Automation in Construction 24 (2012) 52–60

Contents lists available at SciVerse ScienceDirect

Automation in Construction

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

# Concrete compressive strength analysis using a combined classification and regression technique

Jui-Sheng Chou [a,*], Chih-Fong Tsai [b]

[a] Department of Construction Engineering, National Taiwan University of Science and Technology, 43 Sec.4, Keelung Road, Taipei, Taiwan
[b] Department of Information Management, National Central University, 300 Jhongda Road, Jhongli City, Taoyuan County, Taiwan

ARTICLE INFO

*Article history:*
Accepted 16 February 2012
Available online 13 March 2012

*Keywords:*
High performance concrete
Compressive strength
Hierarchical data mining
Classification
Regression

ABSTRACT

High performance concrete (HPC) is a complex composite material, and a model of its compressive strength must be highly nonlinear. Many studies have tried to develop accurate and effective predictive models for HPC compressive strength, including linear regression (LR), artificial neural networks (ANNs), and support vector regression (SVR). Nevertheless, in accordance with recent reports that a hierarchical structure outperforms a flat one, this study proposes a hierarchical classification and regression (HCR) approach for improving performance in predicting HPC compressive strength. Specifically, the first-level analyses of the HCR find exact classes for new unknown cases. The cases are then entered into the corresponding prediction model to obtain the final output. The analytical results for a laboratory dataset show that the HCR approach outperforms conventional flat prediction models (LR, ANNs, and SVR). Notably, the HCR with a 4-class support vector machine in the first level combined with a single ANNs obtains the lowest mean absolute percentage error.



## 1. Introduction

High-performance concrete (HPC) has good workability, high strength and low permeability. The literature shows that permeability is directly related to long-term durability. In fact, long-term durability is more important than high strength in most field applications [29]. Producing HPC requires highly specific component materials, mixture proportions, and placement and curing conditions. The required workability is normally attained by using superplasticizers. Since HPC production does not require exotic materials or complex procedures, most concrete producers are capable of producing HPC. Moreover, certain properties of HPC, which are not fully understood, are unlike those of conventional concrete.

Since the relationships between components and concrete properties are highly nonlinear, mathematically modeling HPC is challenging. That is, traditional models of concrete properties are inadequate for analyzing HPC compressive strength. The main goal of a modeling system that uses large experimental datasets for different concrete mixes is to properly reflect the very nature of physical phenomena such as concrete compressive strength. In such cases, most modeling techniques use experimental data for mathematical modeling with analytical linear or nonlinear functions.

Studies of prediction problems generally apply single (flat) prediction models. However, numerous studies show that a hierarchical structure outperforms a flat structure for solving not only classification problems [8,28,33], but also regression problems [1,13,32,38]. A literature search shows no hierarchical solutions to this domain problem. Moreover, hierarchical estimation approaches proposed in related studies have only proposed various combinations of existing regression models.

This study proposes a hybrid approach that combines classification and regression techniques hierarchically, namely hierarchical classification and regression (HCR), to solve the problem of estimating concrete compressive strength. The first level of the HCR classifies a new unknown case into a specific class such as high or low concrete compressive strength (CCS). This case is then inputted into an estimator trained to predict only high or low CCS before making the final prediction.

The proposed HCR approach applies the divide-and-conquer principle of solving complex problems by first solving subproblems (*i.e.,* simpler tasks) with a classifier in the first level of HCR and with an estimator in the second level of HCR. To verify and validate the proposed HCR model, the prediction performance is compared against single linear regression (LR), artificial neural networks (ANNs), and support vector regression (SVR).

The rest of this paper is organized as follows. Section 2 briefly reviews the related literature, including concrete compressive strength estimation and some well-known estimation techniques. Section 3 presents the proposed hierarchical classification and regression approach. Section 4 gives the experimental results based on real world data. Finally, Section 5 concludes the study and suggests future research directions.

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
*E-mail addresses:* jschou@mail.ntust.edu.tw, rayson.chou@gmail.com (J.-S. Chou), cftsai@mgt.ncu.edu.tw (C.-F. Tsai).

0926-5805/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.autcon.2012.02.001

## 2. Literature review

### 2.1. Concrete compressive strength estimation

Because of its extreme complexity, modeling HPC is difficult. Predicting the mechanical properties of construction materials, such as the 28-day compressive strength of concrete, is an important line of research in material science [31]. As traditional methods perform poorly for modeling HPC and other complex non-linear and uncertain materials, researchers are constantly seeking better prediction tools.

Boukhatem et al. (2011) comprehensively reviewed recent developments in information technology and their applications on concrete mix design [3]. The authors considered simulation models, decision support systems, artificial intelligence, and fuzzy logic useful and powerful because these tools enable to solve complex problems in concrete technology. They concluded that the continuous development of universal systems is needed so as to predict properties of concrete mixing to achieve required performance.

In particular, many studies have proposed the use of ANNs and their variations for mapping non-linear relationships among factors affecting 28-day HPC compressive strength. For instance, Oh et al. (1999) applied ANNs to optimize the proportion of four concrete ingredients (water, cement, fine aggregate and coarse aggregate). They introduced a tool for minimizing uncertainty and errors when calculating the proper proportions of concrete mixes, which is a complex, time-consuming, and uncertain task [27].

Similarly, Mostofi and Samaee (2001) used multi-layer perceptron ANNs to estimate HPC compressive strength [25]. Meanwhile, Yeh (1998) obtained promising results with their proposed model of HPC strength using ANNs as a function of cement, fly ash, blast furnace slag, water, superplasticizer, coarse aggregate, fine aggregate, and age of testing [36].

Kasperkiewicz and Dubrawski (1995) further used fuzzy-ARTMAP ANNs to predict the 28-day compressive strength of HPC with varying mixtures of its six components: (cement, silica, superplasticizer, water, and fine and coarse aggregate) [18]. Fazel Zarandi et al. (2008) combined fuzzy neural networks and polynomial neural networks into six different fuzzy polynomial neural network architectures, each of which had six input parameters (concrete ingredients) and one output parameter (28-day compressive strength of the mix-design) [9].

In Alilou and Teshnehlab's work (2010) [2], they adopted feed-forward neural network to predict the 28-day strength of the concrete using a set of input concrete parameters. Interestingly, in addition to identify the most important parameters, including mix design, water/cement ratio, density, slump, air, silica fumes, super-plasticizer, and age, they found the 3-day strength parameter of concrete an critical index for estimating the 28-day strength of the concrete. By early predicting 28-day strength of the concrete with the 3-day index, the designer and constructor could possibly reduce the duration of the civil project execution.

Based on the concrete parameters determined by non-destructive tests, Hola and Schabowicz (2005) assessed concrete compressive strength using ANNs [14]. Their work indicated Levenberg-Marquardt neural network gives the superior root mean square errors, the differences between the errors, relative testing error, relative error standard deviation, and correlation. In the authors' opinion, one can reliably identify the compressive strength of similar concretes in building structures with a set of data acquired by means of at least non-destructive techniques using the proposed neural network models.

Another interesting research direction is Genetic Operation Tree (GOT) [37], which combines operation tree and genetic algorithm to produce self-organized formulas automatically for predicting HPC compressive strength. Comparison results showed that GOT was more accurate than nonlinear regression formulas but less accurate than neural network models.

Most studies have applied ANNs techniques with only minor modifications, their variations or traditional regression techniques. Therefore, a hybrid model for predicting HPC compressive strength is still needed. Clearly, such a model must not only meet modeling requirements, but must also be sufficiently robust to model involved uncertainties and must be easy to manipulate. Nevertheless, the prediction performance of most of the above approaches have demonstrated either insufficient prediction performance for further generalization due to moderate correlation coefficient and small dataset, or insufficient measures of prediction accuracy, such as mean absolute percentage error or root mean square error. This study combined hierarchical classification and predictive techniques to model HPC compressive strength as a function of its major ingredients and to improve prediction performance.

### 2.2. Estimation techniques

Estimation models or estimators are needed to infer the value of unknown parameters in statistical models. Restated, models are needed to estimate parameter values based on measurement data. The estimator uses measured data as input for parameter estimation. Estimation models such as supervised prediction models learn from a set of training examples that includes output attributes with continuous values [35].

The following subsections describe three typical methods of numeric output estimation: linear regression, artificial neural networks, and support vector regression.

#### 2.2.1. Linear regression

Statistical models of the relationship between dependent variables (response variables) and independent variables (explanatory variables) can be developed using linear regression [26]. The general formula for multiple regression models is

$$Y = \beta_0 + \sum_{j=1}^{n} \beta_j X_j + \varepsilon \quad (1)$$

where $Y$ is a dependent variable, $\beta_0$ is a constant, $\beta_j$ is a regression coefficient ($j = 1, 2, \ldots, n$), and $\varepsilon$ is an error term. Simple regression analyses use only one independent variable ($X_j$) while multiple regression analyses use two or more variables.

A linear regression assumes the following [26]: (a) Existence: For a unique combination of independent variables $X_1, X_2, X_3 \ldots, X_j$, the variable $Y$ (single variable) is a random variable with a fixed average and variance value in specific probability distributions; (b) Independence: Observed values for $Y$ are statistically independent and are not correlated with each other; (c) Linearity of relationships: The average value of the variable $Y$ is a function as the linear combination of $X_1, X_2, X_3 \ldots, X_j$; (d) Homogeneity of variance: The variance in the dependent variable obtained through the linear combination of $X_1, X_2, X_3 \ldots, X_j$ is a constant; and (e) Normality: Normality is obtained by distributing the dependent variable by the linear combination of $X_1, X_2, X_3 \ldots, X_j$.

#### 2.2.2. Artificial neural networks

Artificial neural networks consist of information-processing units similar to neurons in the human brain except that a neural network consists of artificial neurons [12]. Neural networks learn by experience, generalize from previous experiences to new ones, and make decisions. An artificial neural network is a group of neural and weighted nodes, each representing a brain neuron, and the connections among these nodes are analogous to the synapses connecting brain neurons.

Particularly, multilayer perceptron (MLP) network is the standard neural network model. In a MLP network, the input layer contains a set of sensory nodes as input nodes, one or more hidden layers contain computation nodes, and an output layer contains computation nodes. The input nodes/neurons are the feature values of an instance, and the output nodes/neurons function as discriminators between the class of the instance and those of all other instances.

In the multilayer architecture, input vector $x$ passes through the hidden layer of neurons in the network to the output layer. The weight connecting input element $i$ to hidden neuron $j$ is denoted by $W_{ji}$, and the weight connecting hidden neuron $j$ to output neuron $k$ is denoted by $V_{kj}$. The net input of a neuron is obtained by calculating the weighted sum of its inputs, and its output is determined by applying a sigmoid function. Therefore, for the $j$-th hidden neuron

$$net_j^h = \sum_{i=1}^{N} W_{ji}x_i \text{ and } y_i = f\left(net_j^h\right) \tag{2}$$

while for the $k$-th output neuron

$$net_k^o = \sum_{j=1}^{J+1} V_{kj}y_i \text{ and } o_k = f\left(net_k^o\right) \tag{3}$$

The sigmoid function $f(net)$ is the logistic function

$$f(net) = \frac{1}{1+e^{-\lambda net}} \tag{4}$$

where $\lambda$ controls the gradient of the function.

For a given input vector, the network produces an output $o_k$. Each response is then compared to the known desired response of each neuron $d_k$. The weights of the network are then continuously modified to correct or reduce errors until the total error from all training examples is maintained below a pre-defined tolerance level.

For the output layer weights $V$ and the hidden layer weights $W$, the update rules are given by the 5th and 6th equations, respectively.

$$V_{kj}(t+1) = v_{kj}(t) + c\lambda(d_k - o_k)o_k(1-o_k)y_j(t) \tag{5}$$

$$W_{ji}(t+1) = w_{ji}(t) + c\lambda^2 y_j\left(1-y_j\right)x_i(t)\left(\sum_{k=1}^{K}(d_k-o_k)o_k(1-o_k)v_{kj}\right) \tag{6}$$

*2.2.3. Support vector regression*

Support vector machines (SVMs), which were introduced by Vapnik (1998) [34], perform binary classification, *i.e.*, they separate a set of training vectors for two different classes $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$ where $x_i \in R^d$ denotes vectors in a $d$-dimensional feature space and $y_i \in \{-1, +1\}$ is a class label. The SVM model is generated by mapping the input vectors onto a new higher dimensional feature space denoted as $\Phi{:}R^d \rightarrow H^f$ where $d < f$. Then, an optimal separating hyperplane in the new feature space is constructed by a kernel function $K(x_i, x_j)$, which is the product of input vectors $x_i$ and $x_j$ and where $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$.

The SVM can solve both classification and regression problems. Support vector regression (SVR) [30] is a modification of SVM in which a linear model is constructed in the new higher dimensional feature space. The mathematical notation for the linear model (in the feature space) $f(x, \phi)$ is

$$f(x,\phi) = \sum_{j=1}^{m} \phi_j g_j(x) + b \tag{7}$$

where $g_j(x)$, $j = 1, \ldots, m$ denotes a set of nonlinear transformations, and $b$ is the 'bias' term. Estimation quality is also indicated by loss of function $L(y, f(x, \phi))$ where

$$L_\varepsilon(y, f(x,\omega)) = \begin{cases} 0 & if\,|y-f(x,\omega)| \le \varepsilon \\ |y-f(x,\omega)|-\varepsilon & otherwise \end{cases} \tag{8}$$

The objective of support vector regression (SVR) is to compute a linear regression function for the new higher dimensional feature space using $\varepsilon$-insensitive loss while simultaneously reducing model complexity by minimizing $||\phi||^2$. This task can be described by introducing (non-negative) slack variables $\xi_i$, $\xi_i^\phi$ $i = 1\ldots n$, to measure the deviation in training samples outside the $\varepsilon$-insensitive zone. The SVR can therefore be formulated as a minimization of the following function:

$$\begin{aligned} &\min \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &\text{s.t.} \begin{cases} y_i - f(x_i,\omega) \le \varepsilon + \xi_i^* \\ f(x_i,\omega) - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots, n \end{cases} \end{aligned} \tag{9}$$

This optimization problem can be transformed into the dual problem, which is solved by

$$f(x) = \sum_{i=1}^{n_{sv}} (\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) \text{ subject to } 0 \le \alpha_i^* \le C,\ 0 \le \alpha_i \le C \tag{10}$$

where $n_{SV}$ is the number of support vectors and the kernel function is

$$K(\mathbf{x}, \mathbf{x}_i) = \sum_{j=1}^{m} g_j(\mathbf{x})g_j(\mathbf{x}_i) \tag{11}$$

## 3. Hierarchical classification and regression approach

Fig. 1 shows that the proposed HCR approach is a two-level hierarchical structure for predicting HPC compressive strength. The first level of the HCR uses a standard technique for classifying a new unknown case into one of $k$ classes, where $k$ can be fixed as two classes for high/low compressive strength, three classes for high/middle/low compressive strength, *etc*. For the second level of HCR, $k$ regression models are constructed for the prediction task. Thus, after receiving input data classified into one of $k$ classes by the first level of HCR, the input assignment module applies this data as input to its corresponding regression model (*i.e.*, one specific regression model in the second level of HCR) to produce the final output. The section below compares prediction performance between cross-fold and random sampling method for different numbers of classes.

*3.1. Constructing classification and regression models*

The first step of HCR is determining the number of classes (*i.e.*, $k$) needed to construct the classification model in the first level of HCR. Given a training dataset $D$, let $X_i$ be the $i$-th data sample in $D$, where $X_i = (X_1, X_2, \ldots, X_m, Y_{X_i})$ contains $m$ different input variables and $Y_{X_i}$ is the output variable of $X_i$, *i.e.*, the actual concrete compressive strength (CCS). A threshold $T$ is then defined to group the data samples into their corresponding subsets by

$$T = \frac{Y_{\max} + Y_{\min}}{k} \tag{12}$$

where $k$ is the number of classifications and $Y_{\max}$ and $Y_{\min}$ are the data samples in which the output variables are the maximum and minimum CCS, respectively, in $D$.

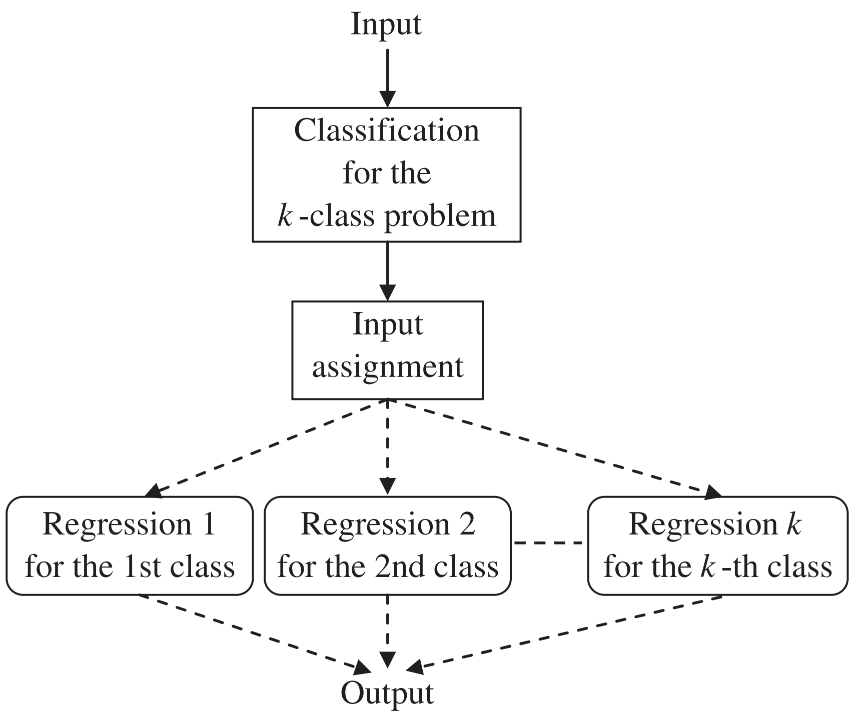


**Fig. 1.** HCR architecture.

Therefore, for the two-class problem, the data samples in $D$ can be classified into class 1 and class 2 subsets according to the following rule:

$$\begin{aligned}&\text{If } Y_{X_i} \le T, \text{ then } X_i \text{ belongs to the Class 1 subset;}\\&\textit{otherwise}, X_i \text{ belongs to the Class 2 subset}\left(i.e., Y_{X_i} > T\right).\end{aligned} \tag{13}$$

Data samples in the three-class problem can be classified into class 1, class 2, and class 3 subsets according to the following rule:

$$\begin{aligned}&\text{If } Y_{X_i} \le T, \text{ then } X_i \text{ belongs to the Class 1 subset;}\\&\text{if } T < Y_{X_i} \le 2T \text{ then } X_i \text{ belongs to the Class 2 subset;}\\&\textit{otherwise}, X_i \text{ belongs to the Class 3 subset}\left(\text{i.e., } 2T < Y_{X_i} \le 3T\right)\end{aligned} \tag{14}$$

Similarly, this rule can be extended to larger numbers of classes. Note that this rule applies even if subset sizes differ.

Once the training dataset $D$ is divided into $k$-class subsets, $Y_{X_i}$ (the original output of $X_i$) is replaced by a discrete class label corresponding to its class subset. This new training dataset $D'$ is then used to construct a specific classifier. The number of data samples and the input variables are identical in $D'$ and $D$, and the output variable of each data sample is a discrete class label.

When constructing the $k$ regression models, however, the Class $j$ subset is used to train the $j$-th regression models ($j=1, \ldots, k$), where the output variable of each data sample in the Class $j$ subset is the actual CCS.

### 3.2. Prediction process

After the first level classifier and the second level regression models are trained and constructed, they can predict the CCS needed for new unknown cases. Fig. 2 shows the pseudo code for the prediction process.

## 4. Experiments

The WEKA (Waikato Environment for Knowledge Analysis) suite [17] is used to test various training techniques and algorithms using the experimental dataset. The WEKA provides various filters for pre-

INPUT: $T_i$ is the $i$-th data sample in the testing dataset where $T_i = (T_1, T_2, \ldots, T_m)$ contains $m$ input variables

OUTPUT: $Y_{T_i}$ is the concrete compressive strength, in MPa

1. $T_i$ is fed into classifier $C$
2. $C$ classifies $T_i$ into the $s$-th class, where $s$ is one of the $k$ trained classes in training dataset $D$
3. The input assignment module assigns $T_i$ as the input to the $s$-th regression model $S$, where $S$ is one of the $k$ trained regression models
4. $T_i$ is fed into $S$
5. $S$ produces $Y_{T_i}$

**Fig. 2.** Pseudo code for predicting unknown cases.

processing, cross-fold and random sampling, model evaluation, visualization and post-processing. The objective of the experiments is to identify the hierarchical and regression techniques with the best prediction performance.

### 4.1. Experimental setup

#### 4.1.1. Dataset description and preparation

The experimental dataset was obtained from a University of California, Irvine data repository [36]. A final set of 1030 samples of ordinary Portland cement containing various additives and cured under normal conditions was obtained from various university research labs [5–7,10,11,15,16,20–24]. All tests were performed on 15-cm cylindrical specimens of concrete prepared using standard procedures. Table 1 shows the experimental dataset of nine attributes used in this study.

Datasets described in the literature often contain unexpected inaccuracies. For instance, the class of fly ash may not be indicated. Another problem is that superplasticizers produced by different manufacturers may have different chemical compositions [18]. Concrete compressive strength is determined not only by w/c ratio, but also by other materials used in the mix. Concrete contains five ingredients other than cement and water. The multiple ingredients, in addition to the nonlinearity of concrete structures, complicate the computation of compressive strength. Thus, the above predictive techniques were applied to the experimental dataset when modeling HPC compressive strength.

#### 4.1.2. Cross-validation

When comparing the predictive accuracy of two or more methods, researchers often use K-fold cross-validation to minimize bias associated with the random sampling of the training and holdout data samples. Since cross-validation requires random assignment of individual cases into distinct folds, a common practice is stratifying the folds themselves. In stratified K-fold cross-validation, the proportions of predictor labels (responses) in the folds are intended to approximate those in the original dataset. Empirical studies show that, compared to regular K-fold cross-validation, stratified cross-validation tends to reduce bias in the comparison results.

Kohavi (1995) showed that ten folds were optimal (*i.e.,* performed validation testing in the shortest time and with acceptable bias and variance) [19]. Thus, to assess model performance, a stratified 10-fold cross-validation approach was used. The entire dataset was divided into ten mutually exclusive subsets (or folds) with class distributions approximating those of the original dataset (stratified). The subsets were extracted in five steps:

1. Randomize the dataset.
2. Extract one tenth of the original dataset size from the randomized dataset (single fold).
3. Remove the extracted data from the original dataset.
4. Repeat steps 1-3 eight times.
5. Assign the remaining portion of the dataset to the last fold (10th fold).

**Table 1**
Attributes of concrete compressive strength.

| Attribute | Unit | Minimum | Maximum | Average | Standard Deviation |
|---|---|---|---|---|---|
| Cement | kg/m$^3$ | 102.0 | 540.0 | 281.168 | 104.506 |
| Blast Furnace Slag | kg/m$^3$ | 11.0 | 359.4 | 107.277 | 61.884 |
| Fly Ash | kg/m$^3$ | 24.5 | 200.1 | 83.862 | 39.989 |
| Water | kg/m$^3$ | 121.8 | 247.0 | 181.567 | 24.354 |
| Superplasticizer | kg/m$^3$ | 1.7 | 32.2 | 8.486 | 4.037 |
| Coarse Aggregate | kg/m$^3$ | 801.0 | 1145.0 | 972.919 | 77.754 |
| Fine Aggregate | kg/m$^3$ | 594.0 | 992.6 | 773.580 | 80.176 |
| Age of testing | Day | 1.0 | 365.0 | 45.662 | 63.170 |
| Concrete Compressive Strength | Mega Pascal (MPa) | 2.3 | 82.6 | 35.818 | 16.706 |

**Table 2**
MAPE and RMSE for LR, MLP, and SVR.

| | SVR | LR | MLP |
|---|---|---|---|
| MAPE | 30.95% | 31.55% | 14.67% |
| RMSE | 10.48 | 10.43 | 5.91 |

This procedure was first used to obtain ten distinct folds. Each fold was used once for performance tests of the single flat and hierarchical prediction models, and the remaining nine folds were used for training, which obtained ten independent performance estimates. The cross-validation estimate of overall accuracy was calculated by simply averaging the K individual accuracy measures for cross-validation accuracy (*CVA*), where K is the number of folds used and $A_i$ is the accuracy measure of each fold.

$$CVA = \sum_{1}^{K} \frac{A_i}{K} \tag{15}$$

#### 4.1.3. Single and hierarchical prediction models

The prediction models obtained by the proposed HCR approach are compared with three single flat regression models for baseline comparisons of performance in predicting HPC compressive strength. The regression models include linear regression (LR), multilayer perceptron (MLP) and support vector regression (SVR).

The SVM is used to develop the first level of hierarchical prediction models based on the HCR approach since SVMs have proven effective in many pattern recognition problems and since they provide better generalisation compared to many other classification techniques [4]. Specifically, the number of classes is set to 2, 3, 4, and 5 in the first level classification. Moreover, the radial basis function (RBF) is used as the kernel function, and the gamma value is set as 1.

The three different prediction models used for the second level of HCR obtained three different hierarchical prediction models for use in comparisons: SVM + LR, SVM + MLP, and SVM + SVR.

#### 4.1.4. Evaluation methods

To assess prediction performance in the three constructed prediction models, mean absolute percentage error (MAPE) and root mean squared error (RMSE) are calculated.

- Mean absolute percentage error (MAPE), a statistical measure of predictive accuracy, is usually expressed as a percentage. The MAPE is widely used for quantitative forecasting because it indicates relative overall fit (goodness-of-fit). The MAPE is given by the following equation:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y - y^{\wedge}}{y} \right| \tag{16}$$

**Table 3**
MAPE in hierarchical prediction models.

| | SVM + SVR | SVM + LR | SVM + MLP |
|---|---|---|---|
| 2-class SVM | 27.9% | 29.07% | 14.14% (1) |
| 3-class SVM | 25.14% | 26.81% | 17.76% (2) |
| 4-class SVM | 22.94% (5) | 23.19% | 19.65% (3) |
| 5-class SVM | 23.16% | 23.44% | 20.5% (4) |

**Table 4**
RMSE in hierarchical prediction models.

| | SVM + SVR | SVM + LR | SVM + MLP |
|---|---|---|---|
| 2-class SVM | 9.76 | 9.94 | 5.75 (1) |
| 3-class SVM | 9.67 | 9.92 | 7.97 (2) |
| 4-class SVM | 9.469 (5) | 9.471 | 8.53 (3) |
| 5-class SVM | 9.52 | 9.52 | 8.78 (4) |

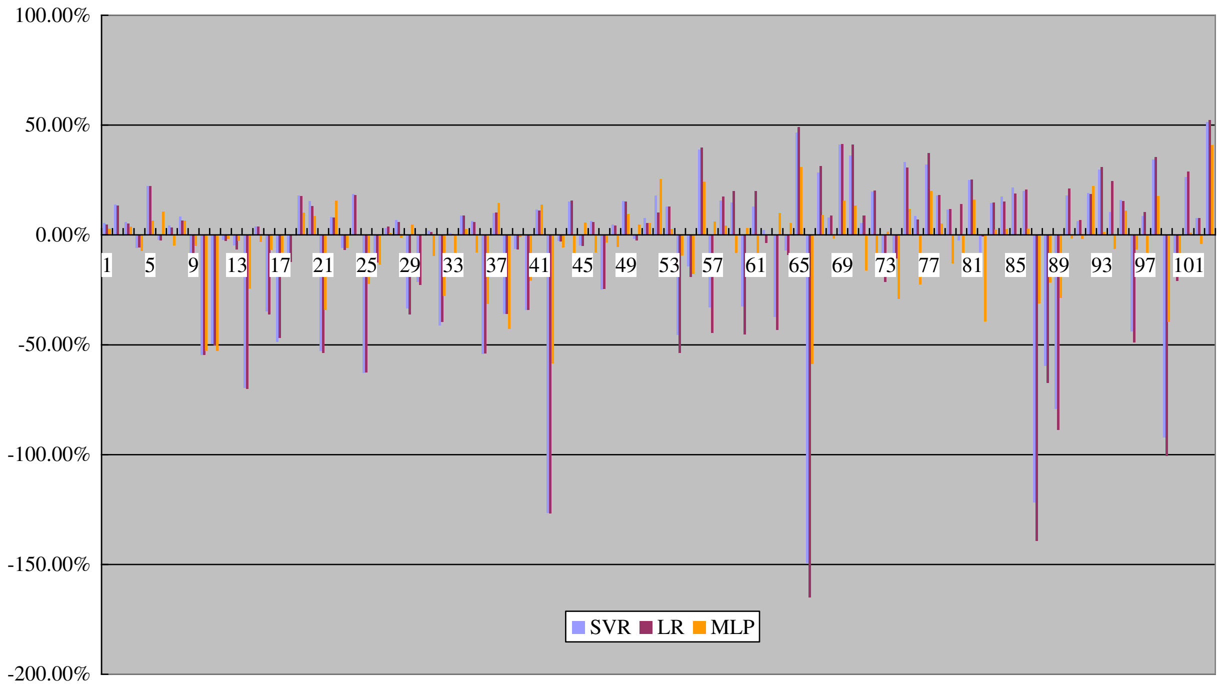


**Fig. 3.** MAPE of 2-class SVM + LR, MLP, and SVR.

- Root mean squared error (RMSE) is the square root of the mean square error, *i.e.*, the average distance of a data point from the fitted line measured along a vertical line, and is calculated by the following equation:

$$RMSE = \sqrt{\frac{\sum (y^{\wedge} - y)^2}{n}} \tag{17}$$

*4.2. Experimental results*

*4.2.1. Single prediction models*

The first experiment compared prediction performance in the three single flat regression models, *i.e.*, LR, MLP, and SVR. Table 2 shows MAPEs and RMSEs for the three models.

The comparison results show that MLP generally outperformed SVR and LR in terms of MAPE and RMSE. Therefore, MLP was set as the baseline for predicting concrete compressive strength.

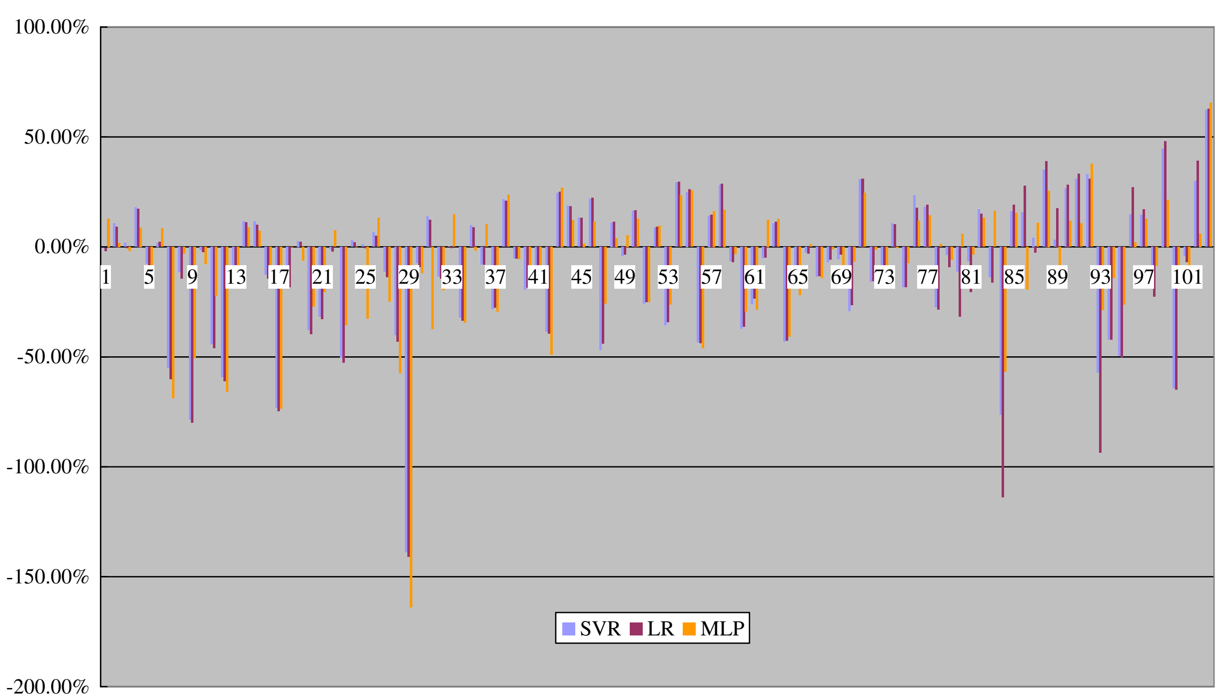


**Fig. 4.** MAPE of 3-class SVM + LR, MLP, and SVR.

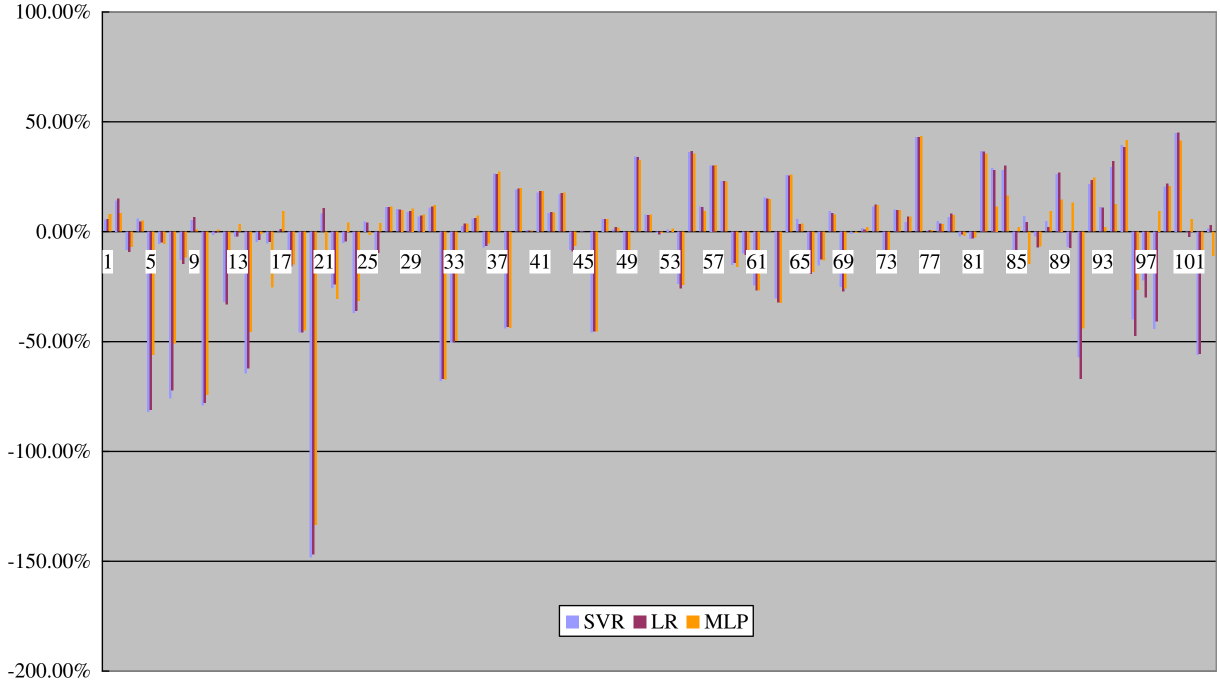


**Fig. 5.** MAPE of 4-class SVM + LR, MLP, and SVR.

Interestingly, however, although SVMs have proven effective in many pattern recognition problems [4], SVR performed only slightly better than LR does in terms of MAPE.

#### 4.2.2. Hierarchical prediction models

The second experiment compared prediction performance in hierarchical prediction models in which SVM is trained by different numbers of classes in the first level of HCR and in which LR, MLP, and SVR are the regression models in the second level of HCR.

Tables 3 and 4 compare the average MAPE and RMSE rates, respectively, for different hierarchical prediction models. The two tables show the five best models in terms of prediction performance. The bracketed numbers indicate the performance ranking.

The comparison results showed that SVM + MLP has the lowest MAPE and RMSE when 2-class classification is used to construct SVM and the regression model is based on MLP. Interestingly, SVM + SVR and SVM + LR have similar MAPE and RMSE.

Moreover, 103 data samples (*i.e.*, 1/10 of the original dataset) were randomly selected from the case dataset for use as another

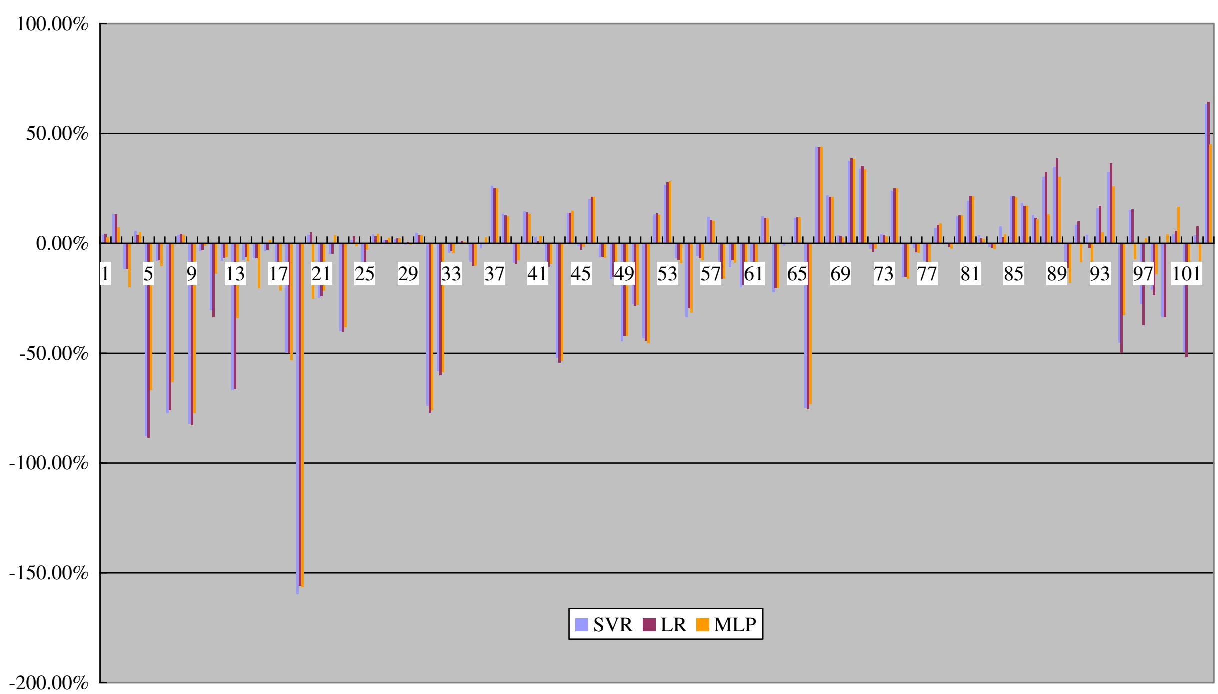


**Fig. 6.** MAPE of 5-class SVM + LR, MLP, and SVR.

**Table 5**
MAPE for 103 randomly selected test data items.

| | SVM + SVR | SVM + LR | SVM + MLP |
|---|---|---|---|
| 2-class SVM | 6.22% (5) | 6.92% | 4.69% (2) |
| 3-class SVM | 8.22% | 9.39% | 7.74% |
| 4-class SVM | 5.26% (3) | 5.39% (4) | 3.62% (1) |
| 5-class SVM | 7.62% | 7.78% | 7.19% |

testing dataset for MAPE comparisons in the hierarchical prediction models. Figs. 3–6 show the results for the hierarchical prediction models based on 2- to 5-class SVMs, respectively. Shorter bars represent better performance, *i.e.*, lower rates of MAPE.

Table 5 shows that MAPE was lower in hierarchical prediction models than in single flat regression models when analyzing the 103 randomly selected test data items. In this case, 4-class SVM + MLP performance was best, and 2-class SVM + MLP performance was second best. Moreover, using 4-class SVM enabled the combined SVR and LR models to achieve the lowest MAPEs in their respective categories (*i.e.*, SVM + SVR and SVM + LR, respectively).

*4.2.3. Comparisons between single and hierarchical models*

Since MLP is the best single flat regression model and SVM + MLP is the best hierarchical prediction model, Fig. 7(a) and (b) compare their predictive accuracy in terms of MAPE and RMSE values, respectively. Similarly, Fig. 7(c–f) compare the MAPE and RMSE values between SVR/LR and SVM + SVR/LR, respectively.

Although 2-class SVM + MLP revealed consistently superior MAPE and RMSE in all tests, it performed only slightly better than the single MLP model did. That is, 2-class SVM + MLP and single MLP perform comparably in predicting concrete compressive strength in terms of MAPE and RMSE. The MAPE differed by only 0.53% between 2-class SVM + MLP and single MLP (*i.e.*, 14.14% vs. 14.67%, respectively) whereas RMSE differed by only 0.16 (*i.e.*, 5.75 vs. 5.91, respectively). Notably, 4-class SVM + MLP and 2-class SVM + MLP achieved even lower MAPEs over the 103 test data items (3.62% and 4.69% respectively).

For SVM + SVR and SVM + LR, however, regardless of the number of classes used to construct SVM in first-level HCR, the hierarchical models clearly outperform the single flat regression models in predicting concrete compressive strength. Specifically, 4-class SVM + SVR and 4-class SVM + LR obtain the lowest MAPE and RMSE, respectively.

To sum up, the prediction models based on the proposed HCR approach had the lowest MAPEs and RMSEs. Restated, prediction models using SVM for first-level classification are superior to those that use single flat regression alone. That is, SVM + MLP, SVM + SVR, and SVM + LR outperform single MLP, SVR, and LR, respectively. Specifically, when analyzing the concrete compressive strength, 2-class SVM + MLP *via* 10-fold cross validation and 4-class SVM + MLP *via* randomly selected data items perform best.

## 5. Conclusion

This study developed a hierarchical approach using a combined classification and regression technique to predicting compressive strength for high performance concrete (HPC). The experimental

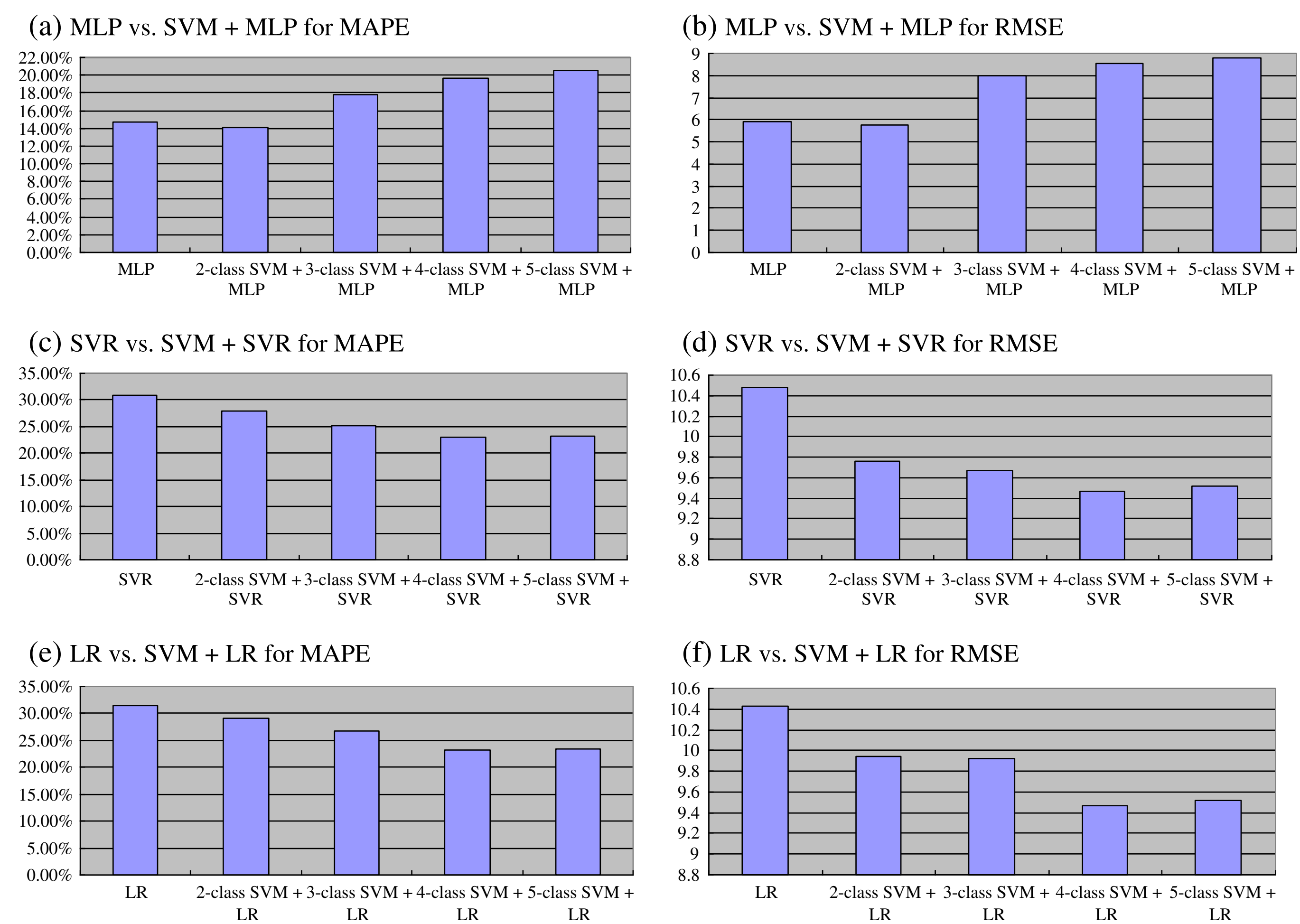


**Fig. 7.** Comparisons of MAPE and RMSE values in terms of prediction performance.

dataset was acquired from a UCI machine learning repository (http://archive.ics.uci.edu/ml/) of a 1030-instance dataset. The predictive techniques proposed in this study were applied to the prepared data by using 9 of the 10 folds for model training and the remaining fold for validation. This procedure was repeated ten times using a different fold as the validation dataset in each experiment.

After cross-fold training to validate the hierarchical classification-regression (HCR) approach, 1/10 of the original dataset were randomly selected from the case dataset for use as test data for MAPE comparisons in the hierarchical prediction models. The comparison results show that the proposed HCR approach to developing the prediction models outperforms single flat regression models in terms of MAPE and RMSE indicators. Specifically, the 4-class SVM classifier for the first level of HCR combined with MLP as the regression model for the second level of HCR (*i.e.,* 4-class SVM + MLP) performs best.

The contribution of this paper to the domain knowledge is to propose and validate the novel HCR technique and through which automate concrete mix design for compressive strength in civil infrastructure and building construction. Several issues remain for further study. The first is the applicability of the HCR approach in different engineering prediction problems, which requires sensitivity analyses over varying dataset sizes and varying numbers of variables. Moreover, other advanced classification techniques such as classifier ensembles can be used for first-level HCR. Similarly, combining different regression models in second-level HCR may improve prediction performance.

## References


[1] C.H. Achen, Two-step hierarchical estimation: beyond regression analysis, Political Analysis 13 (4) (2005) 447–456.
[2] V.K. Alilou, M. Teshnehlab, Prediction of 28-day compressive strength of concrete on the third day using artificial neural networks, International Journal of Engineering 3 (6) (2010) 565–576.
[3] B. Boukhatem, S. Kenai, A. Tagnit-Hamou, M. Ghrici, Application of new information technology on concrete: an overview, Journal of Civil Engineering and Management 17 (2) (2011) 248–258.
[4] H. Byun, S.-W. Lee, A survey on pattern recognition applications of support vector machines, Support Vector Machines for Computer Vision and Pattern Recognition, vol. 17, World Scientific Publishing Co. Pte Ltd, 2003, pp. 459–486.
[5] Ta-Peng Chang, F.-C. Chung, Hung-Chin Lin, A mix proportioning methodology for high-performance concrete, Journal of the Chinese Institute of Engineers 19 (6) (1996) 645–655.
[6] C.Z. Chang, Research on the mix proportion of high flowing eugenic concrete, Department of Civil Engineering, Chung Hua University, 1997.
[7] F.C. Chung, Study on characteristic of coarse aggregate in high-performance concrete, Department of Construction Engineering, National Taiwan University of Science and Technology, 1995.
[8] S. Dumais, H. Chen, Hierarchical classification of web content, SIGIR Forum (ACM Special Interest Group on Information Retrieval) (Compendex), 2000, pp. 256–263.
[9] M.H. Fazel Zarandi, I.B. Türksen, J. Sobhani, A.A. Ramezanianpour, Fuzzy polynomial neural networks for approximation of the compressive strength of concrete, Applied Soft Computing 8 (1) (2008) 488–498.
[10] G. Giaccio, C. Rocco, D. Violini, J. Zappitelli, R. Zerbino, High-strength concretes incorporating different coarse aggregates, ACI Materials Journal 89 (3) (1992) 242–246.
[11] O.E. Gjorv, P.J.M. Monteiro, P.K. Mehta, Effect of condensed silica fume on the steel-concrete bond, ACI Materials Journal 87 (6) (1990) 573–580.
[12] S. Haykin, Neural networks: a comprehensive foundation, 2nd edition Prentice Hall, 1999.
[13] P. Hellier, C. Barillort, E. Memin, P. Perez, Hierarchical estimation of a dense deformation field for 3-D robust registration, IEEE Transactions on Medical Imaging 20 (5) (2001) 388–402.
[14] J. Hoła, K. Schabowicz, Application of artificial neural networks to determine concrete compressive strength based on non-destructive tests, Journal of Civil Engineering and Management 11 (1) (2005) 23–32.
[15] G.E. Hwang, A study on blast furnace slag concrete, Department of Civil Engineering, National Chiao Tung University, 1991.
[16] T.S. Hwang, Compressive strength of blast furnace slag concrete, Department of Civil Engineering, National Chiao Tung University, 1966.
[17] W.H. Ian, F. Eibe, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, New Zealand, 1999.
[18] J. Kasperkiewicz, A. Dubrawski, HPC strength prediction using artificial neural network, Journal of Computing in Civil Engineering 9 (4) (1995) 279–284.
[19] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, International Joint Conference on Artificial Intelligence, vol. 2, Morgan Kaufmann, 1995, pp. 1137–1143.
[20] W.S. Langley, G.G. Carette, V.M. Malhotra, Structural concrete incorporating high volumes of ASTM Class F fly ash, ACI Materials Journal 86 (5) (1989) 507–514.
[21] C.F. Lee, A study on dry shrinkage and creep property of HPC, Department of Construction Engineering, National Taiwan University of Science and Technology, 1994.
[22] M. Lessard, O. Chaalla, P.-C. Aitcin, Testing high-strength concrete compressive strength, ACI Materials Journal 90 (4) (1993) 303–308.
[23] F.Y. Lin, Mixture proportion and quality of HPC, Department of Construction Engineering, National Taiwan University of Science and Technology, 1994.
[24] H.L. Mo, A study on high performance concrete, Department of Construction Engineering, National Taiwan University of Science and Technology, 1995.
[25] D. Mostofi, A. Samaee, HPC Strength prediction using artificial neural network, First International conference of C&D, 2001, pp. 44–49.
[26] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, Fourth ed. IRWIN, 1996.
[27] J. Oh, I. Lee, J. Kim, G. Lee, Applications of neural networks for proportioning of concrete mixes, ACI Material Jorurnal 96 (1) (1999) 51–59.
[28] J. Rousu, C. Saunders, S. Szedmak, Learning hierarchical multi-category text classification models, International Conference on Machine Learning (2005) 744–751.
[29] R.L. Silveira, J.M. Calixto, J.T. Fontoura, Mechanical Properties and Durability of High-Performance Concrete, in: V.M. Malhotra, P. Helene, L.R. Prudencio, D.C.C. Dal Molin (Eds.), Second CANMET/ACI International Conference Gramado, American Concrete Institute, Gramado, RS, Brazil, 1999, pp. 655–670.
[30] A. Smola, B. Schölkopf, A tutorial on support vector regression, NeuroCOLT Technical Report, 1998.
[31] J. Sobhani, M. Najimi, A.R. Pourkhorshidi, T. Parhizkar, Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models, Construction and Building Materials 24 (5) (2010) 709–718.
[32] L.W.G. Strijbosch, J.J.A. Moors, Calculating the accuracy of hierarchical estimation, IMA Journal of Management Mathematics 21 (2010) 303–315 (Compendex).
[33] A. Sun, E.-P. Lim, Hierarchical text classification and evaluation, IEEE International Conference on Data Mining (2001) 521–528.
[34] V.N. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.
[35] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Second edition Morgan Kaufmann, 2005.
[36] I.-C. Yeh, Modeling of strength of high performance concrete using artificial neural networks, Cement and Concrete Research 28 (12) (1998) 1797–1808.
[37] I.C. Yeh, L.-C. Lien, Knowledge discovery of concrete material using Genetic Operation Trees, Expert Systems with Applications 36 (3, Part 2) (2009) 5807–5812.
[38] H. You, T. Ryu, Development of a hierarchical estimation method for anthropometric variables, International Journal of Industrial Ergonomics 35 (2005) 331–343 (Compendex).



**Dr. Jui-Sheng (Rayson) Chou** received his B.S. and M.S. from National Taiwan University, and Ph.D. in Construction Engineering and Project Management from The University of Texas at Austin. Chou is now an associate professor in the Department of Construction Engineering at National Taiwan University of Science and Technology. He has over a decade of practical experience in engineering management and consulting services for the private and public sectors. He has over fifty refereed publications and serves on a number of professional committees. With registered professional engineer licenses in four engineering disciplines, He is a member of several international and domestic professional organizations. His teaching and research interests primarily involve Project Management (PM) related to knowledge discovery in databases (KDD), data mining, decision, risk & reliability, and cost management.

**Dr. Chih-Fong Tsai** obtained a Ph.D. at School of Computing and Technology from the University of Sunderland, UK in 2005 for the thesis entitled “Automatically Annotating Images with Keywords”. He is now an associate professor at the Department of Information Management, National Central University, Taiwan. He has more than 50 refereed publications. He received the Distinguished New Faculty Award from National Central University in 2010 and the Highly Commended Award (Emerald Literati Network 2008 Awards for Excellence) for a paper published in *Online Information Review* (“A Review of Image Retrieval Methods for Digital Cultural Heritage Resources”). His current research focuses on multimedia information retrieval and data mining applications.