Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# Predicting the shrinkage of thermal insulation mortar by probabilistic neural networks*

Yi-qun DENG, Pei-ming WANG†‡

(*Key Laboratory of Advanced Civil Engineering Materials, Ministry of Education, Tongji University, Shanghai 200092, China*)

†E-mail: tjwpm@126.com

Received July 21, 2009; Revision accepted Dec. 2, 2009; Crosschecked Jan. 6, 2010

**Abstract:** This study explored the potential of using probabilistic neural networks (PNN) to predict shrinkage of thermal insulation mortar. Probabilistic results were obtained from the PNN model with the aid of Parzen non-parametric estimator of the probability density functions (PDF). Five variables, water-cementitious materials ratio, content of cement, fly ash, aggregate and plasticizer, were employed for input variables, while a category of 56-d shrinkage of mortar was used for the output variable. A total of 192 groups of experimental data from 64 mixtures designed using JMP7.0 software were collected, of which 120 groups of data were used for training the model and the other 72 groups of data for testing. The simulation results showed that the PNN model with an optimal smoothing parameter determined by the curves of the mean square error (MSE) and the number of unrecognized probability densities (UPDs) exhibited a promising capability of predicting shrinkage of mortar.

**Key words:** Mortar, Shrinkage, Probabilistic neural networks (PNN), Thermal insulation
**doi:**10.1631/jzus.A0900441 **Document code:** A **CLC number:** TU3

## 1 Introduction

Given that thermal insulation mortar is being increasingly used in building energy conservation, as for other cement-based materials, shrinkage is an important concern. Cracks caused by shrinkage reduce the thermal insulation effect severely and may affect the durability and stability of structures. Over the past 30 years, researchers have carried out extensive work concerning the shrinkage of cement-based materials, and various analytical models have been developed to predict the shrinkage in concrete. The American Concrete Institute (ACI) recommended the procedure ACI-209 for predicting the creep and shrinkage of concrete. This model is applicable to both normal and all light-weight concretes using both moist and steam curing (ACI Committee 209, 1982). The CEB-FIP model was proposed for predicting the shrinkage of ordinary structural concretes (CEB-FIP, 1993). RILEM TC-107-GCS recommended the B3 model, which was based on the statistical analysis of creep and shrinkage data in a computerized data bank involving about 15 000 data points and about 100 test series (RILEM TC-107-GCS, 1995). Other models were also proposed, such as the Muller model, the GL2000 model (Lam, 2002; Goel *et al*., 2007). Nevertheless, all of these linear or nonlinear shrinkage models deduced by conventional analytical methods were oversimplified such that they could not illustrate the complex physical and chemical processes during the shrinkage.

In recent years, an extensive application of artificial neural networks (ANN) provided an alternative solution for this problem. This technique, inspired by the mechanism of human brain information processing, is able to consider various inputs without complicated equations, and allows modeling of complex systems without an explicit knowledge or formulation

‡ Corresponding author

* Project (No. 2006BAJ05B03) supported by the National Key Technologies Supporting Program of China during the 11th Five-Year Plan Period

of the relationship existing among the variables (Thibault and Grandjean, 1991). It has been proved that an ANN model comprised of three simplified layers has the ability to approximate to a random curve in arbitrary-precision (Yan and Zhang, 2002). Many researchers have exploited the ability of ANN to predict the properties of concrete (Wang *et al*., 1999; Topçu and Sarıdemir, 2007; Cevik and Guzelbey, 2008; Fazel Zarandi *et al*., 2008; Sarıdemir *et al*., 2009). However, the vast majority of ANN models were based on the back-propagation (BP) learning algorithm and their deficiencies were significant, such as sensitivity to variations in training data, difficulties in deciding the number of hidden nodes and the learning rate, susceptibility to false minima and so forth (Calderon and Cheh, 2002). More importantly, the results predicted by this kind of ANN model were not probabilistic but deterministic, whereas experimental results under the same conditions exhibit a characteristic of normal distribution.

Recently, a probabilistic neural network (PNN) implementation of a Bayesian classifier, while managing to retain simplicity and transparency, seems to have been able to provide a feasible solution to these problems. Probabilistic results can be obtained from the PNN model with the aid of the Parzen non-parametric estimator of the probability density functions (PDFs). So far, the PNN model approach has been applied to many fields, such as bankruptcy prediction (Yang *et al*., 1999), volatility forecasts (Hamid and Iqbal, 2004), stock market prediction (Kim and Chun, 1998), image and pattern recognition (Al-Omari and Al-Jarrah, 2004; Sheng and Chen, 2009), medical/biochemical field applications (Wang *et al*., 1999; Holmes *et al*., 2001), etc. Compared with other ANN models, the advantages are as follows:

1. The PNN model is a type of feed-forward neural networks with a simplified architecture, and as a result it requires less training time than an ANN model based on the BP learning algorithm (Simon and Karim, 2001).

2. There is a unique free parameter (the smoothing parameter $\sigma$) in the PNN model which can be adjusted dynamically without network retraining. Furthermore, a satisfactory prediction precision can be obtained by making decision surfaces with an appropriate smoothing parameter.

3. Training of the PNN model can be incremental as data becomes available and old patterns can be "forgotten" and replaced by new patterns if desired (Steenhoek, 1999).

4. In addition to the gradient descent method used to train ANN models, the Gaussian kernel function is employed to calculate the probability density of input vectors. False minima, therefore, are avoided during the learning process of the PNN model.

5. Most of all, the PNN model provides probabilistic viewpoints as well as deterministic classification results, which are not dependent on the randomization order of training data.

In this paper, a PNN model was used to predict the shrinkage of thermal insulation mortar on 56-d age. Five variables (water-cementitious (w-c) materials ratio, content of cement, fly ash, aggregate and plasticizer) were entered as inputs, while the category of 56-d shrinkage was used as the output variable. To obtain data with good dispersivity and ergodicity, 64 mixtures were designed using JMP7.0 software (serial number 4582731217), and a total of 192 groups of experimental data were collected, of which 120 groups of data were used for model training and the other 72 groups of data were used for testing. An optimal smoothing parameter ($\sigma_{opt}$) was determined by the mean square error (MSE) and the number of unrecognized probability densities (UPDs). Finally, two different mixtures apart from the 64 mixtures were employed for investigating the generalization capability of the PNN model.

## 2 Probabilistic neural networks

### 2.1 Theory of probabilistic neural networks

The PNN is a type of radial basis network originally developed for radar classification. It was first proposed by Specht (1988). The PNN combines feed-forward neural networks and the well-known Bayesian strategy for decision making with the Parzen non-parametric estimator of the PDFs of different categories.

Consider a two-category situation in which the state of nature $\theta$ is known to be either $\theta_A$ or $\theta_B$. If it is desired to decide whether $\theta=\theta_A$ or $\theta=\theta_B$ based on a set of measurements represented by the $p$-dimensional vector $\boldsymbol{X}^p=[x_1, x_2, \ldots, x_p]$, the Bayesian decision rule becomes

$$\begin{cases} d(\boldsymbol{X})=\theta_{\mathrm{A}}, \text{ if } \ h_{\mathrm{A}}l_{\mathrm{A}}f_{\mathrm{A}}(\boldsymbol{X})>h_{\mathrm{B}}l_{\mathrm{B}}f_{\mathrm{B}}(\boldsymbol{X}), \\ d(\boldsymbol{X})=\theta_{\mathrm{B}}, \text{ if } \ h_{\mathrm{B}}l_{\mathrm{B}}f_{\mathrm{B}}(\boldsymbol{X})>h_{\mathrm{A}}l_{\mathrm{A}}f_{\mathrm{A}}(\boldsymbol{X}), \end{cases} \quad (1)$$

where $f_{\mathrm{A}}(\boldsymbol{X})$ and $f_{\mathrm{B}}(\boldsymbol{X})$ are the PDFs for categories $\theta_{\mathrm{A}}$ and $\theta_{\mathrm{B}}$, respectively. $l_{\mathrm{A}}$ is the loss function associated with the decision $d(\boldsymbol{X})=\theta_{\mathrm{B}}$ when $\theta=\theta_{\mathrm{A}}$; $l_{\mathrm{B}}$ is the loss associated with the decision $d(\boldsymbol{X})=\theta_{\mathrm{A}}$ when $\theta=\theta_{\mathrm{B}}$ (the losses associated with correct decisions are taken to be equal to zero). $h_{\mathrm{A}}$ is the a priori probability of occurrence of patterns from category $h_{\mathrm{A}}$, and $h_{\mathrm{B}}=1-h_{\mathrm{A}}$ is the a priori probability that $\theta=\theta_{\mathrm{B}}$. The boundary between the region in which the Bayesian decision $d(\boldsymbol{X})=\theta_{\mathrm{A}}$ and the region in which $d(\boldsymbol{X})=\theta_{\mathrm{B}}$ is given as

$$f_{\mathrm{A}}(\boldsymbol{X})=h_{\mathrm{B}}l_{\mathrm{B}}f_{\mathrm{B}}(\boldsymbol{X})/(h_{\mathrm{A}}l_{\mathrm{A}}). \quad (2)$$

In a simplified case where both losses function and the a priori probability are equal to each other, the category to which each belongs is decided by PDF. The Bayesian decision classifies an input pattern to the category that has its PDF greater than the others. Therefore, the accuracy of the decision boundaries depends on the accuracy with which the underlying PDFs are estimated. Parzen showed that a category of PDF estimators asymptotically approaches the underlying parent density provided that it is smooth and continuous. The PNN replaces the sigmoid activation function that is normally used in ANNs with the Gaussian kernel function, and the multivariate estimates can be expressed as

$$f_{\mathrm{A}}(\boldsymbol{X})=\frac{1}{m(2\pi)^{P/2}\sigma^{P}}\sum_{i=1}^{m}\exp\left(-\frac{\left\|\boldsymbol{X}-\boldsymbol{X}_{\mathrm{A}}\right\|^{2}}{2\sigma^{2}}\right), \quad (3)$$

where $m$ is the pattern number of category A; $\boldsymbol{X}_{\mathrm{A}}$ is the training pattern from category A; and $\sigma$ is the smoothing parameter.

Compared with other types of ANN, the PNN has only a single adjustable smoothing parameter $\sigma$ (or kernel width), along with the members of the training set which defines the PDF for each data category. Each PDF is composed of Gaussian kernels of width $\sigma$ located at each pattern vector and essentially determines the boundaries for categories. The kernel width $\sigma$ is critical because it determines the amount of interpolation that occurs between adjacent pattern vectors.

Fig. 1 illustrates the effect of $\sigma$ on $f(\boldsymbol{X})$ in the case of 1D independent variable $\boldsymbol{X}$. The PDF is obtained from Eq. (3). A value of $\sigma$=0.1 causes the estimated parent density to have 3 distinct modes corresponding to the 3 training samples. $\sigma$=0.5 brings about a greater degree of interpolation between adjacent categories, but the modes remain distinct. When $\sigma$ increased to 1.0, $f(\boldsymbol{X})$ has a single mode and a shape approximating Gaussian distribution. The value causes some flattening of the density function, with spreading out of the tails (Shaffer and Rose-Pehrsson, 1999). In sum, PNN with a larger $\sigma$ has a geometrically less complex decision surface which leads to an increased robustness of the classifier, since the location of the decision surface is less affected by noise and outliers in the data. However, too larger $\sigma$, such as $\sigma\sim\infty$, will cause the decision function to converge to the linear decision surface of the nearest mean (or minimum distance) classifier, which causes a greater possibility of misclassification. Conversely, a small $\sigma$ will reduce the amount of overlap between adjacent data classes, and the effect of classification is correspondingly extreme. But a very small $\sigma$ has obvious disadvantages at the same time. If $\sigma\sim0$ the decision function is essentially equal to the decision function of the 1-nearest neighbor classifier. This is generally a rather complex decision surface which is sensitive to the noise in the data. A tiny change of input vectors may result in unrecognized patterns and UPD of adjacent categories (Kraaijveld, 1996).

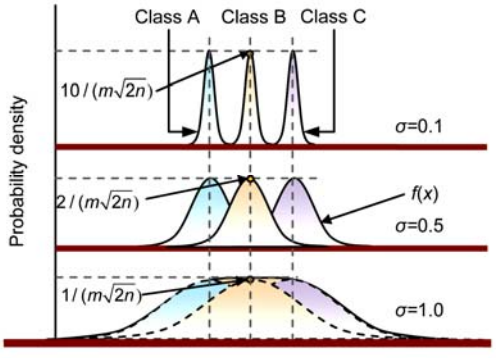


**Fig. 1 Smoothing effect of $\sigma$ on PDF**

## 2.2 Architecture of probabilistic neural networks

PNN is a type of feed-forward neural network and responds to an input pattern by processing the input data from one layer to the next with no feedback paths. The architecture of PNN (Fig. 2) is composed

of four layers: input layer, pattern layer, summation layer and output layer. The input layer does not perform any computation but merely distributes the inputs to the neurons in the pattern layer (step 1). The number of neurons in this layer is decided by the dimensions of the input vectors.

The second layer of PNN is a pattern layer which is the core of the algorithm. The neuron number of this layer is determined by the training set. With input vectors from the input layer, each neuron in the pattern layer computes the distances between the presented input vectors and the training set represented by pattern neurons (step 2). Subsequently, each of the pattern neurons yields dot products with the distances and a threshold value $b_i$ (step 3). Finally, nonlinear operations on $\text{dist}_i \cdot b_i$ are performed before outputting its activation level to the summation unit (step 4). Instead of the sigmoid activation function commonly used for back-propagation, the nonlinear operation used here was length.

Summation layer is the third layer. Pattern category decides the neuron number in this layer. The activations from each neuron of pattern layer are fed into corresponding summation layer neurons and summed within the same categories (step 5). Before the products of the summation layer are forwarded to the output layer, the estimated probability of each category should have dot variable weight $C_i=h_i l_i/n_i$ (step 6). Note that, $n_i$ is the number of training sets from category $i$. If the misclassification cost and prior probabilities of each category are equal, and the categories are mutually exclusive (i.e., no case can be classified into more than one category) and exhaustive (i.e., the training set covers all categories fairly), the activation of the summation neurons will be equal to the posterior probability of each category.

The last layer is the output layer, which contains one neuron, by which trivial threshold discrimination is operated (step 7). Then, the results from the all summation neurons compete with each other and the largest is fed into the output neuron to yield the computed categories and PDFs to which the example belongs.

## 3 PNN for shrinkage prediction of thermal insulation mortar

### 3.1 Properties of materials

1. Cementitious materials

Cement: the cement used in experiment was ordinary Portland cement supplied by Anhui Conch Cement Company Ltd., China. The type of cement was P.O.42.5R (42.5 N/mm) with a specific gravity of 3.15 g/cm$^3$, and the initial and final setting time of the cements was 142 min and 194 min, respectively. Fly ash: the fly ash was obtained from Baoshan Steel Company Ltd. of Shanghai, China. This kind of fly ash contains 12.3% calcium which belongs to high-calcium fly ash according to the National Standard of China. The specific gravity was 2.20 g/cm$^3$, and the specific surface area was 600 m$^2$/kg.

2. Aggregate

The aggregate used here was vitrified micro bubble (VMB) provided by Fudan Argosy Functional Material Company Ltd. of Shanghai, China. VMB is

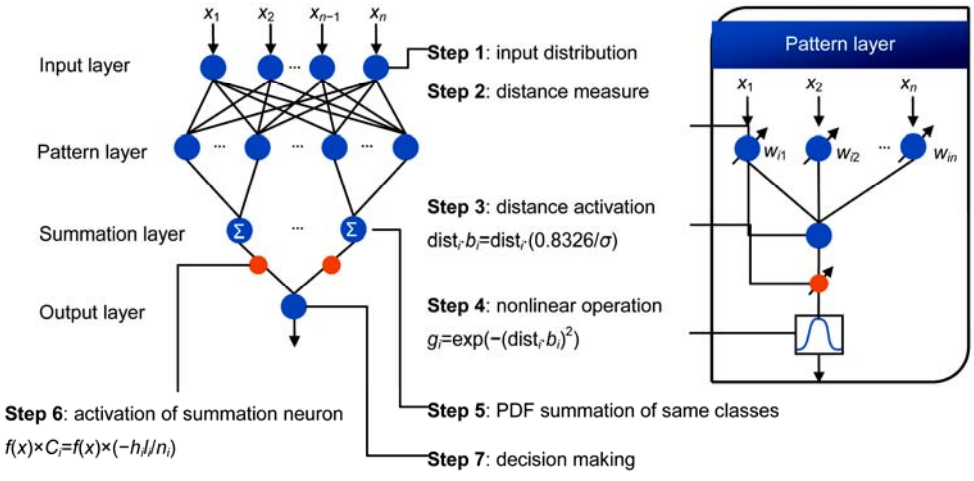


**Fig. 2 Architecture and workflow of PNN**

an internal hollow inorganic microsphere manufactured through pitch stone crashed, sieved and high temperature calcinations. It is often used for aggregate in light thermal insulation materials. The size of VMB was 0.5–1.5 mm with 20%–50% water absorption value. Its relative density at saturated surface dry condition was 120–150 kg/m$^3$ and the thermal coefficient was 0.032–0.045 W/(m·K). The volume loss at 1 MPa was 38%–46%.

3. Chemical admixture

Plasticizer: A carboxylic-type hyper plasticizing admixture (type PCA) was used to maintain the workability of fresh mortar which was obtained from Jiangsu Bote New Materials Company Ltd., China. PCA was a liquid which has the performances of high range water reduction, excellent slump retention and strengthening. The solid content of PCA was 21.7% and the water reduction rate can be up to 35% at high dosage rates. The best dosage was 1.2%–1.8%.

Dispersible polymer powders (DPP): DPP type VINNAPAS RI 551 Z obtained from Wacker Chemie AG (Germany) was used to improve moisture retention, fluidity, workability of mortar.

Hydroxypropyl methyl cellulose (HMC): HMC type 60001 was obtained from Clariant Company (Switzerland).

### 3.2 Test of shrinkage

Shrinkage test of specimens was conducted according to the Standard for Test Method of Basic Properties of Construction Mortar given by Chinese Standard JGJ 70-90 (MOC, 1991). Firstly, the raw materials of mortar were put in a forced mixer simultaneously and mixed for 3 min. Then mixture was cast into test specimens in mold in dimensions of 40 cm×40 cm×160 cm. Specimens were demolded at 7 d and cured in condition of temperature=(20±3) °C and relative humidity ≥90%. When specimens were demolded, the initially lengths of specimens $L_0$ were recorded at once, and the finally lengths $L_t$ were recorded at the age of 56 days. Finally, the shrinkages of mortar were calculated by Eq. (4), and the values were accurate to $1\times10^{-5}$. For all mixtures, the average of experimental results of three identical specimens was used.

$$\zeta = (L_t - L_0)/140. \tag{4}$$

### 3.3 Procedure for predicting the shrinkage of mortar

The general implementation procedure of PNN model as shown in Fig. 3 includes four steps.

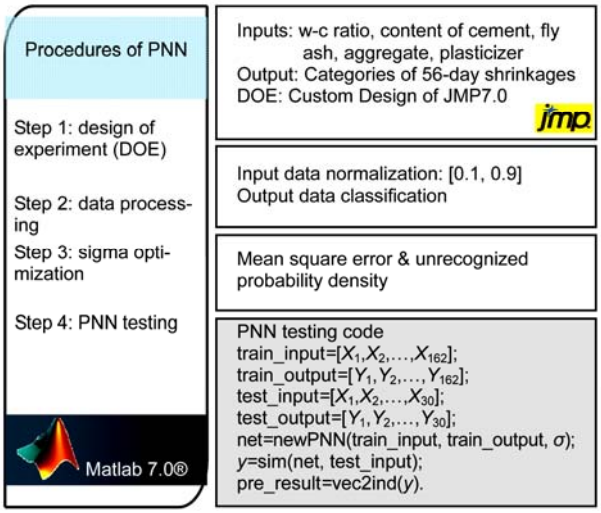


**Fig. 3 Procedure for predicting the shrinkage of mortar**

Step 1: Design of experiments. To obtain data with good dispersivity and ergodicity, JMP7.0 software was used to design experiments. Five variables, including the w-c materials ratio, content of cement, fly ash, aggregate and plasticizer, were employed for factors, while the categories of 56-d shrinkages were defined as the response.

Step 2: Data processing. Due to the difference of units within variables, one may need to normalize input data before proceeding with further statistical analysis. All of the input data should be normalized to the value between 0.1 and 0.9 using Eq. (5), and the outputs were divided into appointed categories with an increment of (max−min)/categories.

$$\overline{x}_{ij} = 0.1 + 0.8 \times \left( \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \right). \tag{5}$$

Step 3: Smoothing parameter optimization. Optimal smoothing parameter $\sigma_{opt}$ was determined by MSE and UPD.

Step 4: PNN model test. For investigating the generalization capability of PNN model with $\sigma_{opt}$, 45 groups of data from two mixtures apart from 64 mixtures were utilized. Then the estimated PDFs of

two mixtures from the optimal PNN model were compared with the experimental values. All of the codes were developed in MATLAB7.0.

### 3.4 Experimental design

JMP7.0 software was used for experimental design. The unique function of Custom Designer for experiment in JMP can meet complex experimental design requirements. The computational method of Custom Designer is an iterative algorithm called coordinate exchange. Each iteration of the algorithm involves testing every value of every factor in the design to determine if replacing that value increases the optimality criterion. If so, the new value replaces the old. Iteration continues until no replacement occurs in an entire iterate. To avoid converging to a local optimum, the whole process is repeated several times using a different random start. The Custom Designer displays the best of these designs (JMP, 2005).

Five continuous factors covered the w-c materials ratio, content of cement, fly ash, aggregate and plasticizer, while 56-d shrinkage of specimens was response. The total weight of DPP and HMC was 24 g per 1000 g cementitious materials. A factor constraint was that the total weight of cementitious materials was equal to 1000 g. For investigating the interactions effect of different factors, 2, 3, 4 and 5 interactions and 2, 4 powers were selected (Table 1). 64 different mixtures were designed (Table 2) and a total of 192 groups of experimental data were collected, of which 120 groups of data were used for model training and the other 72 groups of data were used for testing.

To assess the feasibility of a normal model, a quantile-quantile plot (Q-Q plot) based on the Lilliefors test was applied (Lilliefors, 1967). Fig. 4 is the normal quantile plot which is useful for visualizing the extent to which the variable is normally distributed. On the left of Fig. 4, the histogram is overlaid with the normal distribution curve and the Q-Q plot is to the right. The dark data points represent the mean sample differences. When a variable is normal, the normal quantile plot approximates a diagonal straight line. The dotted curved lines represent the Lilliefors confidence bounds which is a special case of the Kolmogorov-Smirnov goodness-of-fit test. If the data points fall within these bounds, it can be concluded that the data is normally distributed. In this work, there is no abnormal point within a 95% confidence bound. As a result, these data points obtained from the design of experiment (DOE) were reliable when used to the PNN model.

**Table 1 Parameter setting of Custom Design in JMP7.0**

| Response | 56-d shrinkage |
|---|---|
| Factor | |
| w-c ratio | 1.20−1.80 (continuous) |
| Cement | 700−1000 g (continuous) |
| Fly ash | 0−300 g (continuous) |
| Aggregate | 500−800 g (continuous) |
| Plasticizer | 0−18 g (continuous) |
| DPP+HMC | 20+4 g (constant) |
| Factor constraints | Cement+Fly ash=1000 |
| Model | |
| Main effects | 5 |
| Interactions | 2nd, 3rd, 4th, 5th |
| Powers | 2nd, 4th |
| Number of runs | 64 |

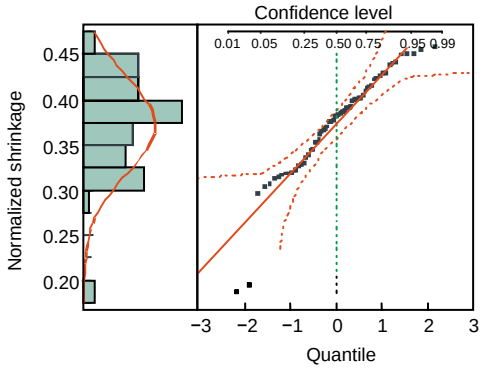


**Fig. 4 Normal quantile plot of 56-d shrinkage**
The dark data points represent the mean sample differences, and the dotted curved lines represent the Lilliefors confidence bounds

### 3.5 PNN model for shrinkage prediction of mortar

#### 3.5.1 Influence of shrinkage categories on the PNN model

The resolution of the PNN model was defined as the minimal prediction range of the PNN model and was calculated by (max−min)/categories. As a result the shrinkage category had great influence on predicting precision of the PNN model. As shown in Fig. 5, PNN models with four types of shrinkage categories (8, 16, 30, 50) and $\sigma$=0.01 were investigated. It was concluded that a high resolution could be obtained when the PNN model had a great shrinkage category but a lower predicting precision. When the

**Table 2 Mortar mixtures designed by JMP7.0**

| No. | w-c ratio | Cement (g) | Fly ash (g) | Aggregate (g) | Plasticizer (g) | No. | w-c ratio | Cement (g) | Fly ash (g) | Aggregate (g) | Plasticizer (g) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.80 | 867 | 133 | 500 | 0.0 | 34 | 1.20 | 808 | 192 | 530 | 16.2 |
| 2 | 1.20 | 867 | 133 | 770 | 0.0 | 35 | 1.20 | 748 | 252 | 500 | 18.0 |
| 3 | 1.80 | 890 | 110 | 500 | 18.0 | 36 | 1.74 | 791 | 209 | 800 | 0.0 |
| 4 | 1.20 | 799 | 201 | 800 | 0.0 | 37 | 1.74 | 884 | 116 | 800 | 16.2 |
| 5 | 1.20 | 797 | 203 | 500 | 0.0 | 38 | 1.20 | 756 | 244 | 800 | 0.0 |
| 6 | 1.20 | 909 | 91 | 800 | 0.0 | 39 | 1.20 | 905 | 95 | 800 | 18.0 |
| 7 | 1.20 | 823 | 177 | 800 | 18.0 | 40 | 1.80 | 811 | 189 | 800 | 0.0 |
| 8 | 1.26 | 851 | 149 | 530 | 1.8 | 41 | 1.80 | 791 | 209 | 530 | 18.0 |
| 9 | 1.80 | 923 | 77 | 800 | 18.0 | 42 | 1.20 | 814 | 186 | 800 | 18.0 |
| 10 | 1.80 | 779 | 221 | 500 | 0.0 | 43 | 1.20 | 911 | 89 | 500 | 16.2 |
| 11 | 1.26 | 865 | 135 | 770 | 18.0 | 44 | 1.80 | 918 | 82 | 800 | 18.0 |
| 12 | 1.20 | 962 | 38 | 800 | 18.0 | 45 | 1.20 | 792 | 208 | 500 | 18.0 |
| 13 | 1.50 | 803 | 197 | 770 | 18.0 | 46 | 1.20 | 931 | 69 | 500 | 0.0 |
| 14 | 1.80 | 779 | 221 | 800 | 18.0 | 47 | 1.50 | 791 | 209 | 800 | 1.8 |
| 15 | 1.50 | 864 | 136 | 500 | 9.0 | 48 | 1.74 | 831 | 169 | 650 | 18.0 |
| 16 | 1.80 | 917 | 82 | 500 | 18.0 | 49 | 1.80 | 768 | 232 | 500 | 0.0 |
| 17 | 1.50 | 879 | 121 | 650 | 0.0 | 50 | 1.20 | 916 | 83 | 500 | 18.0 |
| 18 | 1.20 | 853 | 147 | 500 | 0.0 | 51 | 1.80 | 796 | 204 | 800 | 18.0 |
| 19 | 1.20 | 900 | 100 | 650 | 9.0 | 52 | 1.74 | 889 | 111 | 800 | 0.0 |
| 20 | 1.80 | 733 | 267 | 800 | 0.0 | 53 | 1.80 | 899 | 101 | 650 | 1.8 |
| 21 | 1.80 | 864 | 136 | 800 | 18.0 | 54 | 1.80 | 864 | 136 | 770 | 9.0 |
| 22 | 1.74 | 887 | 113 | 500 | 9.0 | 55 | 1.26 | 881 | 119 | 530 | 18.0 |
| 23 | 1.20 | 829 | 171 | 800 | 0.0 | 56 | 1.80 | 871 | 129 | 800 | 18.0 |
| 24 | 1.20 | 798 | 202 | 500 | 0.0 | 57 | 1.26 | 782 | 218 | 800 | 16.2 |
| 25 | 1.20 | 891 | 109 | 770 | 1.8 | 58 | 1.80 | 813 | 187 | 500 | 18.0 |
| 26 | 1.80 | 923 | 77 | 500 | 0.0 | 59 | 1.80 | 856 | 144 | 500 | 16.2 |
| 27 | 1.20 | 882 | 118 | 500 | 0.0 | 60 | 1.80 | 916 | 84 | 800 | 0.0 |
| 28 | 1.80 | 913 | 87 | 500 | 0.0 | 61 | 1.26 | 884 | 116 | 800 | 0.0 |
| 29 | 1.80 | 741 | 259 | 500 | 18.0 | 62 | 1.20 | 858 | 142 | 500 | 18.0 |
| 30 | 1.80 | 850 | 150 | 800 | 0.0 | 63 | 1.50 | 871 | 129 | 650 | 16.2 |
| 31 | 1.20 | 744 | 256 | 800 | 18.0 | 64 | 1.80 | 898 | 102 | 530 | 18.0 |
| 32 | 1.20 | 883 | 117 | 800 | 16.2 | 1# | 1.30 | 700 | 300 | 800 | 10 |
| 33 | 1.80 | 850 | 150 | 500 | 0.0 | 2# | 1.50 | 900 | 100 | 500 | 5 |

Shade part is the two mixtures used for investigating generalization capability of the PNN model

shrinkage category was 50, the MSE of the PNN model was 0.556 and the resolution was 0.004%, which means this model can predict shrinkage values fluctuating within a range of 0.004%. The MSE was improved from 0.556 to 0.083 when the shrinkage category was 16, but the resolution dropped to 0.011%. The MSE of the PNN model with 8 categories was the same as that of the model with 16 shrinkage categories. In this study, the best option of shrinkage category was 16 if considering the accuracy of shrinkage measured values ($1\times10^{-4}$). In the following sections, the architecture of the PNN model was defined as 5-120-16-1, which means 5 input variables, 120 groups of training data prestored in neurons of the pattern layer, 16 output categories distributed in the summation layer and 1 output.

3.5.2 Smoothing parameter optimization for the PNN model

As above mentioned, the density estimation of the PNN model at a specific location in the pattern space was determined primarily by the nearest class. All other classes were assumed to be so distant that their contributions to the density estimation were minimal. The unique adjustable smoothing parameter $\sigma$ is critical because it determines the amount of interpolation that occurs between adjacent pattern vectors, which eventually affects the predicting precision of the PNN model. A large $\sigma$ has the advantage of producing a smooth PDF, which exhibits good interpolation properties for predicting new pattern vectors but which will simultaneously result in a great possibility of misclassification. However, a too

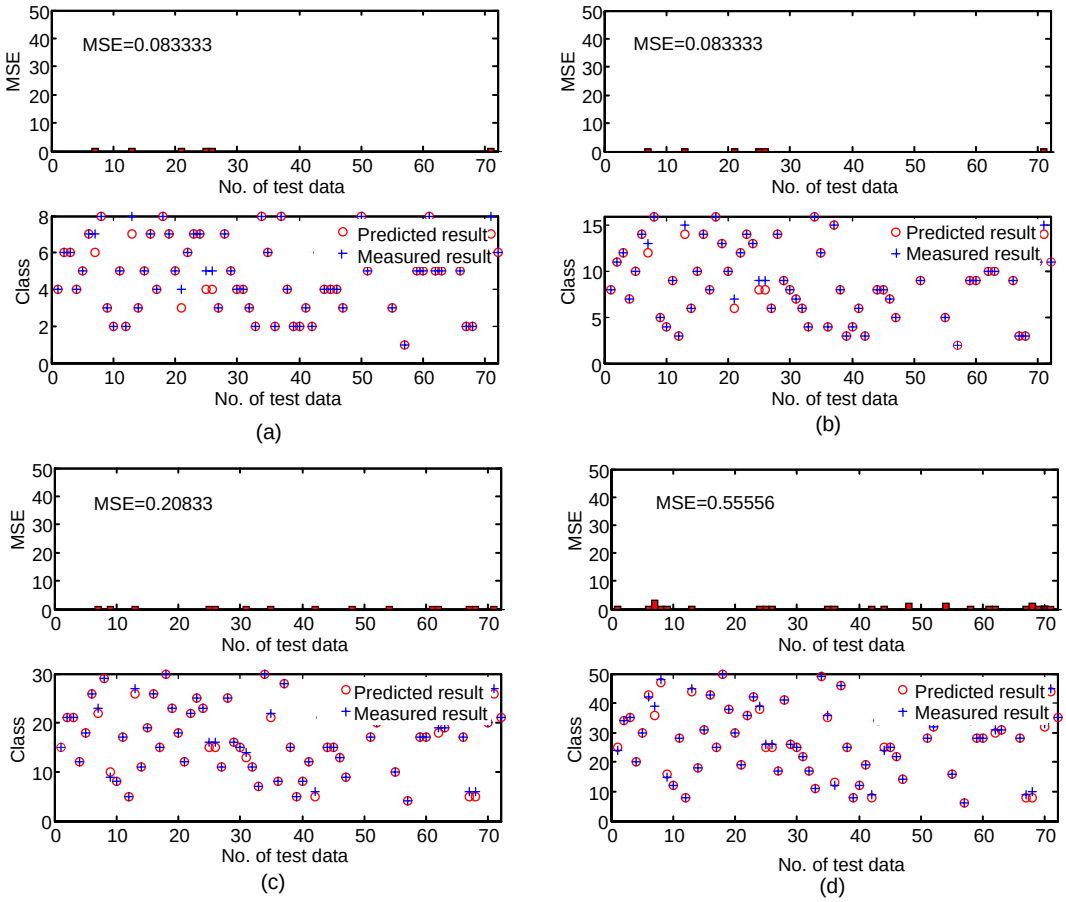


**Fig. 5 Predicted results with different classifications. (a) 8 classes; (b) 16 classes; (c) 30 classes; (d) 50 classes**

small $\sigma$, such as $\sigma$~0, brings about a rather complex decision surface, which is sensitive to the noise in the data. As a result, a tiny difference with the nearest class may lead to UPD.

Research into the theory of PDFs has led to various methods for approximating the optimal $\sigma$. For a typical PNN model, the optimal $\sigma$ was selected by minimizing the training error using a combination of cross-validation and univariate optimization (CV-OPT) (Masters, 1995). However, the time needed to optimize $\sigma$ increases exponentially with the number of training set patterns. Lee *et al*. (2009) utilized the dynamic decay adjustment (DDA) algorithm to automatically calculate the smoothing parameter. Kraaijveld (1996) reported a general approach for Parzen classifiers using a simple distance based calculation. The derivation of this equation was based on maximum likelihood principles and assumes independence of the samples in the training set. Nevertheless, none of these methods took into account the situation of unrecognized category once training set is not sufficient or ergodic in the sample space. In this study, one method with the combination of MSE and UPD was proposed to automatically calculate the optimal $\sigma$. The MSE and UPD of the PNN model were calculated with different $\sigma$ varying from 0.01 to 1.0 during the training processes. Then, all of the MSE and UPD were normalized between 0.1 and 0.9. As shown in Fig. 6, the MSE increased whereas the UPD decreased with an increase in $\sigma$. The optima $\sigma_{opt}$ was determined by the intersection point of MSE and UPD curves, which occurred at 0.0399.

Fig. 7 illustrates the relationships between the predicted values from the PNN model and the measured data. It was concluded that the PNN model with a big $\sigma$ led to a poor predicting precision. When $\sigma$=1.0, the data points showed scatter with the $R^2$=0.6561.

When $\sigma \leq 0.1$ the targets and outputs revealed a very promising behavior of linear correlation characterized by $R^2>0.98$. The results showed that the PNN model with an optimal $\sigma_{opt}$ exhibited better predicting performance with the $R^2=0.9938$ and a sum squared error (SSE) of 6.628, while model with Kraaijveld $\sigma$ 0.9853 and 15.76, respectively.

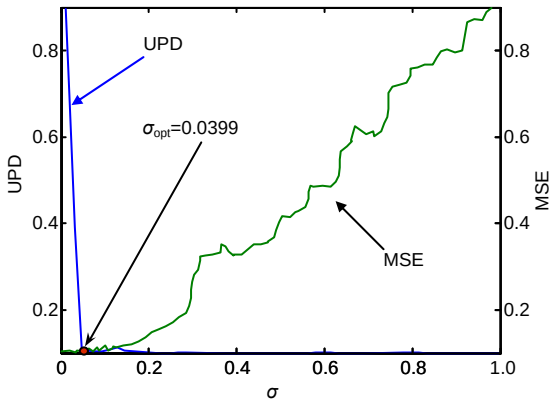


**Fig. 6 $\sigma$ optimized by normalized MSE and UPD**

To test the generalization performance of the proposed PNN model, two mixtures, 1# and 2# showed in Table 2, apart from the former 64 mixtures were employed, and a total of 45 groups of data for each mixture were obtained. Fig. 8 shows a comparison of probability density between the estimated results obtained using the PNN model and the experiment results. It was shown that the density probability plot fluctuated wildly in the PNN model with a large $\sigma$ ($\sigma$=1.0 and $\sigma$=0.1), and the distinction between different categories was not obvious. As a result misclassification might occur. Conversely, the PNN model with a small $\sigma$ (as above-mentioned) exhibits a good classification capability, but runs the risk of unrecognized probability density. A PNN model with $\sigma$=0.0973 calculated using Kraaijveld's equation showed excellent classification capability, but some categories were not recognized simultaneously. It was concluded that the PNN model with $\sigma_{opt}$ provided the best agreement with the distribution of the experiment results since the estimated density function showed the same peak value as the experiment results. In sum, PNN models with an optimal $\sigma$ determined by MSE and UPD had the ability to predict the shrinkage of thermal insulation mortar with a satisfactory predicting precision.

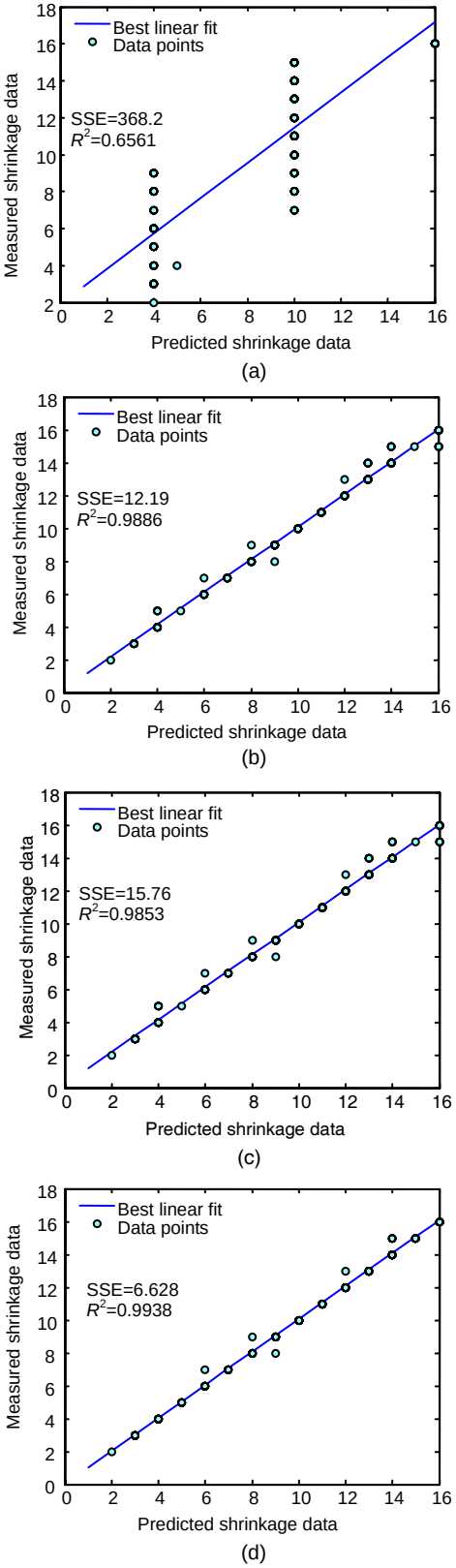


**Fig. 7 Linear fit results with different $\sigma$. (a) $\sigma$=1.0; (b) $\sigma$=0.1; (c) Kraaijveld $\sigma$; (d) $\sigma_{opt}$**

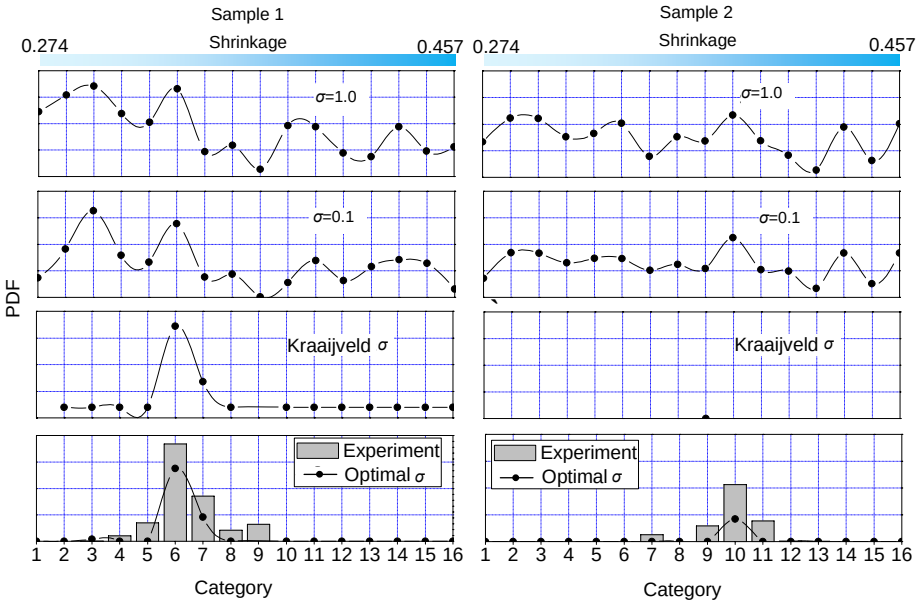


**Fig. 8 Probability density curves of shrinkage with 16 categories under different $\sigma$**

## 4 Conclusion

For the purpose of predicting the shrinkage of thermal insulation mortar without attempting experiments, a model was carried out in PNN. The PNN benefits from the power of ANN as highly flexible nonlinear mapping paradigms and the Bayesian theorem with the aid of the Parzen PDF nonparametric estimator. As a result, a probabilistic point of view can be achieved while a deterministic value is provided by other ANN. Training data with properties of compactness, ergodicity and compatibility were obtained from experimental design by JMP7.0 software, and the simulation results indicated that the PNN model whose architecture was 5-120-16-1 with an optimal $\sigma_{opt}$=0.0399 decided by MSE and UPD had the ability to predict the shrinkage of thermal insulation mortar with a satisfactory precision.

Future research should be directed at three objectives. Firstly, a better experimental design should be accomplished since the performance of training data was very important to PNN. Secondly, additional variables affecting the shrinkage of mortar, such as curing conditions, size and shape of specimen, etc., should be included. Lastly, and the most important, to predict the long term mortar shrinkage with short term shrinkage data, a combination with additional techniques, such as the sliding window algorithm, is another potential extension of the present research.

## References

ACI Committee 209, 1982. Prediction of Creep, Shrinkage and Temperature Effects in Concrete Structures. ACI-209-82, American Concrete Institute, Detroiti.

Al-Omari, F.A., Al-Jarrah, O., 2004. Handwritten Indian numerals recognition system using probabilistic neural networks. *Advanced Engineering Informatics*, **18**(1):9-16. [doi:10.1016/j.aei.2004.02.001]

Calderon, T.G., Cheh, J.J., 2002. A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems*, **3**(4):203-236. [doi:10.1016/S1467-0895(02)00068-4]

CEB-FIP, 1993. CEB-FIP Model Code 1990: Design Code 1994. Thomas Telford, London.

Cevik, A., Guzelbey, I.H., 2008. Neural network modeling of strength enhancement for CFRP confined concrete cylinders. *Building and Environment*, **43**(5):751-763. [doi:10.1016/j.buildenv.2007.01.036]

Fazel Zarandi, M.H., Türksen, I.B., Sobhani, J., Ramezanianpour, A.A., 2008. Fuzzy polynomial neural networks for approximation of the compressive strength of concrete. *Applied Soft Computing*, **8**(1):488-498. [doi:10.1016/j.asoc.2007.02.010]

Goel, R., Kumar, R., Paul, D.K., 2007. Comparative study of various creep and shrinkage prediction models for concrete. *Journal of Materials in Civil Engineering*, **19**(3):249-260. [doi:10.1061/(ASCE)0899-1561(2007)19:3(249)]

Hamid, S.A., Iqbal, Z., 2004. Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, **57**(10):1116-1125. [doi:10.1016/S0148-2963(03)00043-2]

Holmes, E, Nicholson, J.K., Tranter, G., 2001. Metabonomic characterization of genetic variations in toxicological and

metabolic responses using probabilistic neural networks. *Chemical Research in Toxicology*, **14**(2):182-191. [doi:10.1021/tx000158x]

Hsieh, S.H., Chen, C.H., 2009. Adaptive image interpolation using probabilistic neural network. *Expert Systems with Applications*, **36**(3):6025-6029. [doi:10.1016/j.eswa.2008.06.124]

JMP, 2005. Design of Experiments. SAS Institute Inc., Cary, NC, USA, p.80-81.

Kim, S.H., Chun, S.H., 1998. Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index. *International Journal of Forecasting*, **14**(3):323-337. [doi:10.1016/S0169-2070(98)00003-X]

Kim, D.K., Lee, J.J., Lee, J.H., Chang, S.K., 2005. Application of probabilistic neural networks for prediction of concrete strength. *Journal of Materials in Civil Engineering*, **17**(3):353-362. [doi:10.1061/(ASCE)0899-1561(2005)17:3(353)]

Kraaijveld, M.A., 1996. A Parzen classifier with an improved robustness against deviations between training and test data. *Pattern Recognition Letters*, **17**(7):679-689. [doi:10.1016/0167-8655(96)00013-X]

Lam, J.P., 2002. Evaluation of Concrete Shrinkage and Creep Prediction Models. MS Thesis, San Jose State University, USA.

Lee, J.J., Kim, D., Chang, S.K., Nocete, C.F.M., 2009. An improved application technique of the adaptive probabilistic neural network for predicting concrete strength. *Computational Materials Science*, **44**(3):988-998. [doi:10.1016/j.commatsci.2008.07.012]

Lilliefors, H., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62**(318):399-402.

Masters, T., 1995. Advanced Algorithms for Neural Networks. John Wiley, New York, p.55-56.

MOC (Ministry of Construction of the People's Republic of China), 1991. Standard for Test Method of Basic Properties of Construction Mortar. National Standards of People's Republic of China (JGJ 70-90) (in Chinese).

RILEM TC-107-GCS, 1995. Creep and shrinkage prediction models for analysis and design of concrete structures-Model B3. *Material Structure*, **28**(6):357-365. [doi:10.1007/BF02473152]

Sarıdemir, M., Topçu, İ.B., Özcan, F., Severcan, M.H., 2009. Prediction of long-term effects of GGBFS on compressive strength of concrete by artificial neural networks and fuzzy logic. *Construction and Building Materials*, **23**(3):1279-1286. [doi:10.1016/j.conbuildmat.2008.07.021]

Simon, L., Karim, M.N., 2001. Probabilistic neural networks using Bayesian decision strategies and a modified Gompertz model for growth phase classification in the batch culture of Bacillus subtilis. *Biochemical Engineering Journal*, **7**(1):41-48. [doi:10.1016/S1369-703X(00)00102-9]

Shaffer, R.E., Rose-Pehrsson, S.L., 1999. Improved probabilistic neural network algorithm for chemical sensor array pattern recognition. *Analytical Chemistry*, **71**(19):4263-4271. [doi:10.1021/ac990238+]

Specht, D.F., 1988. Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. IEEE International Conference on Neural Networks, San Diego, CA, **1**:525-532. [doi:10.1109/ICNN.1988.23887]

Steenhoek., L.W., 1999. A Probabilistic Neural Network Computer Vision System for Corn Kernel Damage Evaluation. PhD Thesis, Iowa State University, USA, p.12-14.

Thibault, J., Grandjean, B.P.A., 1991. A neural network methodology for heat transfer data analysis. *International Journal of Heat and Mass Transfer*, **34**(8):2063-2070. [doi:10.1016/0017-9310(91)90217-3]

Topçu, İ.B., Sarıdemir, M., 2007. Prediction of properties of waste AAC aggregate concrete using artificial neural network. *Computational Materials Science*, **41**(1):117-125. [doi:10.1016/j.commatsci.2007.03.010]

Wang, J.Z., Ni, H.G., He, J.Y., 1999. The application of automatic acquisition of knowledge to mix design of concrete. *Cement and Concrete Research*, **29**(12):1875-1880. [doi:10.1016/S0008-8846(99)00152-0]

Wang, Y., Adali, T., Kung, S.Y., Szabo, Z., 1998. Quantification and segmentation of brain tissues from MR images—a probabilistic neural network approach. *IEEE Transactions on Image Processing*, **7**(8):1165-1181. [doi:10.1109/83.704309]

Yan, P.F., Zhang, C.S., 2002. Artificial Neural Networks and Evolutionary Computing. Tsinghua University Press, Beijing, China, p.51-60 (in Chinese).

Yang, Z.R., Platt, M.B., Platt, H.D., 1999. Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research*, **44**(2):67-74. [doi:10.1016/S0148-2963(97)00242-7]