

The Association Forecasting of 13 Variants Within Seven Asthma Susceptibility Genes on 3 Serum IgE Groups in Taiwanese Population by Integrating of Adaptive Neuro-fuzzy Inference System (ANFIS) and Classification Analysis Methods

Cheng-Hang Wang · Baw-Jhiune Liu ·
Lawrence Shih-Hsin Wu

Received: 21 December 2009 / Accepted: 28 February 2010
© Springer Science+Business Media, LLC 2010

Abstract Asthma is one of the most common chronic diseases in children. It is caused by complicated coactions between various genetic factors and environmental allergens. The study aims to integrate the concept of implementing adaptive neuro-fuzzy inference system (ANFIS) and classification analysis methods for forecasting the association of asthma susceptibility genes on 3 serum IgE groups. The ANFIS model was trained and tested with data sets obtained from 425 asthmatic subjects and 483 non-asthma subjects from the Taiwanese population. We assessed 13 single-nucleotide polymorphisms (SNPs) in seven well-known asthma susceptibility genes; firstly, the proposed ANFIS model learned to reduce input features from the 13 SNPs. And secondly, the classification will be used to classify the serum IgE groups from the simulated SNPs results. The performance of the ANFIS model, classification accuracies and the results confirmed that the integration of ANFIS and classified analysis has potential in association discovery.

Keywords Adaptive neuro-fuzzy inference system (ANFIS) · Classify analysis · Fuzzy logic · Pediatric asthma · SNPs

C.-H. Wang · B.-J. Liu
Department of Computer Science and Engineering,
Yuan Ze University,
Taoyuan, Taiwan

L. S.-H. Wu (✉)
Institute of Medical Sciences, Tzu Chi University,
No.701, Zhongyang Rd., Sec 3,
Hualien 97004, Taiwan
e-mail: lshwu@hotmail.com

Introduction

The rapid process of diagnosing the patients is an important application of automated information systems, aiming at assisting doctors in making clinical diagnostic decisions. Computerized diagnostic systems have been applied to a variety of medical data, such as ultrasound scanner for images, electroencephalograms (EEGs), and computed tomography images [1–6].

The methods of diagnosing and monitoring the disease depend on detecting the appeal of particular features by doctor's observation. Several techniques for automated diagnostic systems have been developed in the last decade to for a large number of patients who need to be continuously monitored. Such techniques used by transforming the mostly qualitative diagnostic criteria into a more quantitative feature association and classification issues.

In medical decision support technology field, the main concept of it is an inductive inference engine which can learn the decision features of the disease and can be used to diagnose future patients with uncertain disease states within a short period of time.

There are many neural networks models including artificial neural networks (ANN), probabilistic neural networks (PNNs), combined neural networks (CNNs), recurrent neural networks, and support vector machines (SVM) that are being used in medical support systems to assist doctors in disease diagnosis [7].

Artificial neural networks (ANNs) are computational tools for pattern classification, including the diagnosis of diseases. They have been the subject of renewed research interest nearing the past 2 decades [2, 8–10]. Artificial neural networks

(ANNs) have also been used to identify models of complex systems because of their high computational rates, robustness and ability to learn. However, fuzzy set theory plays an important role in dealing with uncertainty when making decisions in medical applications. Therefore, fuzzy sets have attracted growing attention and interest in modern information technology, production techniques, decision making, pattern recognition, diagnostics, data analysis etc. [11–15].

Neuro-fuzzy systems are fuzzy systems which use ANN theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples [16, 17]. A specific approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS) [18], which has shown significant results in modeling nonlinear functions.

In ANFIS, the membership function parameters are extracted from a data set that describes the system behavior. The ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion. Successful implementations of ANFIS in biomedical engineering have been reported, for classification [19, 20] and data analysis [21].

The ANFIS presented in this study was trained with the back propagation gradient descent method in combination with the least squares method, and the data set was obtained from 425 asthmatic subjects and 483 non-asthma subjects from the Taiwanese population. After that training, the classification analysis will be used to classify serum IgE response groups of asthma susceptibility genes from the resulted SNPs. The correction rates and accuracies of the ANFIS and classified analysis integrated model were examined with original clinical patient's records and the performance was reported.

Asthma disease symptom overview

Chronic asthma is a multifaceted disease distressing nearly three hundred million people all over the world [22]. The growing occurrence of asthma and atopic disorders over the past 10 years certifies to the consequence of environmental and lifestyle factors in disease risk assessment [23, 24]. Stable associations of genetic constituents with asthma were supported by family and similar studies [25], and many genes were identified or suspected to be concerned in pathogenesis of asthma [26].

Through the literature that asthma is caused by interaction of multiple genes, some have a protecting effect and others contributing to the disease pathogenesis [27–29]. Consequently, the multifaceted character of asthma phenotypes with substantial locus heterogeneity and environmental influences has made it hard to reveal the genetic factors that trigger asthma [30].

From the report, that a set of seven genes (ADRB2, CD14, FCER1B, IL13, IL4, IL4RA, TNF) are associated with the asthma phenotype [26]. In this study, we selected 13 well-

known functional single-nucleotide polymorphisms (SNPs) from seven genes to perform association studies using 908 Taiwanese subjects. The selected 13 SNPs were singleton SNPs (not haplotype) that have effective gene function and play a significant role in asthma pathogenesis. All these 13 SNPs in seven genes were chosen for genotyping.

Materials and methods

Subjects distribution and assessment

Our study population consists of asthmatic children ranging from age 3 to 12 years old. The study protocol was approved by the Ethical and Clinical Trial Committee of National Cheng-Kung University Hospital. Informed consent forms were required for all participants or their guardians gave consent after having answered a modified British Medical Society respiratory questionnaire, which is identical to the European Community Respiratory Health Survey. These surveys have similar validity as The International Study of Asthma and Allergies in Childhood significant to the diagnosis and assessment of asthma [31, 32].

Pulmonary function was tested using standard methods including spirometry before and after the administration of two puffs of inhaled salbutamol (200 µg/puff). The definition of asthma must meet the following three criteria: (1) history of wheezing and experiencing shortness of breath during or without concurrent respiratory infections, (2) chronic coughing for more than 1 month and being diagnosed by physician of the presence of wheezing episode(s), and (3) bronchodilator test has confirmed the positive response of increased 15% of FEV1.

The diagnosis of childhood asthma below 5 years of age is based mainly on the clinical history, which included more than two episodic attacks of wheezing without concurrent upper respiratory tract infections within 12 months plus a history of atopic disorder as reported by their parents or primary-care physicians. The clinical criterion is consistent with 2006 Global Initiative for Asthma guideline in the diagnosis of childhood asthma under 5 years old (<http://www.ginasthma.com/GuidelineList.asp>) and a good predictor to have bronchial hyper-reactivity as tested by forced oscillation method as recommended by joint American Thoracic Society/European Respiratory Society statement [33].

Other assessment included skin prick test for responsiveness to six common aeroallergens and levels of total serum IgE, as well as IgE specific to house dust and mixed pollens using Unicap system (Pharmacia Diagnostic, Uppsala, Sweden). A positive skin test was defined as the presence of ≥ 1 reaction with a wheal diameter ≥ 5 mm. Total serum IgE was measured by solid-phase immunoassay (Pharmacia IgE EIA; Pharmacia Diagnostics, Uppsala, Sweden).

DNA preparation

Genomic DNA was extracted from blood samples of the study subjects using QIAamp DNA Blood kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. The extracted genomic DNAs were analyzed using garose gel electrophoresis, quantified by spectrophotometer, and stored at -80°C until use.

SNP genotyping

All SNP genotyping were performed using the TaqmanR SNP genotyping assay (ABI: Applied Biosystem Inc. Foster city, CA, USA). The primers and probes of SNPs were from ABI assay on demand (AOD) kit. Reactions were carried out according to the manufacturer's protocol. The probe fluorescence signal detection was performed using the ABI Prism 7900 Real-Time PCR System.

Adaptive neuro-fuzzy inference system (ANFIS)

The methodologies of fuzzy logic and ANNs are described in recent affirmative publications [34]. A Fuzzy Inference System (FIS), also known as a fuzzy-rule based system, is an approach for building qualitative models of human knowledge and reckoning abilities without having to use exact quantitative calculations.

The adaptive neuro-fuzzy inference system (ANFIS) is the achievement of a Fuzzy Inference System in the form of a multi-layer neural network. The fields of fuzzy logic and neural networks have been suggested to give a more powerful tool for dealing with nonlinear and ill-defined problems and to produce better generalization capabilities.

ANFIS architecture

A fuzzy inference system are nonlinear systems that employ fuzzy if-then rules, which can model the qualitative aspects of human knowledge and reasoning processes without employing precise quantitative analyses. Fuzzy logic modeling techniques can be classified into three categories, namely the linguistic [35], the relational equation, and the Takagi-Sugeno-Kang (TSK) [36].

In linguistic models, both the antecedent and the consequence are fuzzy sets while in the TSK model the antecedent consists of fuzzy sets but the consequence is made up of linear equations. Fuzzy relational equation models aim at building the fuzzy relation matrices according to the input-output process data.

The ANFIS study presented in the paper is based on the Sugeno fuzzy model, which was proposed in an effort to develop a systematic approach to generating fuzzy rules and membership function parameters for fuzzy sets from a given

input-output data set. Figure 1 shows the architecture of the ANFIS model. The ANFIS architecture has two sets of trainable parameters: the antecedent parameters, and the consequent parameters. The ANFIS architecture uses a gradient descent algorithm to optimize the antecedent parameters and a least squares algorithm to solve for the consequent parameters. Each rule in the ANFIS model is formed as:

IF x_1 is $A_{1,j}$ AND x_2 is $A_{2,j}$ AND ... AND x_n is $A_{n,j}$
THEN $y = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n$

where $A_{i,j}$ is the j th linguistic term of the i th input variable x_i , and n is the number of inputs. y is the output variable, c_i are consequent parameters to be determined in the training process.

The ANFIS architecture contains six layers.

- Layer 0 is the input layer. It has n nodes where n is the number of input features to the system (13 in the current study).
- Layer 1 represents the membership functions (MFs). A membership function $\mu_{Ai}(x)$ can be any continuous and piecewise differentiable function that transforms the input value x into a membership degree, that is to say a value between 0 and 1. In the current study the Gaussian function (Eq. 1), bell-shaped function (Eq. 2) and triangular-shaped function (Eq. 3) are used in the ANFIS model which are defined as:

$$\mu_{Ai}(x) = \exp\left(\frac{-(x - c_i)^2}{2\sigma_i^2}\right) \quad (1)$$

Where c_i and σ_i are centers and widths of the function and are referred to as the antecedent parameters for the membership function.

$$\mu_{Ai}(x) = \frac{1}{1 + \left[\left(\frac{(x - c_i)}{a_i}\right)^2\right]^{b_i}} \quad (2)$$

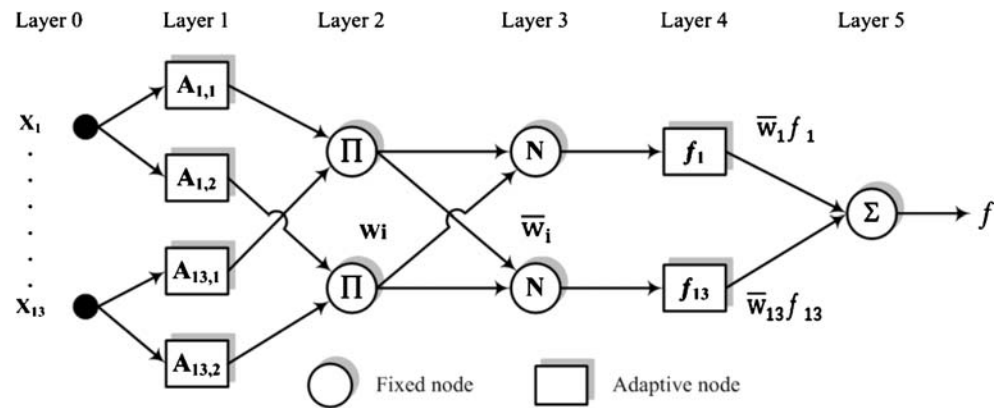
Where a_i , b_i , c_i are antecedent parameters.

$$\mu_{Ai}(x) = \max\left(\min\left(\frac{x - a_i}{b_i - a_i}, \frac{c_i - x}{c_i - b_i}\right), 0\right) \quad (3)$$

Where a_i , b_i , c_i are antecedent parameters, a_i , c_i locate the "feet" of the triangle and b_i locates the peak.

The graphics of these types of membership functions are displayed in Fig. 2.

These parameters are adaptive and they determine the value of the i th membership function for each variable to the fuzzy set A_i . These functions have values ranging from 0 to 1. As the values of the parameters change, the value of the function varies accordingly, therefore indicating the

Fig. 1 ANFIS architecture

degree to which the input variable x satisfies the membership function. In this layer there are $n \times k$ nodes where n is the number of the input variables and k is the number of membership functions.

- Layer 2 provides the strength of each rule by means of multiplication as:

$$W_i = \prod \mu_{A_i}(x_j), i = 1, 2; j = 1, 2, \dots, n.$$

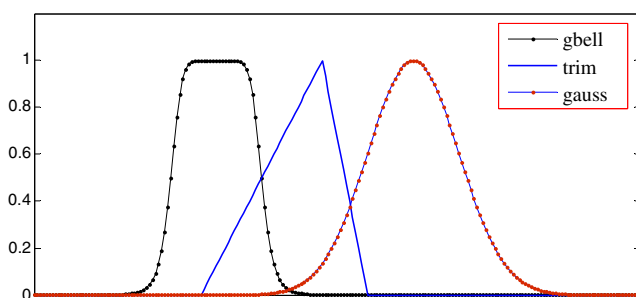
Every node in this layer is the multiplication of the input values from the previous layer. The resulting value represents the firing strength of the i th rule where the variable x_j has linguistic value A_i .

- Layer 3 is the normalization layer where the rule strength is normalized as:

$$\bar{w}_i = \frac{w_i}{\sum w_i}$$

where w_i is the firing strength of the i th rule. The number of nodes in this layer is the same as in the last layer and this layer computes each rule's firing strength to the sum of all rules' firing strengths.

- Layer 4 is an adaptive layer. Every node in this layer is a linear function and the coefficients of the function are adapted through a combination of least squares approximation and back-propagation.

**Fig. 2** Graphics of membership function

$$\bar{w}_i f_i = \bar{w}_i (c_0 + c_1 x_1 + c_2 x_2 \dots + c_n x_n)$$

- Layer 5 is the output layer whose function is the summation of the net outputs of the nodes in Layer 4. The output is computed as:

$$\sum_i \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}$$

Where $\bar{w}_i f_i$ is the output of the node i in layer 4. The overall output is linear, even though the premise parameters are nonlinear.

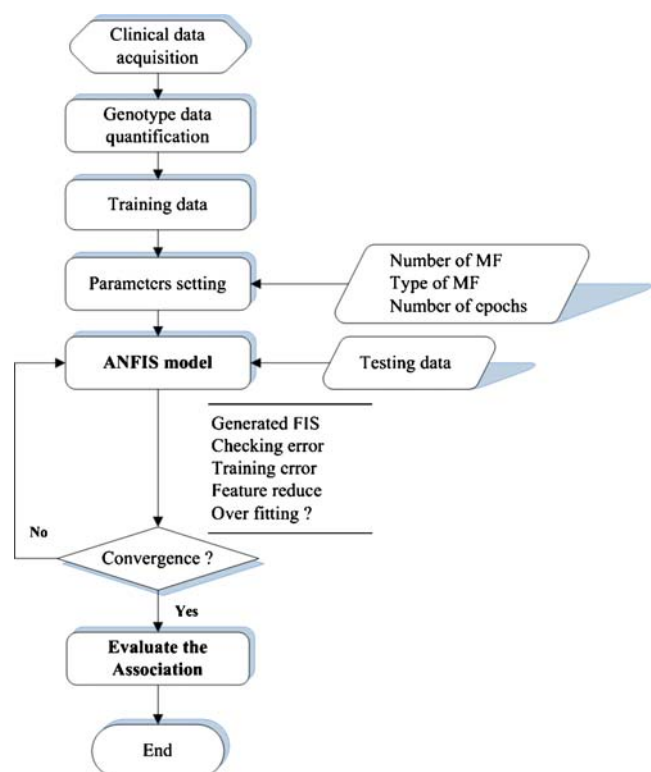
**Fig. 3** Proposed ANFIS approach

Table 1 The measurement for groups

Ige group/number (%)	Clinical subgroup/number (%)	ln(IgE)			
		Min.	Max.	Mean	STD
High/305 (33.6)	AA/204(66.9)	6.2285	10.0962	7.2429	0.6428
	AN /04 (1.3)	6.2529	6.8773	6.4660	0.28
	NA /92 (30.2)	6.2160	8.4935	7.0058	0.5157
	NN /05 (1.6)	6.3262	7.1554	6.6346	0.3401
	Total	6.2160	10.0962	7.1512	0.6168
Median/173 (19.0)	AA /73 (42.2)	5.3127	6.2126	5.8272	0.2615
	AN /08 (4.6)	5.3181	6.0497	5.8198	0.2391
	NA /83 (48)	5.3230	6.2124	5.7676	0.2718
	NN /09 (5.2)	5.3799	6.1420	5.6305	0.2661
	Total	5.3127	6.2126	5.7881	0.2677
Low/430 (47.4)	AA /52 (12.1)	0.4055	5.2953	4.6311	0.7950
	AN/84 (19.5)	1.0986	5.2454	3.8024	0.9837
	NA /74 (17.2)	3.6569	5.2933	4.6513	0.4863
	NN/220(51.2)	0.3365	5.2258	3.3749	1.0901
	Total	0.3365	5.2953	3.83	1.1

$N=908$ observations, Mean=
5.3186, STD=1.7133

AA both asthma and allergy, AN
asthma only, NA allergy only,
NN normal

Performance measurement

The major concerns of the validation are the evaluation of the models, predictability and validity of the models, and consistency of the results. The measurement in our work is Root Mean-Square Error (RMSE) which is a degree of distribution of the data.

The RMSE can be calculated by the below equation where N denotes the number of data patterns, y_i^p is the predicted output for pattern number i , and y_i^e is the experimental output data (Eq. 4).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^p - y_i^e)^2} \quad (4)$$

Proposed integrating approach

In this work, the proposed approach consists of four procedures as in Fig. 3; clinical data acquisition, genotype data quantification, ANFIS model, and validate the accuracy of the ANFIS model by classification which is specifically explained in the next section. The summary role of each procedure is described as follows:

Clinical data acquisition: this procedure is used to attain DNAs from blood samples. Furthermore data processing is also carried out.

Genotype data quantification: based on Hardy–Weinberg principle, transform genotypes data to frequency data.

Table 2 The genotype frequencies for 13 SNPs within seven genes

Codes	SNPs	Genes	Genotype (Frequency)
SNP1	rs1800925	IL13 -1112C/T	TT(0.030), CT(0.260), CC(0.710)
SNP2	rs20541	IL13 Gln130Arg	AA(0.111), AG(0.446), GG(0.443)
SNP3	rs2243250	IL4 -589C/T	TT(0.614), CT(0.344), CC(0.042)
SNP4	rs2070874	IL4 +33C/T	TT(0.612), CT(0.346), CC(0.042)
SNP5	rs1800629	TNF -308G/A	AA(0.011), AG(0.199), GG(0.790)
SNP6	rs569108	MS4A2 Glu237Gly	AA(0.654), AG(0.303), GG(0.043)
SNP7	rs1805010	IL4Ra Ile50Val	AA(0.239), AG(0.471), GG(0.290)
SNP8	rs1801275	IL4Ra Gln551Arg	AA(0.718), AG(0.261), GG(0.021)
SNP9	rs1805011	IL4Ra Glu375Ala	AA(0.859), AC(0.137), CC(0.004)
SNP10	rs1805015	IL4Ra Ser478Pro	TT(0.861), CT(0.133), CC(0.006)
SNP11	rs2569190	CD14 -159C/T	AA(0.354), AG(0.489), GG(0.157)
SNP12	rs1042713	ADRB2 Arg16Gly	AA(0.326), AG(0.485), GG(0.189)
SNP13	rs1042714	ADRB2 Gln27Glu	GG(0.015), CG(0.163), CC(0.822)

Table 3 Simulation results

MF	MF No.	Epoch (RMSE)	CPU(Sec.)	Type of selection	RMSE values			
					Train	Test	Dif.	
a								
gbell	2	11 (1.327)	0.778	SNP06	1.3137	1.3038	0.0099	
			0.930	SNP05 SNP06	1.2996	1.3266	0.0270	
				SNP04 SNP05	1.3004	1.3188	0.0183	
				SNP06 SNP07	1.3051	1.3038	0.0013	
				SNP06 SNP11	1.3064	1.3162	0.0098	
				SNP06 SNP09	1.3069	1.3180	0.0112	
	3	3 (1.327)	5.871	SNP04 SNP05 SNP06	1.2684	1.3613	0.0930	
			0.228	SNP06	1.3137	1.3038	0.0099	
			1.536	SNP05 SNP06	1.2996	1.3266	0.0270	
				SNP04 SNP05	1.3004	1.3188	0.0183	
				SNP06 SNP07	1.3051	1.3036	0.0014	
				SNP06 SNP11	1.3062	1.3176	0.0114	
	4	27 (1.327)		SNP06 SNP09	1.3069	1.3180	0.0112	
			24.788	SNP04 SNP05 SNP06	1.2684	1.4000	0.1316	
			0.234	SNP06	1.3137	1.3038	0.0099	
			2.433	SNP05 SNP06	1.2996	1.3266	0.0270	
				SNP04 SNP05	1.3004	1.3188	0.0183	
				SNP06 SNP07	1.3051	1.3036	0.0014	
				SNP06 SNP11	1.3062	1.3176	0.0114	
				SNP06 SNP09	1.3069	1.3180	0.0112	
112.065			SNP04 SNP05 SNP06	1.2684	1.4346	0.1663		
b								
gauss			2	10 (1.327)	0.416	SNP06	1.3137	1.3038
	0.925	SNP05 SNP06			1.2996	1.3266	0.0270	
		SNP04 SNP05			1.3004	1.3188	0.0183	
		SNP06 SNP07			1.3051	1.3037	0.0014	
		SNP06 SNP11			1.3065	1.3157	0.0092	
		SNP06 SNP09			1.3069	1.3180	0.0112	
	3	12 (1.327)	5.825	SNP04 SNP05 SNP06	1.2684	1.3620	0.0937	
			0.220	SNP06	1.3137	1.3038	0.0099	
			1.536	SNP05 SNP06	1.2996	1.3266	0.0270	
				SNP04 SNP05	1.3004	1.3188	0.0183	
				SNP06 SNP07	1.3051	1.3036	0.0014	
				SNP06 SNP11	1.3062	1.3176	0.0114	
	4	23 (1.327)		SNP06 SNP09	1.3069	1.3180	0.0112	
			24.524	SNP04 SNP05 SNP06	1.2684	1.3988	0.1305	
			0.228	SNP06	1.3137	1.3038	0.0099	
			2.398	SNP05 SNP06	1.2996	1.3266	0.0270	
				SNP04 SNP05	1.3004	1.3188	0.0183	
				SNP06 SNP07	1.3051	1.3036	0.0014	
				SNP06 SNP11	1.3062	1.3176	0.0114	
				SNP06 SNP09	1.3069	1.3180	0.0112	
109.732			SNP04 SNP05 SNP06	1.2684	1.4323	0.1640		
c								
trim			2	1 (1.327)	0.408	SNP06	1.3137	1.3038
	0.922	SNP05 SNP06			1.2996	1.3266	0.0270	

Table 3 (continued)

MF	MF No.	Epoch (RMSE)	CPU(Sec.)	Type of selection	RMSE values		
					Train	Test	Dif.
3	2 (1.327)	5.909	0.213	SNP04 SNP05	1.3004	1.3188	0.0183
				SNP06 SNP07	1.3051	1.3033	0.0017
				SNP06 SNP11	1.3066	1.3152	0.0086
				SNP06 SNP09	1.3069	1.3180	0.0112
				SNP04 SNP05 SNP06	1.2684	1.3610	0.0927
				SNP06	1.3137	1.3038	0.0099
				SNP05 SNP06	1.2996	1.3266	0.0270
				SNP04 SNP05	1.3004	1.3188	0.0183
	10 (1,327)	24.054	0.231	SNP06 SNP07	1.3051	1.3036	0.0014
				SNP06 SNP11	1.3062	1.3176	0.0114
				SNP06 SNP09	1.3069	1.3180	0.0112
				SNP04 SNP05 SNP06	1.2684	1.4109	0.1425
				SNP06	1.3137	1.3038	0.0099
				SNP02 SNP12	1.2992	1.3166	0.0174
				SNP02 SNP13	1.2980	1.3062	0.0082
				SNP05 SNP06	1.2996	1.3266	0.0270
4	110.857	1.487	110.857	SNP04 SNP05	1.3004	1.3188	0.0183
				SNP06 SNP07	1.3051	1.3036	0.0014
				SNP06 SNP11	1.3062	1.3176	0.0114
				SNP06 SNP09	1.3069	1.3180	0.0112
				SNP04 SNP05 SNP06	1.2684	1.4417	0.1733

initial training epoch number: 1, training error goal: 0

initial step size: 0.1, step size decrease rate: 0.5, step size increase rate: 1.5

|Dif. |: ABS(train–test)

MF Member function

ANFIS model: from parameters setting, training and testing data, the model is built, and the most significant SNPs are calculated to reduce the SNPs selection from candidate genes.

Validation: coordinating on SNPs selection from ANFIS model, it will be used to validate the accuracy for forecasting the association on 3 serum IgE groups.

asthma, Immunoglobulin E (IgE) measurement is listed in Table 1.

Genotype data quantification

Based on the idea of the Hardy–Weinberg equilibrium, both allele and genotype frequencies in a population remain constant; however, in this study the genotype data was substituted by its frequency data correspond, which are calculated and summarized in Table 2.

Results and discussion

Clinical data acquisition

DNAs were extracted from blood samples of 425 asthmatic children and 483 non-asthma children. The collection of well-distributed and accurately input data is the main requirement of acquisition, for precisely model measurement, the original data set are divided into three groups based on ln(IgE) and four clinical subgroup for each. The computation of each group between asthma and non-

ANFIS model

The proposed technique involved training the ANFIS model to detect the genes association which most significantly influences the serum IgE levels and the development of asthma when the 13 SNPs given in Table 2 were used as inputs. There are a total of 908 records in this study, of which 425 asthmatic records and 483 non-asthma records. The parameters use in ANFIS model included, three types of membership function each with three different numbers

of function, the number of training epochs is fixed to 400 on the simulation. Training data sets and testing data sets are partitioned from 908 records according to even-indexed and odd-indexed respectively. Through simulation, the number of SNPs selection are limited by RMSE difference value, which were less than 0.03; that means the training and testing are comparable up to 97%. The ANFIS model simulation results are listed in Table 3. The association analysis from simulation results are listed in Table 4.

Discussion

At the end of 377 training periods, the network error (mean square error—MSE) of three types of membership function with three different number of member function are all converged to the same convergence value 1.327. In Table 3 the training epoch numbers of three membership functions under the same type of SNPs selection results, triangular-shaped function with two numbers of functions are more appropriate than the other two membership functions in the ANFIS model simulation of this study, three examples of network error are illustrated in Fig. 4.

After simulation, the types of SNPs selection were used to validate the accuracy of the ANFIS model for forecasting the association of 13 SNPs within seven asthma susceptibility genes on 3 serum IgE groups. Classification results of the classifiers were calculated by a confusion matrix. A confusion matrix contains information about actual and predicted classifications done by a classification system.

The performance of such systems are commonly evaluated using the data in the matrix, and the testing of the classifiers can be decided by the computation of

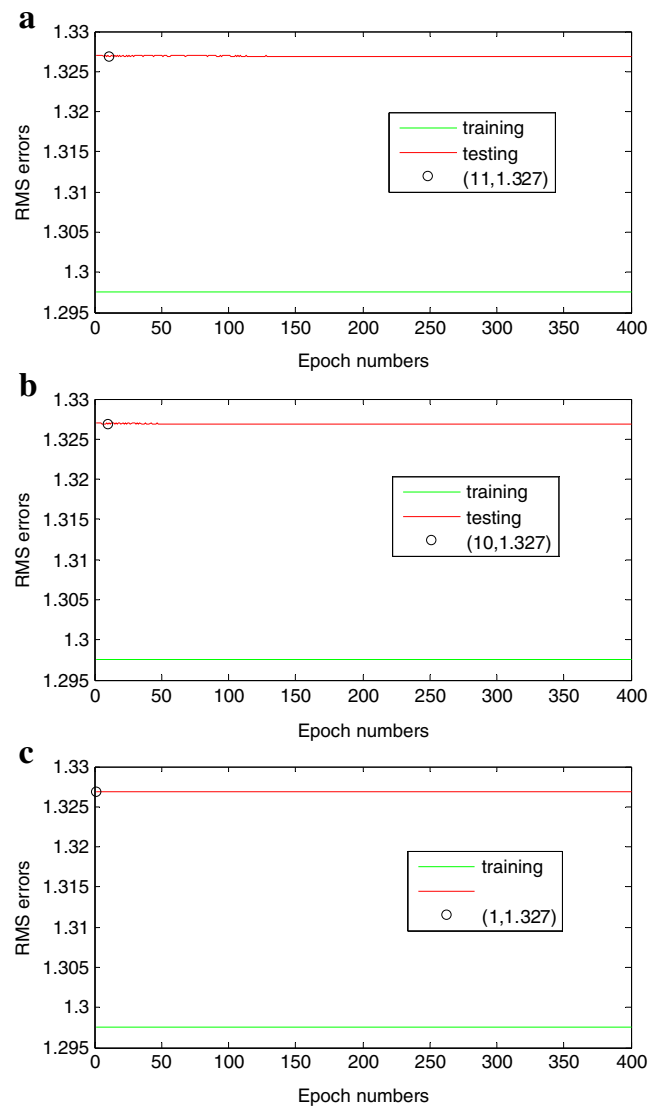


Fig. 4 The curve of network error for gbell, gauss and trim membership function with two numbers of function correspond to **a**, **b** and **c**

Table 4 Confusion matrices of IgE group Low and High for gene pair (a) and triple genes (b)

Actual result	Predicted result		
	High	Low	Total
a			
Gene pair MS4A2 Glu237Gly and IL4Ra Glu375Ala			
High	295	10	305
Low	34	396	430
Total	329	406	735
b			
Triple genes IL4 +33C/T, TNF -308G/A and MS4A2 Glu237Gly			
High	292	13	305
Low	29	401	430
Total	321	414	735

specificity, sensitivity and total classification accuracy. These three statistical parameters are defined as below:

- **Specificity**—the number of true negative decisions divided by the number of actually negative cases
- **Sensitivity**—the number of true positive decisions divided by the number of actually positive cases
- **Total classification accuracy**—the number of correct decisions divided by total number of cases.

The confusion matrix of the appropriated genes combinations showing the classification results of the ANFIS model used for association forecasting is given in Table 4, and the statistical parameters of all genes combinations are shown in Table 5.

In biology, the total serum IgE levels are more likely to be under genetic control, in present study all our selected

Table 5 Association analysis—three groups on serum IgE with accuracy of statistical parameters

Genes combinations			Statistical parameters (%)		
Gene1	Gene2	Gene3	Specificity	Sensitivity	Total classification accuracy
a					
IgE group Low and High					
MS4A2 Glu237Gly	Null	Null	67.2	78.6	74.0
TNF -308G/A	MS4A2 Glu237Gly	Null	90.2	88.4	89.1
IL4 +33C/T	TNF -308G/A	Null	91.0	89.0	90.0
MS4A2 Glu237Gly	IL4Ra Ile50Val	Null	85.0	88.2	87.0
MS4A2 Glu237Gly	CD14 -159C/T	Null	90.5	90.5	90.5
MS4A2 Glu237Gly	IL4Ra Glu375Ala	Null	96.7	92.1	94.0
IL4 +33C/T	TNF -308G/A	MS4A2 Glu237Gly	96.0	93.3	94.3
b					
IgE group Low and Median					
MS4A2 Glu237Gly	Null	Null	66.0	62.0	65.0
TNF -308G/A	MS4A2 Glu237Gly	Null	73.0	94.2	79.0
IL4 +33C/T	TNF -308G/A	Null	75.1	92.5	80.1
MS4A2 Glu237Gly	IL4Ra Ile50Val	Null	68.9	77.5	71.3
MS4A2 Glu237Gly	CD14 -159C/T	Null	72.0	82.1	75.0
MS4A2 Glu237Gly	IL4Ra Glu375Ala	Null	74.5	93.6	80.0
IL4 +33C/T	TNF -308G/A	MS4A2 Glu237Gly	74.5	92.0	79.5
c					
IgE group Median and High					
MS4A2 Glu237Gly	Null	Null	67.2	55.0	63.0
TNF -308G/A	MS4A2 Glu237Gly	Null	70.5	77.0	73.0
IL4 +33C/T	TNF -308G/A	Null	68.2	86.7	75.0
MS4A2 Glu237Gly	IL4Ra Ile50Val	Null	66.0	74.0	69.0
MS4A2 Glu237Gly	CD14 -159C/T	Null	80.0	74.6	78.0
MS4A2 Glu237Gly	IL4Ra Glu375Ala	Null	75.1	63.6	71.0
IL4 +33C/T	TNF -308G/A	MS4A2 Glu237Gly	76.7	82.1	78.7

SNPs have previously been reported to be highly associated with asthma. Besides, from the association analysis in Table 5, the study discovered that variable effects of single SNPs observed by the combined effects of the 2 or 3 polymorphisms which tested in this study can exceed the single effects by far.

Even though the population tested in this study were relatively small for a study of multiple genes interactions, the effects were relative sufficient to reach statistical significance. Among the results, this paper also discovered that MS4A2 Glu237Gly and IL4Ra Glu375Ala were identified as having significance in group Low and High, and in group Median and High was MS4A2 Glu237Gly and CD14 -159C/T. The similar validated results of multiple asthma susceptibility genes were also found in Chinese populations [37] and Japanese populations [38]. In more than two genes co-occurring association cases, the ANFIS

model also discovered that IL4 +33C/T, TNF -308G/A and MS4A2 Glu237Gly triple genes association have the comparable classification accuracy of the MS4A2 Glu237Gly and IL4Ra Glu375Ala.

Conclusions

From the modeling investigation conducted, it is clear that the integration of ANFIS and classification analysis stand an advantage of combining rule base and training ability, fuzzy logic and neural network, all these characters plays an important role in dealing with uncertainty when making decisions in medical applications.

The present ANFIS model combined neural network adaptive capabilities and a fuzzy logic qualitative approach. The classification results and statistical measures were used

for evaluating the ANFIS model. The forecast of the association by serum IgE groups and asthma susceptibility genes were done with a total classification accuracy of 94%.

Therefore, in this study we believe that the advantage of the ANFIS, namely estimation speed, simplicity, error free and capacity to learn from examples can be used to forecast association from asthma susceptibility genes.

Future research may include a more complete set of input features (IL4 (SNPs; G2979T, G3551C, VE6523, G7235T), IL4Ra(SNPs; A148G, T1432C, A1652G)), and applications which automatically assign a different number of MFs at each input; determine specialized function, or substitute the triangular-shaped function by other approaches.

Acknowledgements The authors are grateful to the editors and anonymous referees for their helpful comments on this paper, and also thankful to Miss Rebecca Cheng for clerical assistance.

References

- Mobley, B. A., Schechter, E., Moore, W. E., McKee, P. A., and Eichner, J. E., Predictions of coronary artery stenosis by artificial neural network. *Artif. Intell. Med.* 18:187–203, 2000.
- Miller, A. S., Blott, B. H., and Hames, T. K., Review of neural network applications in medical imaging and signal processing. *Med. Biol. Eng. Comput.* 30:449–464, 1992.
- Übeyli, E. D., Usage of eigenvector methods to improve reliable classifier for Doppler ultrasound signals. *Comput. Biol. Med.* 38:5563–5573, 2008.
- Übeyli, E. D., Comparison of different classification algorithms in clinical decision-making. *Expert Syst.* 24:117–131, 2007.
- Übeyli, E. D., Detection of electrocardiogram beats using a fuzzy similarity index. *Expert Syst.* 24:287–296, 2007.
- Übeyli, E. D., Combining neural network models for automated diagnostic systems. *J. Med. Syst.* 30:6483–6488, 2006.
- Übeyli, E. D., Time-varying biomedical signals analysis with multiclass support vector machines employing Lyapunov exponents. *Digit. Signal Process.* 18:4646–4656, 2008.
- Baxt, W. G., Use of an artificial neural network for data analysis in clinical decision making: The diagnosis of acute coronary occlusion. *Neural Comput.* 2:480–489, 1990.
- Ergün, U., Serhatlioglu, S., Hardalac, F., and Güler, I., Classification of carotid artery stenosis of the patients with diabetes by neural network and logistic regression. *Comput. Biol. Med.* 34:389–405, 2004.
- Güler, I., and Übeyli, E. D., Detection of ophthalmic artery stenosis by least-mean squares back propagation neural network. *Comput. Biol. Med.* 33:333–343, 2003.
- Dubois, D., and Prade, H., An introduction to fuzzy systems. *Clin. Chim. Acta* 270:3–29, 1998.
- Kuncheva, L. I., and Steimann, F., Fuzzy diagnosis. *Art. Intell. Med.* 16:121–128, 1999.
- Nauck, D., and Kruse, R., Obtaining interpretable fuzzy classification rules from medical data. *Art. Intell. Med.* 16:149–169, 1999.
- Güler, I., Hardalac, F., Ergün, U., and Baris-C, N., Application of FFT analyzed cardiac Doppler signals to fuzzy algorithm. *Comput. Biol. Med.* 32:435–444, 2002.
- Güler, I., Hardalac, F., Ergün, U., and Baris-C, N., Classification of aorta Doppler signals using variable coded hierarchical genetic fuzzy system. *Expert Syst. Appl.* 26:321–333, 2004.
- Andrews, R., Diederich, J., and Tickle, A., A survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowl. Base Syst.* 8:373–389, 1995.
- Mcgarra, K., Wermter, S., and Macintyre, J., The extraction and comparison of knowledge from local function networks. *Int. J. Comput. Intell. Appl.* 1:369–382, 2001.
- Jang, J. S. R., ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* 23:3665–3685, 1993.
- Usher, J., Campbell, D., Vohra, J., and Cameron, J., A fuzzy logic controlled classifier for use in implantable cardioverter defibrillators. *Pacing Clin. Electrophysiol.* 22:183–186, 1999.
- Belal, S. Y., Taktak, A. F. G., Nevill, A. J., Spencer, S. A., Roden, D., and Bevan, S., Automatic detection of distorted plethysmogram pulses in neonates and pediatric patients using an adaptive network-based fuzzy inference system. *Artif. Intell. Med.* 24:149–165, 2002.
- Virant-Klun, I., and Virant, J., Fuzzy logic alternative for analysis in the biomedical sciences. *Comput. Biomed. Res.* 32:305–321, 1999.
- Masoli, M., Fabian, D., Holt, S., and Beasley, R., The global burden of asthma: Executive summary of GINA Dissemination Committee Report. *Allergy* 59:469–478, 2004.
- Beasley, R., The burden of asthma with specific reference to the United States. *J. Allergy Clin. Immunol.* 109:S482–S489, 2002.
- Bach, J. F., The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* 347:911–920, 2002.
- Skadhauge, L. R., Christensen, K., Kyvik, K. O., and Sigsgaard, T., Genetic and environmental influence on asthma: A population-based study of 11688 Danish twin pairs. *Eur. Respir. J.* 13:8–14, 1999.
- Ober, C., and Hoffjan, S., Asthma genetics 2006: The long and winding road to gene discovery. *Genes Immun.* 7:95–100, 2006.
- Anderson, G. G., and Cookson, W. O., Recent advanced in the genetics of allergy and asthma. *Mol. Med. Today* 5:264–273, 1999.
- Blumenthal, M. M., What we know about the genetics of asthma at the beginning of the 21st century. *Clin. Rev. Allergy Immunol.* 22:11–31, 2002.
- Holloway, J. W., Beghe, B., and Holgate, S. T., The genetic basis of atopic asthma. *Clin. Exp. Allergy* 29:1023–1032, 1999.
- Huang, J. L., Asthma severity and genetics in Taiwan. *J. Microbiol. Immunol. Infect.* 38:158–163, 2005.
- Burney, P. G., Luczynska, C., Chinn, S., and Jarvis, D., The European Community Respiratory Health Survey. *Eur. Respir. J.* 5:954–960, 1994.
- Pearce, N., Sunyer, J., Cheng, S., Chinn, S., Bjorksten, B., Burr, M., Keil, U., Anderson, H. R., and Burney, P., Comparison of asthma prevalence in the ISAAC and the ECRHS. ISAAC Steering Committee and the European Community Respiratory Health Survey. International Study of Asthma and Allergies in Childhood. *Eur. Respir. J.* 16:420–426, 2000.
- Beydon, N., Davis, S. D., Lombardi, E., Allen, J. L., Arets, H. G. M., Aurora, P., Bisgaard, H., Davis, G. M., Ducharme, F. M., Eigen, H., Gappa, M., Gaultier, C., Gustafsson, P. M., Hall, G. L., Hantos, Z., Healy, M. J. R., Jones, M. H., Klug, B., Carlsen, K. C. L., McKenzie, S. A., Marchal, F., Mayer, O. H., Merkus, P. J. F. M., Morris, M. G., Oostveen, E., Pillow, J. J., Seddon, P. C., Silverman, M., Sly, P. D., Stocks, J., Tepper, R. S., Vilozi, D., and Wilson, N. H., Behalf of the American Thoracic Society/ European Respiratory Society Working Group on Infant and Young Children Pulmonary Function Testing. An official Amer-

- ican Thoracic Society/European Respiratory Society Statement: Pulmonary function testing in preschool children. *Am. J. Respir. Crit. Care Med.* 175:1304–1345, 2007.
34. Terano, T., Asai, K., and Sugeno, M., *Fuzzy systems theory and its applications*. Academic, San Diego, 1992.
 35. Mamdani, E. H., and Assilian, S., An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* 7:1–8, 1975.
 36. Sugeno, M., and Kang, G. T., Structure identification of fuzzy model. *Fuzzy Sets Syst.* 28:15–33, 1988.
 37. Leung, T. F., Tang, N. L. S., Chan, H. I. S., Li, A. M., Ha, G., Lam, C. W. K., and Fok, T. F., Distribution in allele frequencies of predisposition-to atopy genotypes in Chinese children. *Pediatr. Pulmonol.* 34:419–424, 2002.
 38. Unoki, M., Furuta, S., Onouchi, Y., Watanabe, O., Doi, S., Fujiwara, H., Miyatake, A., Fujita, K., Tamari, M., and Nakamura, Y., Association studies of 33 single nucleotide polymorphisms (SNPs) in 29 candidate genes for bronchial asthma: Positive association a T294C polymorphism in the thromboxane A2 receptor gene. *Hum. Genet.* 106:440–446, 2000.