NFI: A Neuro-Fuzzy Inference Method for Transductive Reasoning

Qun Song and Nikola K. Kasabov, *Senior Member, IEEE*

Abstract—This paper introduces a novel neural fuzzy inference method—NFI for transductive reasoning systems. NFI develops further some ideas from DENFIS—dynamic neuro-fuzzy inference systems for both online and offline time series prediction tasks. While inductive reasoning is concerned with the development of a model (a function) to approximate data in the whole problem space (induction), and consecutively—using this model to predict output values for a new input vector (deduction), in transductive reasoning systems a local model is developed for every new input vector, based on some closest to this vector data from an existing database (also generated from an existing model). NFI is compared with both inductive connectionist systems (e.g., MLP, DENFIS) and transductive reasoning systems (e.g., K-NN) on three case study prediction/identification problems. The first one is a prediction task on Mackey Glass time series; the second one is a classification on Iris data; and the last one is a real medical decision support problem of estimating the level of renal function of a patient, based on measured clinical parameters for the purpose of their personalised treatment. The case studies have demonstrated better accuracy obtained with the use of the NFI transductive reasoning in comparison with the inductive reasoning systems.

Index Terms—Adaptive systems, neural-fuzzy inference (NFI), renal function evaluation, time series prediction, transductive reasoning.

I. INDUCTIVE VERSUS TRANSDUCTIVE LEARNING AND REASONING SYSTEMS

MOST learning models and systems in artificial intelligence developed and implemented so far [31], especially in the area of soft computing [10], [11], [13], [19], [28], [32], [37], [40], and particularly—in neuro-fuzzy reasoning systems [5], [6], [8], [20], [21], [30] are based on *inductive inference* methods, where a model (a function) is derived from data representing the problem space and this model is further applied on new data. The model is usually created without taking into account any information about a particular new data vector (test data). An error is measured to estimate how well the new data fits into the model.

The models are in most cases global models, covering the whole problem space. Such models are for example: regression functions; the multilayer perceptron neural network (MLP) used in this paper to compare results with, and also—the ANFIS neuro-fuzzy inference system [20]. These models are difficult
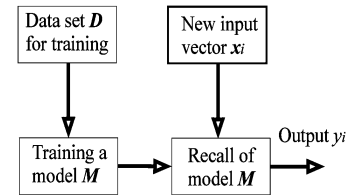
Fig. 1. Block diagram of an *inductive* reasoning system. A global model M is created based on data samples from D and then recalled for every new vector $\mathbf{x}_i$.

to update on new data without using old data, previously used to derive the models. Creating a global model (function) that would be valid for the whole problem space is a difficult task, and in most cases—it is not necessary to solve. In some connectionist and also fuzzy inference systems [11], [16], [23] the global model learned in a system consists of many local models (rules) that collectively cover the whole space and are adjusted incrementally on new data. The output for a new vector is calculated based on the activation of one or several neighboring local models (rules). Such systems are the evolving connectionist systems (ECOS) [23], for example—EFuNN [22] and DENFIS [24].

The inductive learning and inference approach is useful when a global model ("the big picture") of the problem is needed even in its very approximate form. In some models (e.g., ECOS) it is possible to apply incremental, online learning to adjust this model on new data and trace its evolution.

In contrast to the inductive learning and inference methods, transductive inference methods estimate the value of a potential model (function) only in a single point of the space (the new data vector) utilizing additional information related to this point [36]. This approach seems to be more appropriate for clinical and medical applications of learning systems, where the focus is not on the model, but on the individual patient. Each individual data vector (e.g., a patient in the medical area [1], [3]; a future time moment for predicting a time series [9], [15]; or a target day for predicting a stock index [17]) may need an individual, local model that best fits the new data, rather then—a global model. In the latter case the new data is matched into a model without taking into account any specific information about this data.

Transductive inference is concerned with the estimation of a function in a single point of the space only. For every new input vector $\mathbf{x}_i$ that needs to be processed for a prognostic task, the N_i nearest neighbors, which form a sub-data set D_i, are derived from an existing data set D and, if necessary, generated from an existing model M. A new model M_i is dynamically created from these samples to approximate the function in the point $\mathbf{x}_i$ (see Figs. 2 and 3). The system is then used to calculate the output value $\mathbf{y}_i$ for this input vector $\mathbf{x}_i$ (Figs. 2 and 3).
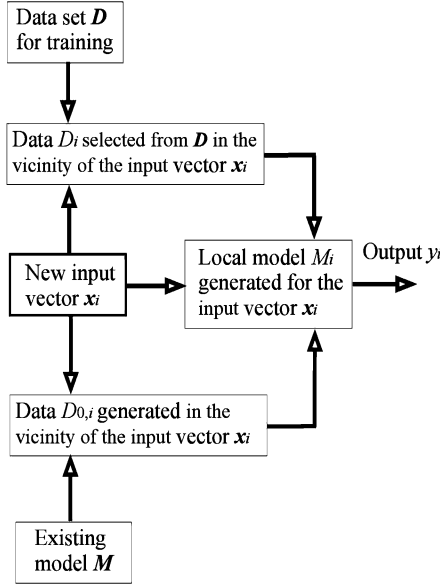
Fig. 2. Block diagram of a *transductive* reasoning system. An individual model M_i is trained for every new input vector $\mathbf{x}_i$ with the use of samples D_i selected from a data set D, and samples $D_{0,i}$ generated from an existing model (formula) M (*if such a model is existing*). The data samples in both D_i and $D_{0,i}$ are similar to the new vector $\mathbf{x}_i$ according to defined similarity criteria.

Before the proposed method, called NFI, is discussed in detail, a very simple transductive inference method—k-nearest neighbor method (K-NN) is briefly introduced to compare with NFI. In the K-NN method, the output value y_i for a new vector $\mathbf{x}_i$ is calculated as the average of the output values of the k nearest samples from the data set D_i. In the weighted K-NN method (WKNN) the output y_i is calculated based on the distance of the K-NN samples to $\mathbf{x}_i$:

$$y_i = \frac{\sum_{j=1}^{Ni} w_j y_j}{\sum_{j=1}^{Ni} w_j} \qquad (1)$$

where y_j is the output value for the sample $\mathbf{x}_j$ from D_i and w_j are their weights measured as

$$w_j = \frac{\max(\mathbf{d}) - [d_j - \min(\mathbf{d})]}{\max(\mathbf{d})}. \qquad (2)$$

In (2), the vector $\mathbf{d} = [d_1, d_2 \dots d_{Ni}]$ is defined as the distances between the new input vector $\mathbf{x}_i$ and N_i nearest neighbors $(\mathbf{x}_j, y_j)$ for $j = 1$ to N_i; $\max(\mathbf{d})$ and $\min(\mathbf{d})$ are the maximum and minimum values in $\mathbf{d}$, respectively. The weights w_j have the values between $\min(\mathbf{d})/\max(\mathbf{d})$ and 1; the sample with the minimum distance to the new input vector has the weight value of 1, and it has the value $\min(\mathbf{d})/\max(\mathbf{d})$ in case of maximum distance. Although both K-NN (or WKNN) and NFI are trainsductive models, for most cases, the NFI has much higher accuracy than K-NN because a K-NN model simply takes the average value of outputs of the nearest neighbors as its output, while the NFI uses the samples to create and train a local fuzzy inference system.

This paper is organized as follows. Section II gives the description of NFI and its learning algorithm and in Section III, different NFI models are applied on two benchmark data sets—*Mackey–Glass* and *Iris*, and on a clinical data set—glomerular filtration rate (GFR) [29]. The results are compared with the results obtained by the use of MLP, DENFIS, ANFIS and WKNN. Section IV presents a detailed comparative analysis on both the accuracy and time complexity of the presented NFI model and one global model—in this case, ANFIS. Conclusions and directions for further research are presented in Section V.

II. NFI: A NEURO-FUZZY INFERENCE METHOD FOR TRANSDUCTIVE REASONING

A. General Principles

NFI is a dynamic neural-fuzzy inference system with a local generalization, in which, either *Zadeh–Mamdani* [7], [38], or *Takagi–Sugeno* [35] type fuzzy inference is used. The local generalization means that in a sub-space of the whole problem space (local area) a model is crated that performs generalization in this area. In the NFI model, *Gaussian* fuzzy membership functions are applied in each fuzzy rule for both antecedent and consequent parts (*Zadeh–Mamdani* type), or for the antecedent part only (*Takagi–Sugeno* type). A back propagation (BP) [2] learning algorithm is used for optimizing the parameters of the fuzzy membership functions (in both *Zadeh–Mamdani* and *Takagi–Sugeno* types). The distance between vectors $\mathbf{x}$ and $\mathbf{y}$ is measured in NFI as the *normalized Euclidean distance* defined as follows (the values are between 0 and 1):

$$\|\mathbf{x} - \mathbf{y}\| = \left[\frac{1}{P} \sum_{j=1}^{P} |x_j - y_j|^2 \right]^{1/2} \qquad (3)$$

where $\mathbf{x}, \mathbf{y} \in \mathbf{R}^P$.

To partition the input space for creating fuzzy rules and obtaining initial values of fuzzy rules, the evolving clustering method (ECM) is applied [24] and the cluster centres and cluster radiuses are respectively taken as initial values of the centres and widths of the *Gaussian* membership functions (for both *Zadeh–Mamdani* and *Takagi–Sugeno* types). The *ECM* performs a scatter partition that has relatively small number of clusters covering the space. The data in a cluster are used for creating a linear function (*Takagi–Sugeno* type fuzzy inference) as a local model for output function evaluation as described here.

B. NFI Learning Algorithm

Suppose that data have been normalized (the values are between 0 and 1) and, for each new data vector $\mathbf{x}_q$, the NFI performs the following learning algorithm, as also shown in Fig. 4.

1) Search in the training data set in the input space to find N_q training examples that are closest to $\mathbf{x}_q$. The value for N_q can be pre-defined based on experience, or—optimized through the application of an optimization procedure. Here we assume the former approach.
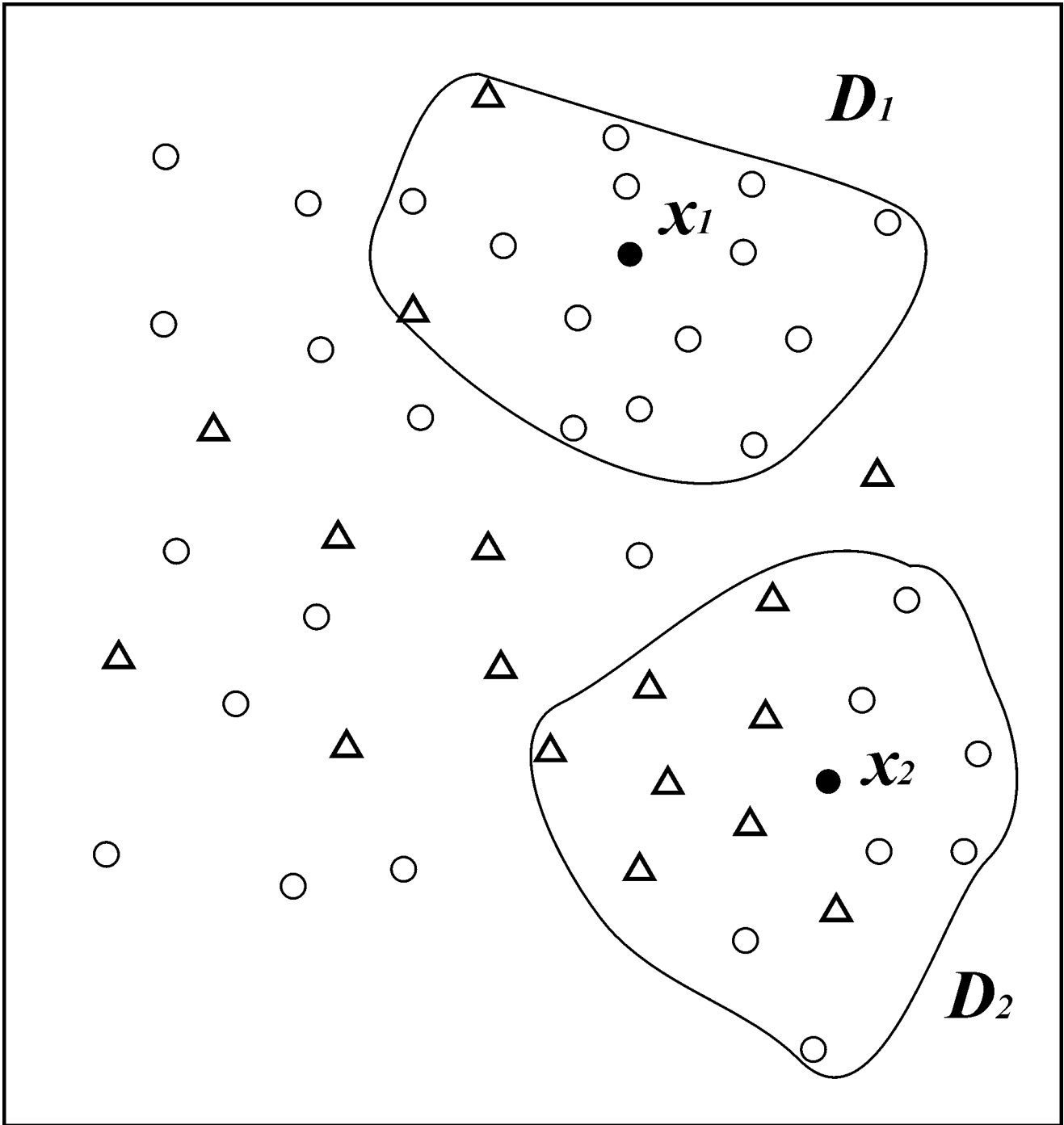
Fig. 3.   In the center of a transductive reasoning system is the new data vector (here illustrated with two of them—$\mathbf{x}_1$ and $\mathbf{x}_2$), surrounded by a fixed number of nearest data samples selected from the training data D and generated from an existing model M.

2) Calculate the distances d_i, $i = 1, 2, \ldots, N_q$, between each of these data samples and $\mathbf{x}_q$. And calculate the weights $w_i = 1 - (d_i - \min(\mathbf{d}))$, $i = 1, 2, \ldots, N_q$, $\min(\mathbf{d})$ is the minimum value in the distance vector $\mathbf{d} = [d_1, d_2, \ldots, d_{Nq}]$.

3) Use the *ECM* clustering algorithm to cluster and partition the input sub-space that consists of N_q selected training samples.

4) Create fuzzy rules and set their initial parameter values according to the *ECM* clustering procedure results; for each cluster, the cluster centre is taken as the centre of a fuzzy membership function (*Gaussian* function) and the cluster radius is taken as the width.

5) Apply the steepest descent method (back-propagation) to optimize the parameters of the fuzzy rules in the local model M_q following (4)–(23).

6) Calculate the output value y_q for the input vector x_q applying fuzzy inference over the set of fuzzy rules that constitute the local model M_q.
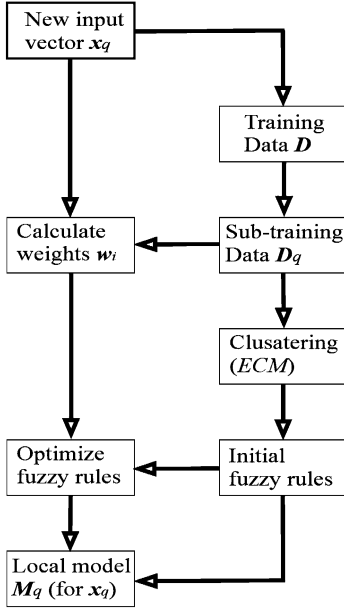
7) End of the procedure.

Fig. 4.   Block diagram of the proposed NFI learning algorithm.

The parameter optimization procedure is described below:

Consider the system having P inputs, one output and M fuzzy rules defined initially through the *ECM* clustering procedure, the lth rule has the form of

$$R_1 : \text{If } x_1 \text{ is } F_{l1} \text{ and } x_2 \text{ is } F_{l2} \text{ and} \ldots x_p \text{ is } F_{lp}$$
$$\text{then } y \text{ is } G_l. \quad (\text{Zadeh- Mamdani type}) \quad (4)$$

or

$$R_1 : \text{If } x_1 \text{ is } F_{l1} \text{ and } x_2 \text{ is } F_{l2} \text{ and} \ldots x_p \text{ is } F_{lp}$$
$$\text{then } y \text{ is } n_l. \quad (\text{Takagi- Sugeno type}). \quad (5)$$

Here, F_{lj} are fuzzy sets defined by the following *Gaussian* type membership function:

$$\text{GaussianMF} = \alpha \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad (6)$$

and G_l are of a similar type as F_{lj} and are defined as

$$\text{GaussianMF} = \exp\left[-\frac{(y-n)^2}{2\delta^2}\right]$$
$$(\text{for Zadeh- Mamdani type}) \quad (7)$$

or

$$n_1 = b_{l0} + b_{l1}x_1 + b_{l2}x_2 + \cdots + b_{lp}x_p :$$
$$(\text{for Takagi- Sugeno type}). \quad (8)$$

Using the *modified centre average defuzzification* procedure [5] the output value of the system can be calculated for an input vector $\mathbf{x}_i = [x_1, x_2, \ldots, x_P]$ as follows:

$$f(\mathbf{x_i}) = \frac{\sum_{l=1}^{M} \frac{G_l}{\delta_l^2} \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{(x_{ij}-m_{lj})^2}{2\sigma_{lj}^2}\right]}{\sum_{l=1}^{M} \frac{1}{\delta_l^2} \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{(x_{ij}-m_{lj})^2}{2\sigma_{lj}^2}\right]}$$
$$(\text{for Zadeh- Mamdani type}) \quad (9)$$

or:

$$f(\mathbf{x_i}) = \frac{\sum_{l=1}^{M} n_l \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{(x_{ij}-m_{lj})^2}{2\sigma_{lj}^2}\right]}{\sum_{l=1}^{M} \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{(x_{ij}-m_{lj})^2}{2\sigma_{lj}^2}\right]}$$
$$(\text{for Takagi- sugeno type}). \quad (10)$$

Suppose the NFI is given a training input-output data pair $[\mathbf{x}_i, t_i]$, the system minimizes the following objective function (a weighted error function):

$$E = \frac{1}{2} wi \left[f(xi) - ti\right]^2 \quad (w_i \text{ are defined in step 2}). \quad (11)$$

The steepest descent algorithm (BP) [2] is used then to obtain the formulas for the optimization of the parameters G_l, δ_l, α_{lj}, m_{lj} and σ_{lj} of *Zadeh–Mamdani* type NFI such that the value of E from (11) is minimized

$$G_l(k+1) = G_l(k) - \frac{\eta_G}{\delta_l^2(k)} wi\Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right] \quad (12)$$

$$\delta_l(k+1) = \delta_l(k) - \frac{\eta_\delta}{\delta_l^3(k)} w_i\Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - G_l(k)\right] \quad (13)$$

$$\alpha_{lj}(k+1) = \alpha_{lj}(k) - \frac{\eta_\alpha}{\delta_l^2(k)\alpha_{lj}(k)} w_i\Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[G_l(k) - f^{(k)}(\mathbf{x}_i)\right] \quad (14)$$

$$m_{lj}(k+1) = m_{lj}(k) - \frac{\eta_m}{\delta_l^2(k)\sigma_{lj}^2(k)} w_i\Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - ti\right]$$
$$\times \left[G_l(k) - f^{(k)}(\mathbf{x}_i)\right]$$
$$\times [x_{ij} - m_{lj}(k)] \quad (15)$$

$$\sigma_{lj}(k+1) = \sigma_{lj}(k) - \frac{\eta_\sigma}{\delta_l^2(k)\sigma_{lj}^3(k)} w_i\Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[G_l(k) - f^{(k)}(\mathbf{x}_i)\right]$$
$$\times [x_{ij} - m_{lj}(k)]^2 \quad (16)$$

here

$$\Phi(\mathbf{x}_i) = \frac{\prod\limits_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{[x_{ij}(k)-m_{lj}(k)]^2}{2\sigma_{lj}^2(k)}\right\}}{\sum\limits_{l=1}^{M} \frac{1}{\delta_l^2} \prod\limits_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{[x_{ij}(k)-m_{lj}(k)]^2}{2\sigma_{lj}^2(k)}\right\}}. \quad (17)$$

The steepest descent algorithm (BP) is also used to obtain the formulas for the optimization of the parameters $\mathbf{b}_l$, α_{lj}, m_{lj} and σ_{lj} of the *Takagi–Sugeno* type NFI such that the value of E from (11) is minimized

$$b_{l0}(k+1) = b_{l0}(k) - \eta_b w_i \Phi(\mathbf{x}_i)\left[f^{(k)}(\mathbf{x}_i) - t_i\right] \quad (18)$$

$$b_{lj}(k+1) = b_{lj}(k) - \eta_b x_{ij} w_i \Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right] \quad (19)$$

$$\alpha_{lj}(k+1) = \alpha_{lj}(k) - \frac{\eta_\alpha}{\alpha_{lj}(k)} w_i \Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[n_l(k) - f^{(k)}(\mathbf{x}_i)\right] \quad (20)$$

$$m_{lj}(k+1) = m_{lj}(k) - \frac{\eta_m}{\sigma_{lj}^2(k)} w_i \Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[n_l(k) - f^{(k)}(\mathbf{x}_i)\right]$$
$$\times [x_{ij} - m_{lj}(k)] \quad (21)$$

$$\sigma_{lj}(k+1) = \sigma_{lj}(k) - \frac{\eta_\sigma}{\sigma_{lj}^3(k)} w_i \Phi(\mathbf{x}_i)$$
$$\times \left[f^{(k)}(\mathbf{x}_i) - t_i\right]$$
$$\times \left[n_l(k) - f^{(k)}(\mathbf{x}_i)\right]$$
$$\times [x_{ij} - m_{lj}(k)]^2 \quad (22)$$

here

$$\Phi(\mathbf{x}_i) = \frac{\prod\limits_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{[x_{ij}(k)-m_{lj}(k)]^2}{2\sigma_{lj}^2(k)}\right\}}{\sum\limits_{l=1}^{M} \prod\limits_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{[x_{ij}(k)-m_{lj}(k)]^2}{2\sigma_{lj}^2(k)}\right\}} \quad (23)$$

where η_G, η_δ, η_b, η_α, η_m and η_σ are learning rates for updating the parameters G_l, δ_l, $\mathbf{b}_j$, α_{lj}, m_{lj} and σ_{lj} respectively.

In the NFI training algorithm, the following indexes are used
- training data samples: $i = 1, 2, \ldots, N$;
- tnput variables: $j = 1, 2, \ldots, P$;
- tuzzy rules: $l = 1, 2, \ldots, M$;
- tearning epochs: $k = 1, 2, \ldots$.

III. APPLICATIONS OF THE TRANSDUCTIVE NFI METHOD ON THREE CASE STUDY PROBLEMS: PREDICTION, CLASSIFICATION, AND IDENTIFICATION

A. Time Series Modeling and Prediction

In this section, the NFI model is applied for modeling and predicting the future values of a chaotic time series—the Mackey–Glass (MG) data set [12], [15] which has been used as
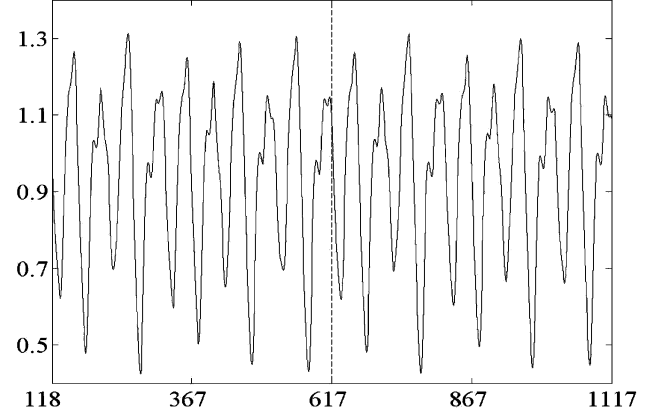

Fig. 5. Mackey–Glass case study data: the first half (118–617) is used as training data, and the next half (618–1117)—as testing data.

TABLE I
TEST RESULTS ON MG DATA

Model	Test NDEI
WKNN	0.06
CC - NN model	0.06
6th-order Polynomial	0.04
MLP (BP)	0.02
HyFIS [27]	0.01
ANFIS [20]	0.007
DENFIS [24]	0.006
NFI	**0.004**

a bench-mark problem in the areas of neural networks, fuzzy systems and hybrid systems. Here, a *Takagi–Sugeno* type NFI is used. This time series is created with the use of the MG time-delay differential equation defined here

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t). \quad (24)$$

To obtain values at integer time points, the *fourth-order Runge–Kutta* method was used to find the numerical solution to the previous MG equation. Here we assume that: the time step is 0.1; $x(0) = 1.2$; $\tau = 17$; and $x(t) = 0$ for $t < 0$. The task is to predict the values $x(t + 6)$ from input vectors $[x(t - 18), x(t - 12), x(t - 6), x(t)]$ for any value of the time t. The following experiment was conducted: 1000 data points, from $t = 118$ to 1117, were extracted, the first 500 data was taken as the training data and another 500 as testing data. For each of the testing data sample a local transductive model is created and tested on this data. Fig. 5 displays the target data. Table I lists the simulating results represented by NDEI—*nondimensional error index* which is defined as the *root mean square error* (RMSE) divided by the standard deviation of the target series. For the purpose of a comparative analysis, we have quoted the prediction results on the same data produced by some other methods [27], which are also listed in Table I.

TABLE II
CLASSIFICATION TEST RESULTS ON IRIS DATA

Model	Nodes Rn	Test Error
ECF [25]	12	4
ECMC [23]	9.8	3.8
ZISC 39]	14.4	6.1
MLP	12	4.6
NFI	**6.3**	**3.3**

The NFI transductive reasoning system performs better than the other inductive reasoning models. This is a result of the fine tuning of each local model in NFI for each tested example, derived according to the NFI learning procedure. The finely tuned local models achieve a better local generalization.

B. Transductive NFI For Classification

In this section, the transductive NFI (*Zadeh–Mamdani* type) is used as a classifier on a benchmark data set—*Iris* data. *Iris* is perhaps the best known data set to be found in the pattern recognition and classification literature. The data set contains three classes of 50 instances with four numeric, predictive attributes: 1) sepal length; 2) sepal width; 3) petal length; 4) petal width. Each class refers to a type of *iris* plant. One class, *Iris* Setosa, is linearly separable from the other two classes, *Iris* Versicolor and *Iris* Virginica; the latter are not linearly separable from each other.

To compare the performance of the NFI method we conducted the following experiments. We used ECMC (evolving clustering based classifier) [34], ECF (evolving classifying function) [25], ZISC [39] and a multiplayer perceptron (MLP) on the same data. All experimental results listed in Table II are based on 10 experiments with the same model and parameters and the results (*Rn*—number of neurons or rules, and *TestError*—the number of errors on the testing data) are averaged. In each experiment 50% of the whole data set is randomly selected as training data and another 50% as testing data.

The results from Table II show that NFI outperforms the other classification methods.

C. Transaductive NFI For Personalised Modeling in Medical Decision Support Systems

The transductive NFI method can be applied for the creation of a personalised model of a patient for a medical diagnosis, disease prognosis, and treatment planning. Personalised medicine is a growing area of research where this method can be utilized broadly.

Here, the transductive NFI (*Zadeh–Mamdani* type) is applied for the creation of personalised (individualised) models for the evaluation (identification) of renal function of patients in a renal clinic. Real data is used collected in a clinical environment.

The accurate evaluation of renal function is fundamental to sound nephrology practice. The early detection of renal impairment will allow for the institution of appropriate diagnostic and therapeutic measures, and potentially maximize preservation of intact nephrons [18].

TABLE III
COMPARISON BETWEEN THE TEST ERROR FOR GFR RENAL FUNCTION EVALUATION WITH THE USE OF THE PROPOSED NFI TRANSDUCTIVE REASONING METHOD VERSUS THE MDRD FORMULA, MLP, DENFIS, AND WKNN MODELS

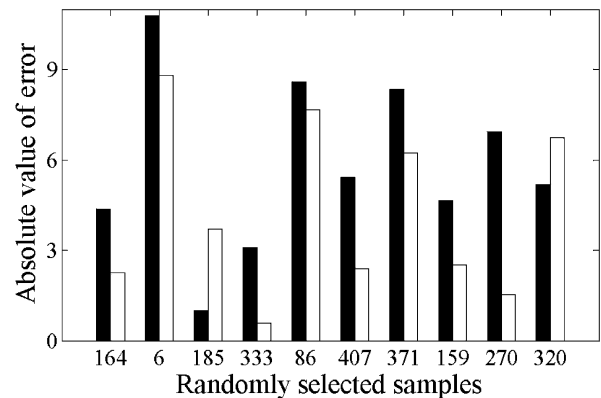| Model	Nodes Rn	MAE	RMSE
MDRD (inductive)[29]	NA	5.88	7.74
WKNN (transductive)	NA	6.43	8.36
MLP (inductive)	12	5.75	8.44
ANFIS (inductive) [20]	27	5.48	7.49
DENFIS (inductive) [24]	25	5.35	7.25
NFI (transductive)	**8.6**	**5.17**	**7.13**

Fig. 6. Comparison of the absolute error of a global ANFIS model trained for the prediction of GFR and tested on 10 new randomly selected samples, and the error of personalised NFI models built individually for each of these 10 samples (dark color bars represent ANFIS errors, and light color bars represents NFI errors).

Glomerular filtration rate (GFR) is traditionally considered the best overall index to determine renal function in healthy and in diseased people. Most clinicians rely upon the clearance of creatinine (CrCl) as a convenient and inexpensive surrogate for GFR. CrCl can be determined by either timed urine collection, or from serum creatinine using equations developed from regression analyzes such as that by Cockcroft–Gault formula, but the accuracy of CrCl is limited by methodological imprecision and the systematic bias [29].

Recently, the modification of diet in renal disease (MDRD) study group developed a new formula to more accurately evaluate the GFR [29]. The formula uses six input variables: age, gender, Screat, Race, Salb and Surea and is defined as follows:

$$\text{GFR} = 170 \times \text{Screat}^{-0.999} \times \text{Age}^{-0.176} \times 0.762(\text{ if female})$$
$$\times 1.18(\text{ if race is black}) \times \text{Surea}^{-0.17} \times \text{Salb}^{0.318}. \quad (25)$$

In this formula, Screat (Serum creatinine) is a protein which is expected to be filtered in the kidneys and the residual of it—released into the blood. The creatinine level in the serum is determined by the rate it is being removed in the kidney and is also a measure of the kidney function. Surea (Serum urea) is a substance produced in the liver as a means of disposing of ammonia from protein metabolism. It is filtered by the kidney and can be reabsorbed to the bloodstream. Salb (Serum albumin) is the protein of the highest concentration in plasma. Decreased

TABLE IV
AVERAGE EXPERIMENTAL RESULTS OF THE ANFIS GLOBAL MODEL AND THE NFI PERSONALIZED MODELS ON THE 10 RANDOMLY SELECTED
SAMPLES FORM FIG. 6

Model	Number of training samples	Number of training iterations	Number of average fuzzy rules	Average training time (CPU time)	RMSE	MAE
ANFIS	440	50	64	477.6 sec.	6.45	5.84
NFI	32	50	7.1	29.9 sec.	5.04	4.24

serum albumin may result from kidney disease, which allows albumin to escape into the urine. Decreased albumin may also be explained by malnutrition or liver disease.

However, prediction of GFR with the use of the existing formulas that constitute global and fixed models, can be misleading as to the presence and progression of renal disease [29]. Here, the transductive NFI method is applied for the prediction of the GFR through the construction of an individual model for each new patient.

Using NFI on a GFR data set (447 samples), collected from hospitals in New Zealand and Australia, we have obtained more accurate results than with the use of the MDRD formula or the use of other connectionist and statistical techniques. The results are listed in Table III. For comparison, the results produced by the MDRD formula (a global regression model), the MLP (a globally trained connectionist model), the ANFIS (a global TSK fuzzy model) and DENFIS (a global model that is a set of adaptive local models), all—inductive reasoning systems, along with the results produced by using the transductive WKNN method, are also listed in the table.

All experimental results reported here are based on 10 cross-validation experiments with the same model and parameter values, and the results are averaged. In each experiment, 70% of the whole data set is randomly selected as training data and another 30% as testing data. The results are listed in Table III. They include: Rn—the number of fuzzy rules (for ANFIS, DENFIS and for NFI); number of neurons in the hidden layer (for the MLP model), the test root mean square error (RMSE), and the test mean absolute error (MAE).

The transductive NFI system gives the best accuracy of the GFR evaluation for each individual patient and overall—for the whole data set.

IV. COMPARATIVE ANALYSIS AND DISCUSSIONS

Here, we compare the transductive NFI method with global modeling techniques in terms of accuracy of the model and the time required. It was shown in Table III that a personalised model is at average more accurate for the prediction of the output values for new input vectors $\mathbf{x}$ than a general model, e.g., ANFIS, MLP, when recalled on $\mathbf{x}$. This is illustrated more in detail in Fig. 6 and Table IV on randomly selected ten samples from the GFR renal function identification problem. For each of the 10 test samples a global ANFIS model is created (using the rest of the samples), along with a personalised NFI model (using only some nearest samples). The error in the predicted values is compared between the two models for each of the ten
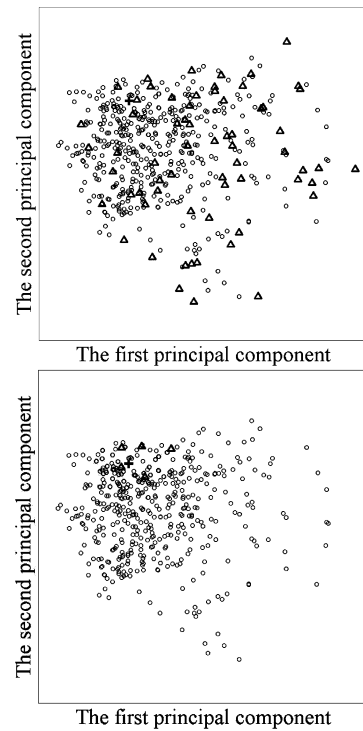

Fig. 7. (a) Training samples and rule nodes of ANFIS in 2D PCA space. Denotation: "o:" all 446 samples; "+:" sample no. 270; "$\triangle$:" 64 rule nodes. (b) Training samples and rule nodes of NFI in the 2-D PCA space. Denotation: "o:" samples (light); "o:" selected 32 nearest samples (dark); "+:" sample no. 270; "$\triangle$:" five rule nodes (centres of Gaussian membership functions). The corresponding rules are given in Table VI.

individual samples. The error produced by the NFI approach in most of the cases is smaller than the error of the ANFIS model. Accuracy of prediction is the single most important issue in medical prognostic systems as that saves lives and resources due to a more accurate and personalised treatment and medication of each patient.

Transductive NFI modeling may require a longer time for the creation of a model for a single input vector $\mathbf{x}$ than recalling a global model (e.g., ANFIS, MLP) on this vector. But the time for a personalised model creation is not at all long as the model is created with the use of only a subset of samples from the whole problem space as it is shown in Fig. 7(a) and (b). The time for the first phase of searching for the nearest samples to the new one depends on the size of the data set and the search algorithm. In terms of time required, the time to create a personalised model is much shorter (e.g., an order of magnitude) than the time for the creation of a global model, if for every

TABLE V
EXPERIMENTAL RESULTS OF ANFIS AND NFI ON PATIENT DATA SAMPLE NO. 270

Patient data (No.270)	x_1 Age	x_2 Gender	x_3 Screat	x_4 Surea	x_5 Face	x_6 Salb	y GFR (desired)
	59.7	Female	0.22	21.8	White	43	30.0

Model	Training samples	Training iterations	Number of rules	Training time (CPU time)	Prediction results
ANFIS	446	50	64	476.4 sec.	36.9
NFI	32	50	5	21.2 sec.	28.5

new sample a global model needs to be created (see Table IV). This is the case when a global model cannot be adapted on incoming data continuously. Real time use of personalised modeling would be possible when dealing with not very large data sets, otherwise some prefiltering of the data may be needed. For the GFR case study problem, the average time for the creation of a NFI model is about 20 sec on a standard personal computer (2.4 MHz Pentium; using an NFI MATLAB uncompiled code). The time for a global ANFIS model creation is evaluated above 450 sec when a compiled MATLAB code of ANFIS is run.

A major advantage of the personalised modeling is that for a single vector (a single patient data) an individual model is created that can be used to explain the calculated output value. Tables V and VI illustrate this for sample no. 270 from the ten randomly selected samples in Fig. 6. Table V compares the results from the NFI personalised model and the ANFIS global model, while Table VI shows the five fuzzy rules (first-order Takagi–Sugeno type) that comprise this NFI model. Each rule represents an output function for each of the five clusters that are defined for the 32 nearest samples to the sample no. 270. The rules constitute also a good explanation on the predicted value for this patient.

If a single linear regression was used for personalized modeling instead of the NFI method, one linear function would have been derived for the whole set of 32 samples, that would result in a less accurate prediction.

V. CONCLUSION AND FURTHER RESEARCH

This paper presents a transductive neuro-fuzzy inference method—NFI, for prediction, classification, and identification tasks. Both *Zadeh–Mamdani* and *Takagi–Sugeno* type fuzzy models are developed. The NFI transductive inference method has several advantages when compared to the previously developed inductive inference methods on the some data sets.

1) It performs a better local generalization over new data as it develops an individual model for each data vector that takes into account the new input vector location in the space. This type of modeling can be called "personalised," and it is promising for medical decision support systems.

2) It works well in a large dimensional space of many variables (for example—thousands of genes, proteins and clinical variables), but with a relatively small number of examples. According to [36] a sample size is considered

TABLE VI
EXTRACTED FUZZY RULES (FIRST-ORDER TAKAGI–SUGENO TYPE) FROM THE PERSONALISED NFI MODEL ON PATIENT DATA SAMPLE NO. 270. ALL VARIABLES ARE NORMALIZED IN $[0, 1]$ INTERVAL AND EACH GAUSSIAN MEMBERSHIP FUNCTION HAS PARAMETERS (α, m AND σ) DEFINED IN (6)

Rule 1: if x_1 is GaussianMF(0.90   0.72   0.29)
   x_2 is GaussianMF(0.88   1.00   0.30)
   x_3 is GaussianMF(0.93   0.08   0.30)
   x_4 is GaussianMF(0.90   0.28   0.30)
   x_5 is GaussianMF(0.89   0.00   0.30)
   x_6 is GaussianMF(0.91   0.70   0.29)
then $y = 0.41 - 0.047x_1 + 0.085x_2 - 0.570x_3 - 0.197x_4 - 0.027x_5 + 0.036x_6$

Rule 2: if x_1 is GaussianMF(0.97   0.67   0.30)
   x_2 is GaussianMF(0.78   1.00   0.30)
   x_3 is GaussianMF(0.99   0.11   0.30)
   x_4 is GaussianMF(0.97   0.40   0.32)
   x_5 is GaussianMF(0.98   0.00   0.30)
   x_6 is GaussianMF(0.96   0.90   0.26)
then $y = 0.40 - 0.04x_1 + 0.084x_2 - 0.552x_3 - 0.201x_4 - 0.029x_5 + 0.047x_6$

Rule 3: if x_1 is GaussianMF(0.96   0.65   0.29)
   x_2 is GaussianMF(0.95   1.00   0.30)
   x_3 is GaussianMF(0.96   0.21   0.28)
   x_4 is GaussianMF(0.98   0.51   0.24)
   x_5 is GaussianMF(0.90   0.00   0.30)
   x_6 is GaussianMF(0.97   0.73   0.32)
then $y = 0.37 - 0.025x_1 + 0.078x_2 - 0.533x_3 - 0.171x_4 - 0.01x_5 + 0.01x_6$

Rule 4: if x_1 is GaussianMF(0.97   0.56   0.27)
   x_2 is GaussianMF(0.96   1.00   0.30)
   x_3 is GaussianMF(0.97   0.24   0.28)
   x_4 is GaussianMF(0.99   0.28   0.31)
   x_5 is GaussianMF(0.98   0.00   0.30)
   x_6 is GaussianMF(0.97   0.73   0.32)
then $y = 0.39 - 0.065x_1 + 0.081x_2 - 0.544x_3 - 0.226x_4 - 0.074x_5 + 0.059x_6$

Rule 5: if x_1 is GaussianMF(0.94   0.88   0.26)
   x_2 is GaussianMF(0.90   1.00   0.30)
   x_3 is GaussianMF(0.94   0.16   0.30)
   x_4 is GaussianMF(0.98   0.41   0.31)
   x_5 is GaussianMF(0.94   0.00   0.30)
   x_6 is GaussianMF(0.97   0.70   0.29)
then $y = 0.38 - 0.013x_1 + 0.086x_2 - 0.541x_3 - 0.167x_4 - 0.006x_5 + 0.035x_6$

small if the ratio $N/M < 20$, where N is the size of the data set and M is the VC (Vapnik-Chervonenkis) dimension—an estimate of the possible number of functions in the space for the defined problem and for the available data set.

3) The transductive NFI is an adaptive model, in the sense that input-output pairs of data can be added to the data set

continuously and immediately made available for transductive inference of local models.

4) The NFI method can be used at different times on different number of variables (different dimensions) and over data vectors characterized by missing values. A personalised model can be evolved with the use of only the available variables in the new vector that is not possible when using global models.

As the NFI method creates a unique submodel for each data sample, it usually needs more performing time than inductive models, especially in the case of training and simulating on large data sets. Although there are some simulating data samples that are same or very similar to each other, the NFI model will create the same or very similar models for them repeatedly. It is therefore advantageous to use both incremental, inductive reasoning (e.g., ECOS) to reveal a global model (the "big picture"), and the NFI transductive reasoning for accurate personalised inference and decision making, or to store some already evolved personalised models for a further use, which is one of our current research topics. Time complexity of the method depends mainly on the search algorithm—while searching for similar data to the new vector in a data base. The problem of the efficiency of search algorithms is beyond the scope of this paper.

Further directions for research include: 1) NFI system parameter optimization such as optimal number of nearest samples; 2) Using some signal processing techniques to improve the NFI model for dealing with noisy data; 3) Investigating different distance measure methods for the selection of the neighboring samples; 4) Developing nonlinear local models for each cluster of the selected samples; and 5) Applications of the NFI transductive inference method to other decision support problems, such as: cardio-vascular risk prognosis [3]; biological processes modeling and prediction based on gene expression micro-array data [4], [14]; business decision support.

REFERENCES

[1] A. I. Akl *et al.*, "Artificial intelligence: A new approach for prescription and monitoring of hemodialysis therapy," *Amer. J. Kidney Diseases*, vol. 38, no. 6, pp. 1277–1283, 2001.

[2] S. Amari, *Mathematical Foundations of Neuro-Computing*. New York: IEEE Press, 1990, vol. 78.

[3] K. M. Anderson *et al.*, "Cardiovascular disease risk profiles," *Amer. Heart J.*, pt. 2, vol. 121, no. 1, pp. 293–298, 1990.

[4] P. Baldi and S. Brunak, *Bioinformarics—A Machine Learning Approach*. Cambridge, MA: MIT Press, 2001.

[5] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[6] ——, *Analysis of Fuzzy Information*. Boca Raton, FL: CRC, 1987.

[7] J. C. Bezdek and S. K. Pal, Eds., *Fuzzy Models For Pattern Recognition—Methods That Search for Structures in Data*. New York: IEEE Press, 1992.

[8] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[9] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden Day, 1970.

[10] G. Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991.

[11] S. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, 1994.

[12] R. S. Crodder, *Predicting the Mackey-Glass Timeseries With Cascade-Corelation Learning*. Pittsburgh, PA: Connectionist Models Summer School, Carnegie Mellon Univ., 1990.

[13] G. Cybenko, "Approximation by super-positions of sigmoidal function," *Math. Control, Signals, Syst.*, vol. 2, pp. 303–314, 1989.

[14] P. D'Haeseleer *et al.*, "Gene expression data analysis and modeling, session on gene expression and genetic networks," in *Proc. Pacific Symp. Biocomputing*, Mauna Lani, HI, 1999, pp. 41–52.

[15] J. D. Farmer and J. J. Sidorowitch, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol. 59, no. 8, pp. 845–848, 1987.

[16] B. Fritzke, "A growing neural gas network learns topologies," *Adv. Neural Inform. Process. Syst.*, vol. 7, pp. 625–632, 1995.

[17] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[18] J.-Y. Guh *et al.*, "Prediction of equilibrated postdialysis BUN by an artificial neural network in high-efficiency hemodialysis," *Amer. J. Kidney Diseases*, vol. 31, no. 4, pp. 638–646, 1998.

[19] T. M. Heskes and B. Kappen, "On-line learning processes in artificial neural networks," *Math. Found. Neural Networks*, pp. 199–233, 1993.

[20] R. Jang, "ANFIS: Adaptive network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, Jun. 1993.

[21] N. Kasabov, *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*. Cambridge, MA: MIT Press, 1996.

[22] ——, "Evolving fuzzy neural networks for on-line supervised/unsupervised, knowledge-based learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 6, pp. 902–918, Dec. 2001.

[23] ——, *Evolving Connectionist Systems—Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines*. New York: Springer-Verlag, 2002.

[24] N. Kasabov and Q. Song, "DENFIS: Dynamic, evolving neural-fuzzy inference systems and its application for time-series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Apr. 2002.

[25] ——, "GA-Optimization of evolving connectionist systems for classification with a case study from bioinformatics," in *Proc. ICONIP'2002*, Singapore, Nov. 2002, pp. 602–605.

[26] V. Kecman, *Learning and Soft Computing—Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. Cambridge, MA: MIT Press, 2001.

[27] J. S. Kim and N. Kasabov, "HyFIS: Hybrid connectionist fuzzy inference for adaptive dynamic systems," *Neural Networks*, vol. 12, no. 9, pp. 1301–1321, 2001.

[28] T. Kohonen, *Self-Organizing Maps*. New York: Springer-Verlag, 1997.

[29] A. S. Levey *et al.*, "A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation," *Ann. Internal Med.*, vol. 130, no. 6, pp. 461–470, 1999.

[30] C. T. Lin and C. S. G. Lee, *Neuro Fuzzy Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[31] M. T. Mitchell, *Machine Learning*. New York: MacGraw-Hill, 1997.

[32] G. A. Rummery and M. Niranjan, *On-Line Q-Learning Using Connectionist Systems*: Cambridge University Engineering Department, 1994.

[33] D. Saad, *On-Line Learning in Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[34] Q. Song and N. Kasabov, "Weighted data normalizations and feature selection for evolving connectionist systems proceedings," in *Proc. 8th Australian and New Zealand Intelligence Information Systems Conf. (ANZIIS2003)*, Sydney, Australia, 2003, pp. 285–290.

[35] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan. 1985.

[36] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[37] R. R. Yager and L. A. Zadeh, Eds., *An Introduction to Fuzzy Logic Applications in Intelligent Systems*. Norwell, MA: Kluwer, 1992.

[38] L. A. Zadeh, "Fuzzy sets," *Inform. Control*, vol. 8, pp. 338–353, 1965.

[39] (2001) ZISC Manual, Silicon Recognition. [Online]. Available: www.silirec.com

[40] J. M. Zurada, *Introduction to Artificial Neural Systems*. Singapore: Info Access Distribution Pte Ltd., 1992.
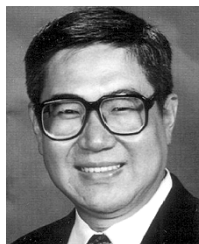
Qun Song received the B.E., M.M.M.S., and Ph.D. degrees from the Textile University of China, Tokyo University of Mercantile Marine, Japan, and Otago University, New Zealand, respectively.

The research fields he specializes in are AI systems, neural networks, fuzzy systems, and their applications. He is with the Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland, New Zealand, as a Senior Research Fellow.

Nikola K. Kasabov (SM'97) received the Ph.D. degree from TU Sofia, Bulgaria, in 1975.

He is currently the Director of the Knowledge Engineering and Discovery Research Institute and Personal Chair of Knowledge Engineering in the School of Information Technology, Auckland University of Technology, Auckland, New Zealand. He has published over 350 works, among them journal papers, text books, edited research books and monographs, conference papers, book chapters, edited conference proceedings, and patents and authorship certificates in the area of intelligent systems, connectionist and hybrid connectionist systems, fuzzy systems, expert systems, speech recognition, and bioinformatics.

Prof. Kasabov is a Fellow of the Royal Society of New Zealand and the New Zealand Computer Society, a member of the Board of Governors of the International Neural Network Society (INNS) since 1994, a Past President of the Asia Pacific Neural Network Assembly (APNNA), a member of the TC12 group on Artificial Intelligence of IFIP, and also a member of INNS, ENNS, and the IEEE Computer Society. He is the General Chairman of a series of biannual international conferences on Neurocomputing in New Zealand. He received The Royal Society of New Zealand Silver Medal for contribution to Science and Technology in 2001. He is an Associate Editor of two international journals and a leader of a National Research Programme funded by FRST.