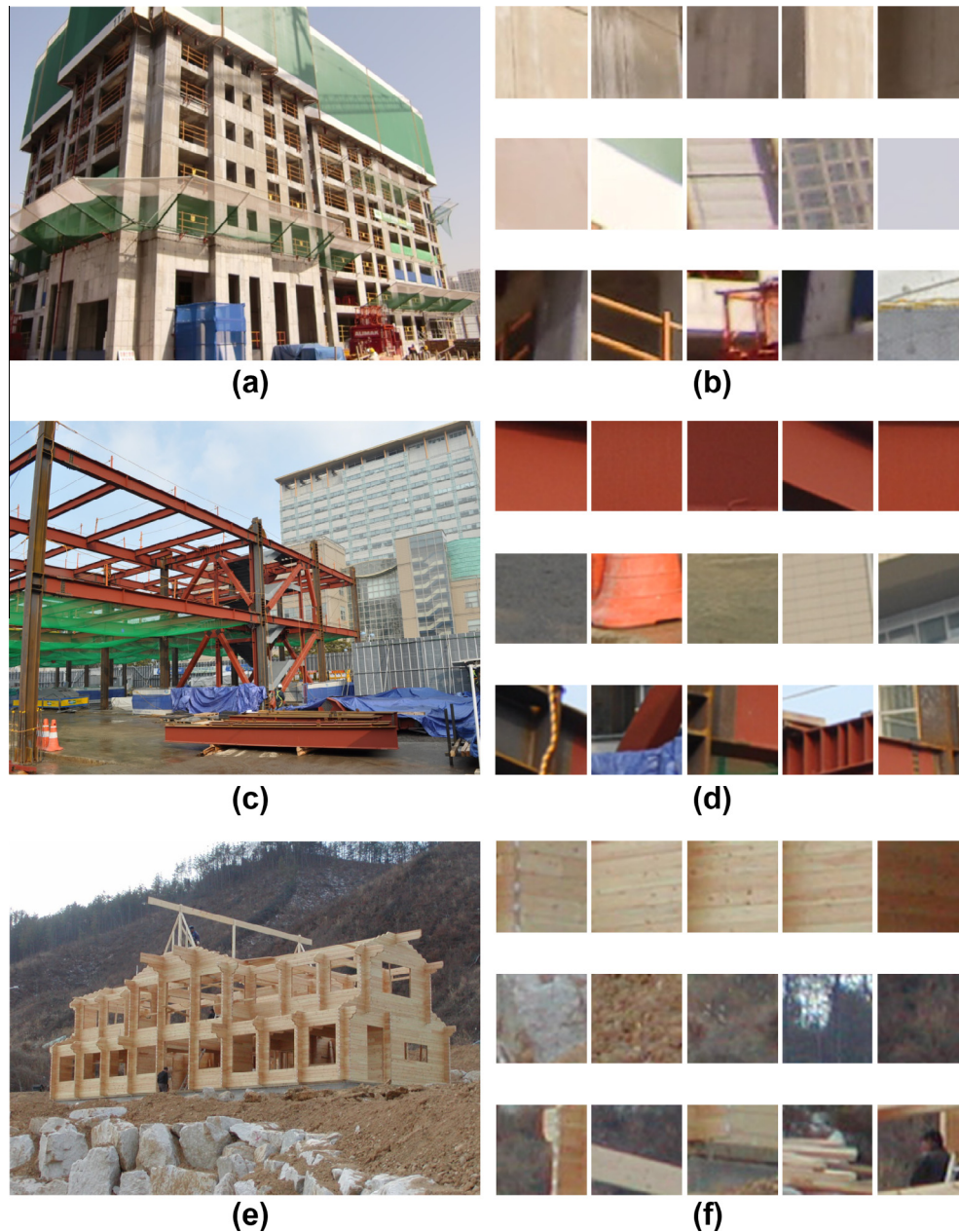
### 2.1. Data collection

Without a comprehensive data set, it cannot be concluded that an ensemble classifier yields better accuracy than single classifiers in detecting construction materials in construction environment. Because comprehensive data sets for construction material detection were not readily available, a total of three data sets (one each for concrete, steel, and wood) were generated.

The appearance of a construction material's surface colors can be affected by environmental factors such as changes in the direction and intensity of illumination. Since most construction sites are outdoors, the intensity of illumination varies unpredictably and uncontrollably, depending on the time of day, seasonal variations, and weather conditions (sunny, cloudy, or foggy), thereby resulting in large variations in the appearance of a construction material's surface colors. To account for such variations, 108 photographs were taken at a total of 50 construction sites for concrete detection, 91 photographs were taken at a total of 80 construction sites for steel detection, and 50 photographs were taken at a total of 14 construction sites for wood detection. Fig. 1(a), (c), and (e) present examples of construction site images for concrete, steel, and wood detection. Digital cameras with resolutions ranging from 3 megapixels to 12 megapixels were used when collecting data. The photographs were intended to contain either of the three structural components (concrete, steel, and wood) in images of actual construction-site scenes in order to validate the effectiveness and robustness of the proposed method for use in applications such as the generation of 3D as-built models and progress monitoring. Therefore, the photographs were taken at a distance from structures so that images contained the entire structures.

Each photograph was then divided into sub-regions of either  $25 \times 25$  pixels or  $50 \times 50$  pixels. Each sub-region was categorized and labeled as either a material of interest or the background, or as unable to say whether the sub-region was the material of interest or the background. Fig. 1(b), (d), and (f) display examples of the sub-regions for concrete, steel, and wood data sets. For example, the first, second, and third rows in Fig. 1(b) show the examples of sub-regions that were labeled as the concrete, the background, and unable to say whether the sub-region was the concrete or the background. Of the sub-regions, 48.2% from the concrete data set, 22.8% from the steel data set, and 49.0% from the wood data set were categorized as unable to say whether the sub-region was the material of interest or the background. These sub-regions were then excluded from the data set. As a result, every data set consisted of material or non-material pixels. The former are pixels associated with objects made of materials such as concrete, steel, and wood, while the latter are pixels related to the background. To assess whether the heterogeneous ensemble classifiers perform better than single classifiers, this study made a particular effort to collect and include as many materials as possible with color properties similar to those of the materials of interest. In each data set, the background included all kinds of scenery—bricks, construction equipment, fences, forms, pipes, safety nets, the sky, soil, traffic signs, trees, windows, and other construction-related materials. For example, the background of the concrete data set included objects made of materials such as steel and wood in order to evaluate results in the presence of different construction materials.

In total, the data collected from the concrete, steel, and wood sub-regions and their background sub-regions amounted to over 113 million pixels for concrete detection, 95 million pixels for steel detection, and 35 million pixels for wood detection. The first data set contained approximately 44 million pixels of concrete and 69 million pixels of background. The second data set consisted of



**Fig. 1.** (a) An example of construction site images for concrete detection, (b) examples of sub-regions for concrete detection, (c) an example of construction site images for steel detection, (d) examples of sub-regions for steel detection, (e) an example of construction site images for wood detection, and (f) examples of sub-regions for wood detection (? Note: The first, second, and third rows in this figure (b), (d), and (f) show the examples of sub-regions that were labeled as the material of interest, the background, and unable to say whether the sub-region was the material of interest or the background, respectively.).

9 million pixels of steel and approximately 85 million pixels of background. The third data set had 10 million pixels of wood and approximately 25 million pixels of background. The characteristics of the three data sets are provided in Table 1. In summary, the data sets were well balanced in terms of time of day, season, and weather. The percentages of data collected during a.m. and p.m. hours were roughly comparable. There was greater variation in the percentages of data collected in different seasons, as well as in the percentages of data collected and under different weather conditions, as shown in Table 1.

## 2.2. Data pre-processing

One issue that must be carefully considered in order to take full advantage of color-based material detection is how to effectively re-

duce the effect that variations in illumination which occur in outdoor environments will have on material color. **The values of colors in the RGB (red, green, and blue) color space, the most prevalent choice for computer graphics, are particularly subject to deterioration as a result of changes in illumination [25].** Such variations caused by factors in outdoor environments may dramatically affect color properties, potentially impacting detection performance [26]. To deal with these potential artifacts, it is important to represent color in a way that minimizes the effect of variations in illumination. In this study, in order to represent color in an appropriate form, **the color values in the RGB color space were converted to the HSI color space,** which separates chrominance and luminance information. The HSI color space consists of three components: hue ( $H$ ), saturation ( $S$ ), and intensity ( $I$ ). Hue and saturation are related to color, or chromaticity, and are illumination-independent components.



**Table 1**  
Distribution of data.

Data set			
Category	Concrete (%)	Steel (%)	Wood (%)
<i>Time of day</i>			
AM	56.9	33.0	30.0
PM	43.1	67.0	70.0
<i>Season</i>			
Spring	23.2	9.9	24.0
Summer	16.7	16.5	28.0
Fall	46.8	32.9	24.0
Winter	13.2	40.7	24.0
<i>Weather</i>			
Cloudy	13.1	18.7	44.0
Foggy	11.5	8.8	4.0
Sunny	75.4	72.5	52.0

These three components were used as input features to the classifiers. The HSI color space is defined as follows [27]:

$$H = \arctan\left(\frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}\right) \tag{1}$$

$$S = 1 - \frac{\min(R, G, B)}{I} \tag{2}$$

$$I = \frac{(R + G + B)}{3} \tag{3}$$

### 3. Research methodology

#### 3.1. Single classifiers

We considered six different types of classifiers: a support vector machine (SVM), an artificial neural network (ANN), commercial version 4.5 (C4.5), naïve Bayes (NB), logistic regression (LR), and *k*-nearest neighbors (KNN). In all of our experiments, we employed the algorithms from WEKA release 3.6.7 [28], which is a Java-based machine learning tool. Default parameter values in WEKA release 3.6.7 were used in these six classifiers. What follows is a brief description of these six classifiers.

##### 3.1.1. Support vector machine (SVM)

The SVM, inspired by statistical-learning theory, is one of the most powerful machine-learning techniques for solving a large number of complex binary classification problems [29]. It acts as a linear classifier in a high-dimensional feature space transformed through a projection from the original input feature space by using non-linear kernel functions of the original data set [30]. The high-dimensional space could theoretically be infinite, in which case linear discrimination is almost possible. In general, however, the resulting classifier is non-linear in the input space. An SVM learns the data in the high-dimensional feature space and achieves good generalization performance by finding a hyperplane that maximizes the margin between the classes with the help of multiplier parameters, for example, Lagrange multipliers [31]. For further details, see Vapnik [32] and Vladimir and Vapnik [33]. In this study, the radial basis function (RBF) was selected as the kernel function. With the RBF kernel function, the parameters *C* and  $\gamma$  were set to 1 and 0, respectively.

##### 3.1.2. Artificial neural network (ANN)

In this study, a back-propagation (BP) neural network is used for classification. This is a feed-forward network that can have one or more hidden layers. It utilizes an iterative gradient-search

technique designed to minimize the mean-square error between the actual and desired net outputs. The units in the hidden layers sum their inputs, add a constant, and take a fixed function of the result [34]. The output units are of the same form but with an output function. A three-layer network with one hidden layer was proven to be capable of computing any continuous likelihood function required in a classifier and of solving complex binary classification problems [35,36]. In this study, a three-layer network with one hidden layer was used for experimentation. It has been shown that a multiple hidden-layer neural network can approximate any function [37]. The number of nodes for the hidden layer was set to 3. The sigmoid transfer function was selected as the activation function. The learning rate, which determines the amount of weights that are updated, was set to 0.3; the momentum that was applied to the weights during updating was set to 0.2; and the training time, which indicates the number of epochs through which to train, was set to 500.

##### 3.1.3. Commercial version 4.5 (C4.5)

C4.5 is a supervised-learning algorithm that is the advanced version of the decision tree algorithm ID3 [38]. It is used to generate a set of rules from the data. As such, it employs a divide-and-conquer approach [39,40]. C4.5 works in three main steps. First, the node at the top of the tree (the “root node”) considers all samples and passes information about them to the “branch nodes” (the nodes that have at least one node of the tree below them). The branch nodes generate rules for a group of samples based on an entropy measure. At this stage of the process, C4.5 constructs a tree by considering all attribute values and finalizes the decision rule by pruning. It uses a heuristic approach for pruning based on the statistical significance of splits. After fixing the best rule, the branch nodes send the final target value to the lowest node on each branch, which is called a “leaf node” [38,41]. In this study, a confidence factor, which is used for pruning, was set to 0.25, and the minimum number of instances per leaf node was set to 2.

##### 3.1.4. Naïve Bayes (NB)

The NB classifier is a simple linear classifier that is based on the classical statistical “Bayes theorem.” The term “naïve” refers to the fact that it calculates the maximum posterior probability based on the assumption that the attributes of the training samples are independent and that there are no hidden or latent attributes which influence the prediction procedures [42]. This classifier is based on theoretically well-founded mathematical models used to predict an unseen case given a training sample [43]. The NB classifier will assign the most likely class to a given example as described by its feature vector. It assumes that the decision problem is posed in probabilistic terms and that all of the relevant probability values are known. This type of classifier is simple to implement, as neither numerical optimization nor matrix algebra is required for its use. NB is also efficient to train and use and is easy to update with new data.

##### 3.1.5. Logistic regression (LR)

The LR is a generalization of linear regression [44]. It is used primarily to predict binary or multi-class dependent variables. Because the response variable is discrete, it cannot be modeled directly by linear regression. Therefore, rather than predicting a point estimate of the event itself, it builds a model to predict the odds of its occurrence. In a two-class problem, odds greater than 50% would mean that the case is assigned to the class designated as 1; otherwise, it is assigned to the class designated as 0. Although logistic regression is a powerful modeling tool, it assumes that the response variable (which is the logarithm of the odds, not the odds per se) is linear in the coefficients of the predictor variables. Furthermore, the modeler, based on his or her experience with the

data and data analysis, must choose the right inputs and specify their functional relationship to the response variable.

### 3.1.6. *k*-Nearest neighbors (KNN)

KNN is an instance-based learning approach. It finds the *k*-nearest neighbors and uses a majority vote to determine the class label [45]. The basic principle of this algorithm is that each as-yet-unseen instance is always compared with existing instances using a distance metric. Based on the distance metric, the closest existing instance is used to assign the class for the test sample [28]. The most straightforward approach is to assign the majority class among the nearest neighbors to the query [46]. The advantage of the KNN is its robustness when using noisy training data sets [41,47]. The major drawback of the KNN algorithm is the high computational cost, as it has to compute the distance of every query instance for all the training data. In this study, Euclidean distance was used as the distance metric, and *k*, which indicates the number of neighbors to use, was set to 1.

### 3.2. Voting-based ensemble classifier

#### Indicadores

In this study, an ensemble classifier used a voting scheme which internally makes use of several single (base) classifiers and combines them in an effort to produce better results. Next, a determination was made of the single classifiers to choose for incorporation into a proposed ensemble classifier. The motivation for using a voting-based ensemble classifier was to take advantage of information outputs generated by a number of different base classifiers, instead of relying on the output of a single classifier alone [48]. The output generated by the ensemble classifier was defined as the average of the prediction probabilities generated by the base classifiers. This was done because there is a consensus in the literature that the classification performance of this type of heterogeneous ensemble classifier is better than that of even the best-performing individual classifier, and that biased decisions can be avoided more easily as a result [49,50].

When we design heterogeneous ensemble classifiers, we must select the subset of classifiers that can be combined to achieve better accuracy [51]. It is easy to see that such an optimal subset could be obtained by exhaustive enumeration, that is, by assessing on a validation set the classification accuracy provided by all possible subsets, and then choosing the subset exhibiting the best performance. In our case, there are six single classifiers; thus, the number of possible subsets is equal to 57. If the number of classifiers increases, the number of possible subsets also increases. Unfortunately, the evaluation of all possible combinations of base classifiers is impractical and requires high computational cost. Therefore, different strategies have been proposed in order to limit the computational complexity of the selection process [51]. Accordingly, measures for evaluating the base classifiers forming an ensemble classifier have been proposed for classifier selection purposes [52].

Generally speaking, to have an effective ensemble classifier that is more accurate than any of its base classifiers, one needs base classifiers which are diverse (in the sense that they predict differently) but accurate [23,53–56]. In other words, heterogeneous ensemble classifiers are more effective when the base classifiers are diverse and have acceptable individual performances, and consequently are more likely to have a more robust generalization performance in terms of the principle of bias–variance tradeoff [57,58]. If individual base classifiers are rather inaccurate, their combination is meaningless and might not produce an accurate ensemble classifier. On the other hand, if there is no diversity among the classifiers, (that is, if the base classifiers all produce essentially the same outputs), then their combination will not sig-

nificantly contribute to improved performance but will serve mainly to increase the complexity of the process.

It is rather difficult to achieve both high diversity and accuracy in base classifiers, since high diversity means substantial disagreement among the outputs of the base classifiers, and high accuracy indicates significant agreement among the outputs of the base classifiers [54,56,59]. The two key ingredients to achieve an effective ensemble classifier are thus in contradiction with each other. Therefore, it is crucial to select base classifiers with a good tradeoff between the two characteristics. In this study, the **weighted count of errors and correct results (WCEC)** was employed for this purpose. The WCEC is one of the commonly employed measures of diversity of errors.

#### 3.2.1. Selection of base classifiers for ensemble classifier

The weighted count of errors and correct results (WCEC) is one of a number of measures of diversity of errors that are in use in assessing the choice of base classifiers to be used in an ensemble classifier [52]. The measures of diversity of errors emphasize that differences within the errors made by the member classifiers truly affect performance [60,52]. In the previous literature (see, for example, [52,61]), the WCEC measure was proven to be effective in assessing the choice of base classifiers to be used in an ensemble classifier. The WCEC measure not only gives a measure of the correctness of classifiers but also considers the diversity between two classifiers [61]. By using the WCEC, we were able to select base classifiers that are both diverse as well as accurate from among a number of competing classifiers.

The main concept of the WCEC is that adopting an incorrect classifier generally has a negative effect on the ensemble classifier, whereas adopting the correct classifier is beneficial. The agreements generally have a significantly greater positive impact on the ensemble classifier's performance than the disagreements. Based on this concept, the resulting ensemble classifier often yields more accurate classification results than the single classifiers, because it aggregates the benefits of multiple classifiers [62].

The WCEC takes information on both incorrect results and correct results into account, with more emphasis placed on cases where the classifiers agree on the result, regardless of whether that result is correct or incorrect. One can subdivide the occurrences of the different combinations of outputs of the base classifiers as follows [52]:

$$WCEC_{a,b} = N^{11} + \frac{1}{2}(N^{01} + N^{10}) - N^{00}_{\text{different}} - 5N^{00}_{\text{same}}, \quad (4)$$

where *N* is the total number of instances,  $N^{11}$  is the number of instances in which both classifiers *a* and *b* are correct,  $N^{10}$  (resp.  $N^{01}$ ) is the number of instances in which just the first (resp. second) classifier is correct, and  $N^{00}$  is the number of instances in which both classifiers *a* and *b* are incorrect. Furthermore,  $N^{00}_{\text{same}}$  is the number of instances in which both classifiers *a* and *b* are incorrect and have the same output, and  $N^{00}_{\text{different}}$  is the number of instances in which both classifiers *a* and *b* are incorrect but have different outputs. This study deals with a binary classification problem; thus, the value of  $N^{00}_{\text{different}}$  is zero. Thus  $N = N^{11} + N^{10} + N^{01} + N^{00}$  and  $N^{00} = N^{00}_{\text{same}} + N^{00}_{\text{different}}$ . According to Aksela and Laaksonen [52], who proposed the WCEC measure, the weighting is arbitrary, and the presented values were chosen based on their suitability to penalize errors, and especially the identical errors. This study adopted the weight values, as suggested by Aksela and Laaksonen [52].

Using WCEC as the measure of diversity of errors in order to devise a diverse ensemble classifier, we trained six classifiers (SVM, ANN, C4.5, NB, LR, and KNN) on the three different data sets. Then the WCEC was computed for every combination of two to six of the

classifiers for each of the three data sets. For each combination of just two classifiers, the pairwise values were used as inputs to the WCEC. For combinations of three or more classifiers, the mean of all the pairwise values of the classifiers were used as inputs to the WCEC. Then the combination of classifiers that had the maximum value of the WCEC was chosen for the ensemble classifier.

### 3.2.2. Combination rule

The use of an ensemble classifier based on the voting scheme starts with training of multiple base classifiers (so-called member classifiers). The information from multiple base classifiers is then combined to produce a final decision. Here, the information from the base classifiers included their predictions (binary decisions) for each instance and their prediction probabilities for each target class. In the case of the concrete data set, for example, the information from the base classifiers included a prediction probability that an object is made of concrete a prediction probability that it is not made of concrete, and a decision on whether it is or is not made of concrete. The prediction probabilities convey not only prediction information but also confidence information [63]. Once the base classifiers were selected, based on the measure of the weighted count of errors and correct results, they were combined to produce a final decision (prediction) for each instance and prediction probabilities for each target class.

The average of the prediction probabilities for the different base classifiers was used as the prediction probability for the overall classification result. Use of the average of prediction probabilities, rather than the average of binary decisions, is known to have improved the performance of some ensemble classifiers [64]. In addition, it is a simple, low-cost, but effective approach that has been shown to be quite successful in various classification problems [65–67].

Let  $D = \{D_1, \dots, D_L\}$  denote the ensemble classifier, where  $L$  is the number of base classifiers of which it is composed. Let  $\Omega = \{\omega_1, \omega_2\}$  be the set of class labels. The individual outputs are estimates of the posterior probabilities; that is, the output  $d_{ij}(x)$  of classifier  $D_i$ , in support of the hypothesis that an instance  $x$  comes from class  $\omega_j$ , is an estimate of  $P(\omega_j|x)$ ,  $i = 1, \dots, L$ ;  $j = 1, 2$  [60]. In this study, the conditional probability  $\mu_j(x)$  that  $x$  is in class  $j$  was defined as the average of the posterior probabilities output by the base classifiers:

$$\mu_j(x) = \frac{1}{L} \sum_{i=1}^L d_{ij}(x) \quad (5)$$

Then instance  $x$  was assigned to the class with the larger of the values of  $\mu_1(x)$  and  $\mu_2(x)$ .

### 3.3. Evaluation

Various approaches have been suggested for evaluating the performance of classifiers. Six performance evaluation measures were employed in this study: accuracy, precision, sensitivity, specificity, area under the receiver operating characteristic curve (AUC), and overall average performance score (S). The first five measures have seen extensive use in evaluation of the performance of classification models [68–70]. They can be calculated by computing four quantities: the number of correctly predicted pixels of the material of interest (true positives,  $TP$ ), the number of correctly predicted pixels that belong to the background (true negatives,  $TN$ ), the number of pixels that were incorrectly assigned to the material of interest (false positives,  $FP$ ), and the number of pixels of the material of interest that were incorrectly assigned to the background (false negatives,  $FN$ ) [71]. The fifth measure (AUC) indicates the area under the receiving operating characteristic (ROC) curve. It is the most common quantitative index describing the ROC curve [72–

74]. In this study, AUC is calculated according to Hanley and McNeil [72]. The sixth measure (S) is calculated by averaging the values of the first five measures, which are given by the following formulas:

$$\text{Accuracy} = \left( \frac{TP + TN}{TP + TN + FP + FN} \right) \quad (6)$$

$$\text{Precision} = \left( \frac{TP}{TP + FP} \right) \quad (7)$$

$$\text{Sensitivity} = \left( \frac{TP}{TP + FN} \right) \quad (8)$$

$$\text{Specificity} = \left( \frac{TN}{FP + TN} \right) \quad (9)$$

For the evaluation of the classification performance of each of the seven classifiers (the six single classifiers and the ensemble classifier),  $k$ -fold cross validation was used. This method is known for its tendency to yield minimal bias and variance in comparison to all other validation methods, including the leave-one-out method [75]. Extensive studies on numerous data sets with different classifiers have demonstrated that 10-fold cross validation is optimal in terms of computation and estimation of error, and there is theoretical evidence backing this up [44]. Thus, 10-fold cross validation was used to assess the performance of the seven classifiers.

In 10-fold cross validation, the data set is split into 10 approximately equal “folds” (sets), with each fold in turn being used for testing and the remaining 9 folds being used for training. Thus every instance is used exactly once for testing. The cross validation estimate of the overall performance is then calculated by simply averaging the 10 individual performance evaluation measures.

## 4. Experimental results and discussion

### 4.1. Classification results

For an ensemble classifier to be considered superior to its base classifiers in the detection of construction materials, it should be verified that the ensemble classifier outperforms single classifiers in distinguishing those materials from others. Table 2 lists the statistical results regarding the effectiveness and usefulness of our ensemble classifier for the detection of concrete, steel, and wood. The results were generated by applying 10-fold cross validation to the six single classifiers on the three data sets (one each for concrete, steel, and wood) separately. Based on the results generated by the six single classifiers, the optimal diverse ensemble classifier for each of three types of material detection turned out to be the same one comprised of the SVM, C4.5, and KNN; this determination was made by choosing the combination with the maximum value the weighted count of errors and correct results. After six single classifiers' training via 10-fold cross validation was conducted, the selection process of base classifiers for an ensemble classifier using the value of the WCEC on the three data sets (one each for concrete, steel, and wood) took 2.57 s, 2.94 s, and 2.97 s, respectively.

Regardless of the type of classifier, the results imply that concrete, steel, and wood detection can be achieved using color information alone, ranging from 82.28% to 96.70% in accuracy, 74.38% to 96.00% in precision, 87.68% to 98.62% in sensitivity, 66.08% to 95.98% in specificity, 89.39% to 97.24% in AUC, and 82.60% to 96.41% in S. These results strongly indicate that color information is effective and useful in construction material detection for objects made of materials such as concrete, steel, and wood. On all three data sets, the ensemble classifier demonstrated the best

**Table 2**

Classification results for each data set with different methods.

Data set	Method	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)	S
Concrete	SVM	89.56	87.75	91.96	87.16	95.61	90.41
	ANN	87.08	86.30	88.15	86.01	92.94	88.10
	C4.5	89.47	86.84	93.05	85.90	93.54	89.76
	NB	84.20	79.66	91.87	76.54	90.66	84.59
	LR	82.46	79.39	87.68	77.24	91.48	83.65
	KNN	91.16	90.92	91.45	90.87	92.24	91.33
	Ensemble	92.64	90.42	95.38	89.90	97.00	93.07
Steel	SVM	94.18	91.13	97.90	90.47	96.65	94.07
	ANN	90.92	87.12	96.04	85.80	94.33	90.84
	C4.5	93.93	91.85	96.42	91.44	95.03	93.73
	NB	89.00	83.05	98.01	79.99	92.76	88.56
	LR	86.37	83.85	90.07	82.66	91.60	86.91
	KNN	96.18	96.00	96.39	95.98	96.63	96.24
	Ensemble	96.70	95.23	98.31	95.08	96.74	96.41
Wood	SVM	90.00	86.89	94.22	85.78	94.10	90.20
	ANN	89.76	85.34	96.01	83.51	90.96	89.12
	C4.5	90.39	87.66	94.03	86.76	92.84	90.34
	NB	82.28	74.38	98.49	66.08	91.78	82.60
	LR	87.64	80.86	98.62	76.66	89.39	86.63
	KNN	91.24	91.43	91.01	91.47	92.49	91.53
	Ensemble	92.19	88.73	96.67	87.72	97.24	92.51

performance in terms of accuracy, sensitivity, AUC, and S, except in terms of sensitivity on the wood data set, where LR fared best. In addition, KNN yielded better results in terms of precision and specificity on all three data sets (see Table 2). Among single classifiers, KNN in particular seemed to perform quite well alone except in terms of sensitivity and AUC on all three data sets. Although the differences in performance between the ensemble classifier and the single classifiers were not large, the ensemble classifier performed better overall than the single classifiers, especially in terms of accuracy, sensitivity, AUC, and S. These results confirm that an ensemble classifier can effectively deal with detection of construction materials using color information alone and improve classification performance relative to that of single classifiers.

In this study, the processing time incurred in classification by using the proposed ensemble classifiers was quantified. The test time for the proposed ensemble classifiers was 38.53 min, 37.44 min, and 41.12 min for concrete, steel, and wood detection, respectively. For applications such as the generation of 3D as-built models and progress monitoring, it is important to achieve accurate detection because it is a precedent process in such applications. Poor detection performance may negatively influence the quality of subsequent process steps. While some applications, such as automated equipment control and safety enhancement-related ones, would require less processing time, less than a second or several seconds for real-time and continuous updates. When we consider applications such as the generation of 3D as-built models and progress monitoring, a processing time of about 40 min is sufficient and applicable. The results demonstrate that the proposed method is appropriate to the applications by achieving accurate detection as well as reaching an appropriate balance between accuracy and processing time.

#### 4.2. Statistical comparison of classifiers for the three data sets

For all three data sets, McNemar's test was used to assess whether the ensemble classifier significantly outperformed the single classifiers. McNemar's test is a statistical process that can evaluate the significance of the difference in the classification performance between two models [76]. As a non-parametric test for two related samples, this test is particularly useful for detecting differences in outputs [77,78]. Table 3 shows the results of McNemar's test for statistical comparisons of the classification performance of the ensemble classifier to that of each of the six single classifiers on the three datasets. The numbers in parentheses are the *p*-values. If a *p*-value is less than 0.05, the difference in the classification performance between two classifiers is statistically significant [79]. As shown in Table 3, all of the *p*-values are 0 (to three decimal places). Therefore, according to McNemar's test, the ensemble classifier significantly outperformed the single classifiers.

Receiver operating characteristic (ROC) curves were used to visually compare the classification performances of the six single classifiers (SVM, ANN, C4.5, NB, LR, and KNN) to that of the proposed ensemble classifier. The ROC curve is a two-dimensional graph in which the sensitivity  $[TP/(TP + FN)]$ , is plotted on the vertical axis and the quantity  $1 - \text{specificity}$   $[FP/(FP + TN)]$  is plotted on the horizontal axis. Sensitivity is the fraction of pixels of the material of interest that are correctly predicted as pixels of the material of interest, and  $1 - \text{specificity}$  is the fraction of pixels of the background that are incorrectly predicted as pixels of the material of interest. A larger area under the ROC curve means better classification performance, because a larger value of sensitivity can be achieved for each value of  $1 - \text{specificity}$ . For each of the three data

**Table 3**McNemar values (and corresponding *p*-values) for pairwise comparison of performance.

	SVM	ANN	C4.5	NB	LR	KNN
Concrete	42.28	577.09	199.58	3944.50	883.91	2,602.55
Ensemble	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Steel	2,489.51	3,388.54	378.77	12054.26	715.45	1,915.70
Ensemble	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wood	18,488.79	12,176.59	7356.33	30,252.80	21,720.44	8133.74
Ensemble	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)



sets, ROC curves for the seven classifiers are shown in Fig. 2. The ROC curve for the ensemble classifier lies above the ROC curves for the single classifiers. This means that the area under the ROC curve for the ensemble classifier is larger than the areas under the ROC curves for the single classifiers. Furthermore, if one ROC curve completely encloses another ROC curve, the classifier that corresponds to the latter curve can be discarded [80]. Therefore, it has been clearly shown that better classification performance was achieved by the ensemble classifier than by the single classifiers.

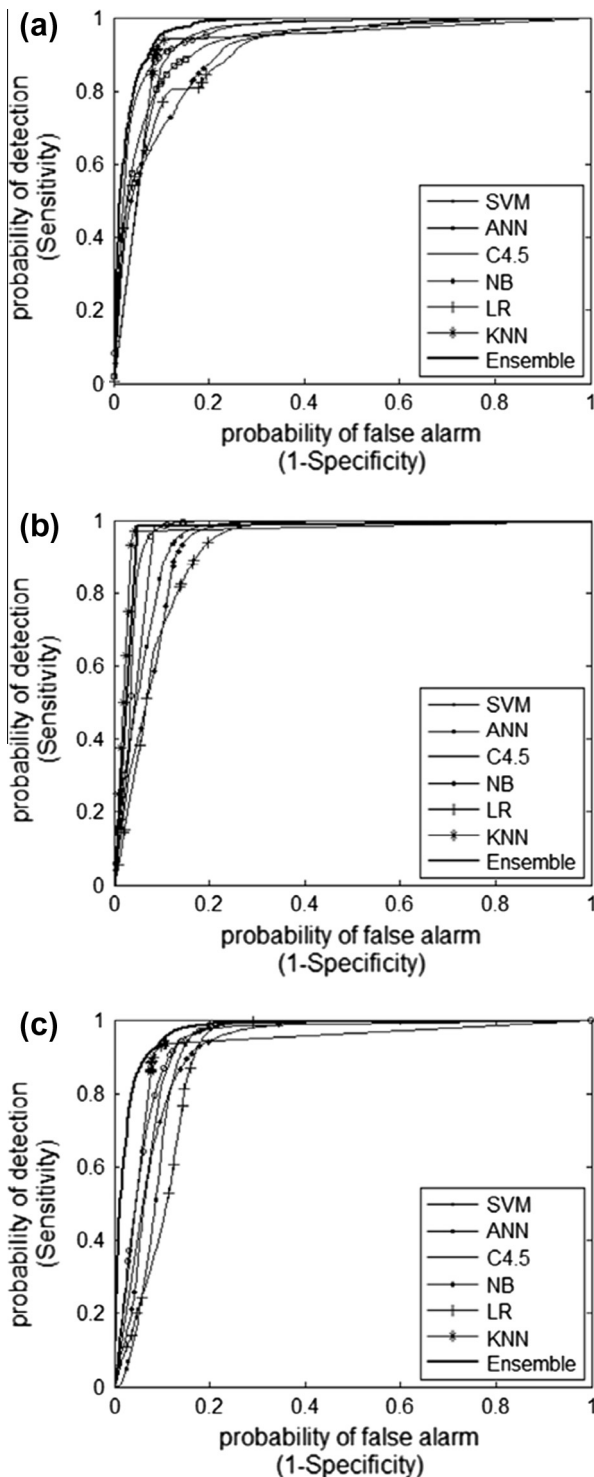


Fig. 2. ROC curves for detection of (a) concrete, (b) steel, and (c) wood.

## 5. Conclusions and further research directions

This paper proposed the use of a heterogeneous ensemble classifier to improve the detection accuracy of major construction materials such as concrete, steel, and wood on construction sites. In order to take advantage of single classifiers in designing the heterogeneous ensemble classifier and produce better results than any of the single classifiers, this study employed the diversity measure by taking into account the computational efficiency and the combination rule that we have presented. This study demonstrated that the heterogeneous ensemble classifier can be more accurate than an excellent single classifier in the detection of major construction materials; herein lies the main contribution of this study.

To validate the performance of the proposed ensemble model, an experimental comparison of the ensemble model and six single classification models (SVM, ANN, C4.5, NB, LR, and KNN) was carried out. The experimental results validated our initial hypothesis that the ensemble model will achieve greater accuracy than single classifiers in detecting construction materials in a construction environment. Thus, the proposed ensemble strategy can be useful in a variety of construction applications based on the use of images acquired on a construction site.

The results obtained with ensemble classifiers thus far are encouraging and suggest that it would be worthwhile to proceed with this line of research. In this study, the proposed ensemble classifier was validated on images of objects made of materials such as concrete, steel, or wood. In future work, the performance of the proposed ensemble classifier on images of objects made of other construction materials or composite materials made from two or more different types of materials will be investigated. Moreover, recent studies have shown that images acquired on a construction site can be used to detect defects in construction materials (see, for example, [1–5]). Further developments in the use of ensemble classifiers will be directed toward defect detection, such as crack detection in concrete and rust detection in steel. In addition, use of other types of ensemble classifiers, such as bagging classifiers and boosting classifiers, and ensemble classifiers comprised of different combinations of base classifiers other than those employed in this study, will be investigated in future studies.

## Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023229).

## References

- [1] M. Golparvar-Fard, F. Peña-Mora, C.A. Arboleda, S. Lee, Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs, *Journal of Computing in Civil Engineering* 23 (2009) 391–404.
- [2] Y.M. Ibrahim, T.C. Lukins, X. Zhang, E. Trucco, A.P. Kaka, Towards automated progress assessment of workpackage components in construction projects using computer vision, *Automation in Construction* 23 (2009) 93–103.
- [3] H. Son, C. Kim, C. Kim, A comparative study on color model-based concrete image retrieval in different invariant color spaces, in: *Proceedings of the 27th International Symposium on Automation and Robotics in Construction*, Bratislava, Slovakia, 2010, pp. 355–363.
- [4] I. Brilakis, S. German, Z. Zhu, Visual pattern recognition models for remote sensing of civil infrastructure, *Journal of Computing in Civil Engineering* 25 (2011) 388–393.
- [5] H. Son, C. Kim, C. Kim, Automated color model-based concrete detection in construction-site images by using machine learning algorithms, *Journal of Computing in Civil Engineering* 26 (2012) 421–433.
- [6] S.-F. Chang, J.R. Smith, Extracting multidimensional signal features for content-based visual query, in: *Proceedings of Visual Communications and Image Processing*, Taipei, Taiwan, 1995, pp. 995–1006.



- [7] A. Mojsilovic, J. Hu, A method for color content matching of images, in: *Proceedings of IEEE International Conference on Multimedia and Expo*, Yorktown Heights, NY, 2000, pp. 649–652.
- [8] M.S. Banu, K. Nallaperumal, Analysis of color feature extraction techniques for pathology image retrieval system, in: *Proceedings of IEEE International Conference on Computational Intelligence and Computing Research*, Virudhunagar District, India, 2010, pp. 1–7.
- [9] J.A. Neto, D. Ardit, M.W. Evens, Using colors to detect structural components in digital pictures, *Computer-Aided Civil and Infrastructure Engineering* 17 (2002) 61–67.
- [10] J. Zou, H. Kim, Using hue, saturation, and value color space for hydraulic excavator idle time analysis, *Journal of Computing in Civil Engineering* 21 (2007) 238–246.
- [11] H. Son, C. Kim, 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring, *Automation in Construction* 19 (2010) 844–854.
- [12] Z. Zhu, I. Brilakis, Parameter optimization for automated concrete detection in image data, *Automation in Construction* 19 (2010) 944–953.
- [13] T.K. Ho, J.J. Hull, S.N. Srihari, On multiple classifier systems for pattern recognition, in: *Proceedings of the 11th International Conference on Pattern Recognition*, Buffalo, NY, 1992, pp. 84–87.
- [14] L. Xu, A. Krzyzak, C.Y. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, *IEEE Transactions on Systems, Man Cybernetics* 22 (1992) 418–435.
- [15] T.K. Ho, J.J. Hull, S.N. Srihari, Decision combination in multiple classifier systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (1994) 66–75.
- [16] B. Achermann, H. Bunke, Combination of face recognition classifiers for person identification, in: *Proceedings of the 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 416–420.
- [17] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 20 (1998) 226–239.
- [18] L.I. Kuncheva, Using diversity measures for generating error-correcting output codes in classifier ensembles, *Pattern Recognition Letters* 26 (2005) 83–90.
- [19] P.R. Kumar, V. Ravi, Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review, *European Journal of Operational Research* 180 (2007) 1–28.
- [20] L. Breiman, Bagging predictors, *Machine Learning* 24 (1996) 123–140.
- [21] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Machine Learning* 29 (1997) 131–163.
- [22] E. Bauer, R. Kohavi, An empirical comparison of voting classification algorithms: bagging, boosting, and variants, *Machine Learning* 36 (1999) 105–139.
- [23] L.I. Kuncheva, C.J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Machine Learning* 51 (2003) 181–207.
- [24] L.I. Kuncheva, C.J. Whitaker, C.A. Shipp, R.P.W. Duin, Limits on the majority vote accuracy in classifier fusion, *Pattern Analysis and Applications* 6 (2003) 22–31.
- [25] B. Bascle, O. Bernier, V. Lemaire, Illumination-invariant color images correction, in: *Proceedings of the International Workshop on Intelligent Computing in Pattern Analysis/Synthesis*, Xi'an, China, 2006, pp. 359–368.
- [26] S. Doyle, M. Feldman, J. Tomaszewski, A. Madabhushi, A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies, *IEEE Transactions on Biomedical Engineering* 59 (2012) 1205–1218.
- [27] H.D. Cheng, X.H. Jiang, Y. Sun, J.L. Wang, Color image segmentation: advances and prospects, *Pattern Recognition* 34 (2001) 2259–2281.
- [28] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, second ed., Morgan Kaufmann, San Francisco, 2005.
- [29] C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning* 20 (1995) 273–297.
- [30] D. Meyer, Support vector machines, *R News* 1 (2001) 23–26.
- [31] S. Ali, K.A. Smith, On learning algorithm selection for classification, *Applied Soft Computing* 6 (2006) 119–138.
- [32] V. Vapnik, The support vector method of function estimation, in: J.A.K. Suykens, J. Vandewalle (Eds.), *Nonlinear Modeling: Advanced Black-Box Techniques*, Kluwer Academic, Dordrecht, The Netherlands, 1998, pp. 55–85.
- [33] V.N. Vladimir, V. Vapnik, *The Nature of Statistical Learning Theory*, first ed., Springer, New York, 1995.
- [34] B.D. Ripley, *Pattern Recognition and Neural Networks*, first ed., Cambridge University Press, Cambridge, 1996.
- [35] R.P. Lippmann, An introduction to computing with neural nets, *IEEE ASSP Magazine* 4 (1987) 4–22.
- [36] W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S*, fourth ed., Springer, New York, 2002.
- [37] B. Irie, S. Miyake, Capabilities of three-layered perceptrons, in: *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, CA, 1988, pp. 641–648.
- [38] J.R. Quinlan, *C4.5 Programs for Machine Learning*, first ed., Morgan Kaufman, San Mateo, 1993.
- [39] J.R. Quinlan, *C5.0 and See5: Illustrative Examples*, RuleQuest Research, 1997. <<http://www.rulequest.com>>.
- [40] Y.M. Chae, S.H. Ho, K.W. Cho, D.H. Lee, S.H. Ji, Data mining approach to policy analysis in a health insurance domain, *International Journal of Medical Informatics* 62 (2001) 103–111.
- [41] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
- [42] G.H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, in: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, San Mateo, CA, 1995, pp. 338–345.
- [43] I. Rish, An empirical study of the naïve Bayes classifier, in: *Proceedings of the IJCAI-01 Workshop on Empirical Methods in AI*, Seattle, WA, 2001, pp. 41–46.
- [44] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, first ed., Springer, New York, 2001.
- [45] C.C. Holmes, N.M. Adams, A probabilistic nearest neighbour method for statistical pattern recognition, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (2002) 295–306.
- [46] P. Cunningham, S.J. Delany, k-Nearest Neighbour Classifiers, Technical Report UCD-CSI-2007-4: University College of Dublin, 2007.
- [47] N. El Gayar, F. Schwenker, G. Palm, A study of the robustness of KNN classifiers trained using soft labels, *Artificial Neural Networks in Pattern Recognition* 4087 (2006) 67–80.
- [48] J. Sun, H. Li, Financial distress prediction using support vector machines: ensemble vs. individual, *Applied Soft Computing* 12 (2012) 2254–2265.
- [49] S. Hou, K. Ramani, Classifier combination for sketch-based 3D part retrieval, *Computers and Graphics* 31 (2007) 598–609.
- [50] H. Zhou, X. Li, D. Huang, Y. Yang, F. Ren, Voting-based ensemble classifiers to detect hedges and their scopes in biomedical texts, *IEICE Transactions on Information and Systems* E94-D (2011) 1989–1997.
- [51] F. Roli, G. Giacinto, G. Vernazza, Methods for designing multiple classifier systems, *Multiple Classifier Systems* 2096 (2001) 78–87.
- [52] M. Aksela, J. Laaksonen, Using diversity of errors for selecting members of a committee classifier, *Pattern Recognition* 39 (2006) 608–623.
- [53] D. Ruta, B. Gabrys, Classifier selection for majority voting, *Information Fusion* 6 (2005) 63–81.
- [54] J.J. Rodriguez, L.I. Kuncheva, C.J. Alonso, Rotation forest: a new classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006) 1619–1630.
- [55] S. Peddabachigari, A. Abraham, C. Grosan, J. Thomas, Modeling intrusion detection system using hybrid intelligent systems, *Journal of Network and Computer Applications* 30 (2007) 114–132.
- [56] Y. Zhu, J. Liu, S. Chen, Semi-random subspace method for face recognition, *Image and Vision Computing* 27 (2009) 1358–1370.
- [57] L. Yu, K.K. Lai, S. Wang, W. Huang, A bias-variance-complexity trade-off framework for complex system modeling, in: *Proceedings of the International Conference on Computational Science and Its Applications*, Glasgow, United Kingdom, 2006, pp. 518–527.
- [58] S. Bian, W. Wang, On diversity and accuracy of homogeneous and heterogeneous ensembles, *International Journal of Hybrid Intelligent Systems* 4 (2007) 103–128.
- [59] G. Yu, G. Zhang, Z. Yu, C. Domeniconi, J. You, G. Han, Semi-supervised ensemble classification in subspaces, *Applied Soft Computing* 12 (2012) 1511–1522.
- [60] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, first ed., John Wiley & Sons Inc., Hoboken, 2004.
- [61] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, S. Liu, Multiple classifier system for remote sensing image classification: a review, *Sensors* 12 (2012) 4764–4792.
- [62] L.A. Alexandre, A.C. Campilho, M. Kamel, On combining classifiers using sum and product rules, *Pattern Recognition Letters* 22 (2001) 1283–1289.
- [63] A.K. Seewald, How to make stacking better and faster while also taking care of an unknown weakness, in: *Proceedings of the 19th International Conference on Machine Learning*, San Francisco, CA, 2002, pp. 554–561.
- [64] D.M.J. Tax, M.V. Breukelen, R.P.W. Duin, J. Kittler, Combining multiple classifiers by averaging or by multiplying?, *Pattern Recognition* 33 (2000) 1475–1485.
- [65] G. Giacinto, R. Perdisci, M.D. Rio, F. Roli, Intrusion detection in computer networks by a modular ensemble of one-class classifiers, *Information Fusion* 9 (2008) 69–82.
- [66] R. Perdisci, D. Ariu, P. Fogla, G. Giacinto, W. Lee, McPAD: a multiple classifier system for accurate payload-based anomaly detection, *Computer Networks* 53 (2009) 864–881.
- [67] A. Agrawal, S. Misra, R. Narayanan, L. Polepeddi, A. Choudhary, Lung cancer survival prediction using ensemble data mining on SEER data, *Scientific Programming* 20 (2012) 29–42.
- [68] C. Ferri, J. Hernández-Orallo, R. Modroiu, An experimental comparison of performance measures for classification, *Pattern Recognition Letters* 30 (2009) 27–38.
- [69] M.-H. Horg, Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ML, RBFNN and SVM classifiers, *Expert Systems with Applications* 37 (2010) 4146–4155.
- [70] Y.S. Kim, Performance evaluation for classification methods: a comparative simulation study, *Expert Systems with Applications* 37 (2010) 2292–2306.
- [71] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Information Processing and Management* 45 (2009) 427–437.
- [72] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiving operating characteristic (ROC) curve, *Radiology* 143 (1982) 29–36.
- [73] D.R. Lovell, C.R. Dance, M. Niranjan, R.W. Prager, Limits on the Discrimination Possible with Discrete Valued Data, with Application to Medical Risk Prediction, first ed., Cambridge University Press, Cambridge, 1996.
- [74] B. Farina, C. Bartoli, A. Bono, A. Colombo, M. Lualdi, Multispectral imaging approach in the diagnosis of cutaneous melanoma: potentiality and limits, *Physics in Medicine and Biology* 45 (2000) 1243–1254.

- [75] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the International Joint Conference on Artificial Intelligence, Montréal, Canada, 1995, pp. 1137–1143.
- [76] C.-M. Tan, Y.-F. Wang, C.-D. Lee, The use of bigrams to enhance text categorization, *Information Processing and Management* 38 (2002) 529–546.
- [77] K.-J. Kim, Financial time series forecasting using support vector machines, *Neurocomputing* 55 (2003) 307–319.
- [78] J.K. Bae, J. Kim, Integration of heterogeneous models to predict consumer behavior, *Expert Systems with Applications* 37 (2010) 1821–1826.
- [79] X. Huang, L. Zhang, An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, *IEEE Transactions on Geoscience and Remote Sensing* 51 (2013) 257–272.
- [80] D.M.W. Powers, The problem of area under the curve, in: Proceedings of the 2012 International Conference on Information Science and Technology, Wuhan, China, 2012, pp. 567–573.