Evolutionary learning-based modeling for warfarin dose prediction in Chinese

Yanyun Tao
School of urban railway transportation
Soochow University
Suzhou, China, 215137

Yuzhen Zhang*
Cardiology department,
The first affiliated Hospital of Soochow University,
Suzhou, China, 215006
zhangyuzhen@suda.edu.cn

Bin Jiang
Cardiology department,
The first affiliated Hospital of Soochow University,
Suzhou, China, 215006

ABSTRACT

The perdition on the warfarin dosage daily requirement is important to thromboembolic treatment with warfarin. In order to improve the generalization and accuracy of warfarin-dose predictive model, an evolutionary learning modeling method, called ELM, is developed. ELM runs multiple artificial neural network (ANNs) and support vector regression (SVR) as leaning members to build candidate models. A genetic algorithm (GA) in the outer loop of ELM optimizes the parameters of multiple ANNs and SVR for exploring better predictive models. The models discovered by leaning members were trained, validated and evaluated on the dataset of 100 Chinese patients provided by The First affiliated Hospital of Soochow University. In the experiment, ELM is compared with ANNs, SVR, linear regression models, and the predictive calculator of IWPC. The results show that the models developed by ELM present the best mean squared error(*MSE*) in these cases. ELM outperformed the other comparable methods 13.9~31.5% in terms of *MSE*. It is also noted that the variation in R^2 of ELM discovered models from training set to test set is even no decrease. This illustrates that the ELM models have better generalization than that of the other methods

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**→ Modeling and simulation→ modeling methodologies; learning paradigms → Supervised learning →Supervised learning by regression[1]

GECCO '17 Companion, July 15-19, 2017, Berlin, Germany

ACM ISBN 978-1-4503-4939-0/17/07…$15.00

KEYWORDS

warfarin dose prediction, genotype, machine learning, predictive model, neural network, support vector regression, genetic algorithm

1 INTRODUCTION

The prevention and treatment of thromboembolic disorders is largely dependent on anticoagulants. Warfarin is the most widely used anticoagulant in this field. IWPC claimed that clinical factors account for 26% of the variability in dose, which is improved to 43% by incorporation of CYP2C9 and VKORC1 genotypes [1]. Due to large inter-individual variability in dose requirements, the warfarin dosage in anticoagulation is usually complicated and unpredictable [2]. This actually increases the difficulty in giving an accurate prediction of warfarin-dose. In anticoagulation therapy, the international normalized ratio (INR) of prothrombintime should be closely monitored to ensure an adequate yet safe warfarin dose is taken[3]. Although some dose predictive models such as IWPC calculator and pharmacogenetics-based dosing model have been developed and used by some clinicians [1, 4], the prediction of warfarin maintenance dose to keep INR within the target therapeutic range is still mostly depending on the experience of clinicians.

Machine learning (ML) algorithm is to acquire the pattern by automatic analysis on the data, and makes use the pattern to predict the unknown data. A variety of modeling methods based on ML techniques are developed, where Bayesian network (BN), artificial neural network (ANN), and support vector machine (SVM) are the most popular techniques. Wright et.al aimed to develop a BN-based dose individualization tool for warfarin, which has been incorporated into the freely available TCIWorks for use in the clinic[5]. However, BN cannot certainly give a high-accurate prediction on dosage. ANNs has been widely used in the modeling for warfarin-dose prediction [6-7]. For ANNs, the poor generalization ability is a common disadvantage. SVR is a structured risk minimization-learning modeling method [8-

http://dx.doi.org/10.1145/3067695.3082492

10]. SVR approximates the function not only by minimizing the empirical risk, but also the incredible risk [10]. Disguised with ANNs, SVR can be working with a relatively small number of training data called support vectors.

Any a kind of ML techniques applied in modeling of function has its own disadvantage. The fact that diverse ensemble members outperform a single such model has been proved, especially in complex system modeling [11-13]. This is because the diversity is able to explore a large search space and makes the probability of finding a generalized solution significantly increased. In this study, diversity can be achieved by two ways. One is to create candidate predictive model by using multiple ML techniques (i.e. ANNs and SVR); the other way is to generate a variety of candidate models by using a ML technique with different parameters, which can be changed and optimized by Evolutionary Algorithms (EAs). EAs are population-based search algorithms, which have been successfully applied in complex optimization problem [15-17]. In the family of EAs, genetic algorithm (GA) is a parallel and global search algorithm, which can evolve a variety of candidates to produce the best adaptive solution. Therefore, GA integrated with different ML techniques can be an effective way to overcome overfitting and provide a generalized predictive model.

2 Backgrounds

2.1 Genetic algorithm

Genetic Algorithm (GA) is a heuristic method based on "survive for fittest". In GA, chromosome typically encodes solution to a binary string. Selection, crossover (P_c) and mutation (P_m) are main genetic operators for search space exploration. Roulette wheel together with elite strategy is widely used selection strategy. Crossover changes a pair of chromosomes by swapping their genes at probability P_c, and mutation performs the variation on a chromosome by change its genes at probability P_m. The flow of GA is as follow.

1. Create initial random population of chromosomes(i.e. potential solution);
2. Evaluate fitness of each chromosome ;
3. Perform selection, crossover and mutation on chromosomes to create a new population;
4. When the termination is achieved or optimal solution is found, GA stop; else, go to the step "Evaluate..".

2.2 Artificial Neural Network

In the family of ANNs, back-propagation (BP) network and radial basis function (RBF) network are two most popularly used algorithms in solving mathematical modeling problems.

BP network is a multi-layer ANN (i.e. normally three layers) with back-propagation learning algorithm. Mathematically, a widely used type of composition is the nonlinear weighted sum, which can be denoted as (1).

$$y(x) = f(\sum_{i=1}^{n} w_i g(x_i) - \theta) \tag{1}$$

Where x_i is a training sample, n is the number of neurons in the hidden layer, θ is a threshold, $g_i(x)$ denotes transfer functions, w_i indicates the weight of function g_i, $f(x)$ is an activation function, which can be defined as a linear or nonlinear function. The weights are modified to minimize the mean squared error between the desired and actual outputs of the network.

A RBF network is typically a three layers (i.e. input, hidden, and output) network uses radial basis functions as activation functions. Let input be a vector of real numbers, $x \in \mathbf{R}^n$. The output is a linear combination of radial basis functions of the input vector y: $\mathbf{R}^n \rightarrow \mathbf{R}$ and neuron parameters, and is given by

$$y(x) = \sum_{i=1}^{n} w_i g(\| x - c_i \|) \tag{2}$$

Where c_i is the center vector for i-th neuron, and w_i is the weight of i-th neuron in the linear output neuron. Radial basis function that depends only on the distance from a center vector is radially symmetric about that vector. The radial basis function is commonly taken to be Gaussian $g(\| x\text{-}c_i \|)=\exp[-\beta \| x\text{-}c_i \|^2]$, where $\lim_{x \rightarrow c_i} g(\| x\text{-}c_i \|)=0$.

2.3 Support vector regression

Support vector regression (SVR) is a version of SVM for regression, which uses linear models to implement nonlinear regression by mapping the input space to a higher dimensional feature space using kernel functions. It inherits all the advantages from SVM. The mathematical function of SVR represents the nonlinear system can be denoted by (3).

$$y(x) = \sum_{i=1}^{l} \lambda \cdot k(\langle x, x_i \rangle) + b \tag{3}$$

Where x_i is a training sample with target value y_i, λ are the coefficients onto the kernel functions, $\langle x, x_i \rangle$ is the inner product plus, $k(*)$ indicates the kernel function(e.g. linear, poly, rbf and so on), and l is the size of sample data.

Disguised from ANNs, SVR simultaneously minimizes the fitting error in the training data and incredible risk (i.e. the model complexity). This can be written as a multi-objective optimization problem, denoted as following.

$$\min \ \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{4}$$

$$\text{Subject to} \begin{cases} \langle w, x_i \rangle + b - y_i < \varepsilon + \xi \\ y_i - \langle w, x_i \rangle - b < \varepsilon + \xi^* \\ \xi, \xi^* > 0 \end{cases}$$

Where C indicates the penalty factor, ξ and ξ^* are slack variables, and ε is a free parameter that serves as a threshold. When C is large, the points that have large deviation from the actual output cannot be tolerated; On the contrary, when C is set to a small value, deviations larger than threshold are tolerated.

3 METHOD

3.1 Predictive model building

ANNs (i.e. BP and RBF) use a neurons interconnection structure to deal with modeling by simulating the human brain, and SVR can yield predictive models with good generalization under small size dataset. Over fitting is a difficult issue to resolve when using ANNs for modeling because of its super function representation. It is known that the number of neurons in layers and basis functions has largely influence on the performance of ANNs. As compared to ANNs, SVR is adaptive to small size of dataset because it can fit continuous complex functions with only support vectors. Using a support vectors is a pragmatic way to enhance the effectiveness of fitting with the limited size dataset. In order to make use of respective advantages of ANNs and SVR, we run multiple ML techniques (i.e. BP, RBF and SVR) simultaneously to build diverse candidate models.

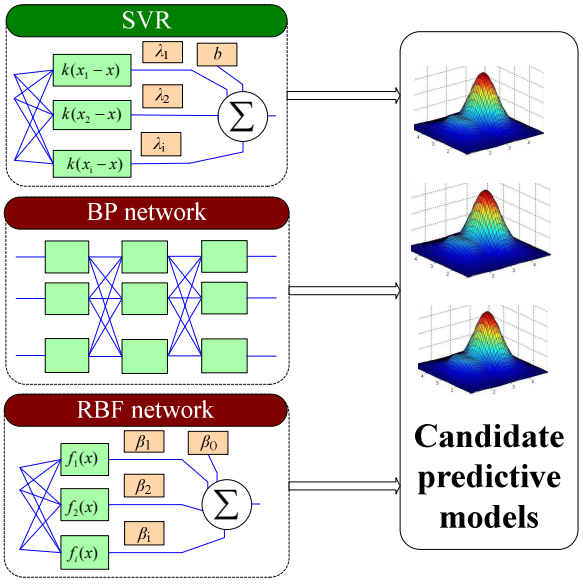

Figure.1 candidate model building with ML algorithms

Fig 1 gives the building candidate models using three basic ML techniques. The candidate models built by three techniques have different structures and parameters. The models created by SVR have the structure denoted as (3). BP models actually have a three-layer, input, hidden and output layer, where the hidden-layer realizes the non-linear complex function map. RBF models are represented by a structure with Gaussian basis functions and coefficients.

The evaluation function for a BP|RBF|SVR built model is based on the mean square error (*MSE*) of model output and the predicted output. The function *z* is denoted as (5).

$$z = \frac{1}{n}\sum_{j=1}^{n}[y(X_j) - Y_j]^2 \qquad (5)$$

Where X_i and Y_i indicate the *i*-th input vector and output value of training dataset, $y(X_i)$ denotes the predicted output of model under the input vector X_i, and *n* denotes the size of training dataset. The goal is to build a model with *z* minimized on the training set.

3.2 Framework for warfarin dose prediction

In this subsection, the evolutionary learning modeling (ELM) is described. The inputs to ELM are a set of clinical variables and genotypes, a set of original data of patients, and the parameters of GA. The output of ELM is the final solution, which is a predictive model discovered by one of BP, RBF and SVR algorithms.

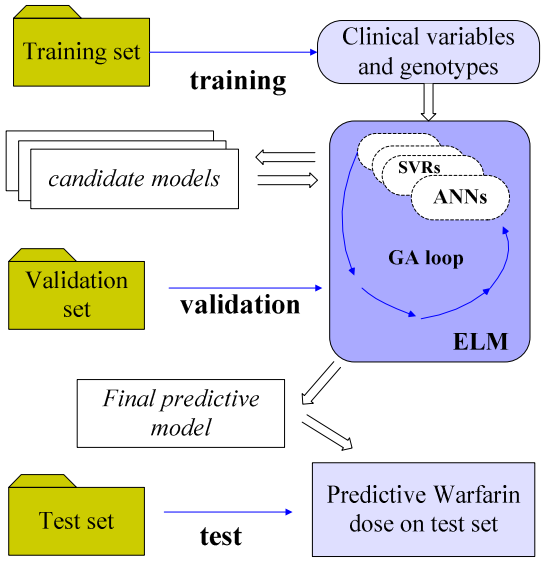

Figure.2 ELM framework for warfarin dose prediction

In ELM, the candidate models are built by BP, RBF and SVR. GA is an outer-loop for optimization on BP, RBF and SVR. GA utilizes the genetic operators, selection, crossover, and mutation, to optimize the evolutionary process of ANNs with *N* (*e.g* =20) iterations. For GA, chromosome is a real-valued string, which is defined by encoding the parameters of ANNs and SVR, and each chromosome corresponds to a process of ANN or SVR. Chromosome can be denoted as follow.

$$chr = \{C, ker, loss, e, svr\}|$$

$$\{max\_epcho, lr, [S_i], [TF_i], BTF, BLF, network\}|$$

$$\{max\_neuron, spread, eg, network\}$$

Where,

- *C*: penalty factor; *ker*: kernel functions of SVR, e.g. 'linear', 'poly', and 'rbf'; *e*: insensitivity; *loss:* loss function; *svr:* store the support regression model;
- *max_epcho*: the maximum number of iterations; $[S_i]$: the number of nodes in layers; *BTF*: training function; *BLF*: weight learning function; $[TF_i]$: transfer function; *lr*: the learning rate; *network:* the BP network model;

- *max_neuron*: the maximum number of neurons; *spread*: spread constant; *eg*: training error; *network:* the RBF network models;

Fig 2 briefly describes ELM framework for warfarin dose prediction, which uses ANN/SVR models and clinical data of patients treated with warfarin. Corresponding to Fig.2, the ELM framework consists of three main tasks, as follows:

1. Represent the collected data as feature vectors based on the clinical variables and genotypes used in predictive models;

2. Use ANNs/SVR to train a variety of nonlinear candidate models on the training dataset, and use GA to improve the performance of ANNs/SVR for better models;

3. Apply the discovered predictive model to predict warfarin dose on test set.

3.2 Flow of ELM

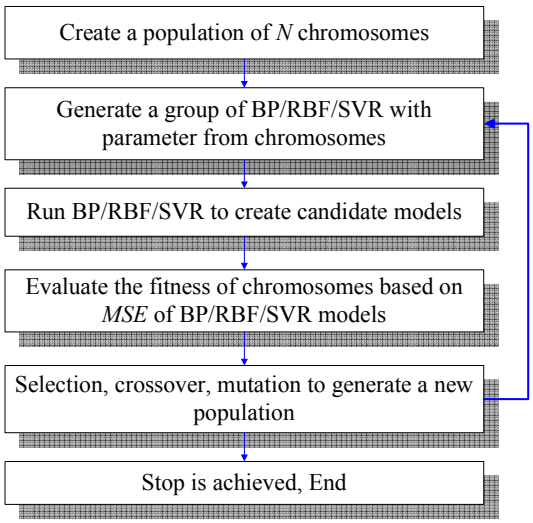

Figure.3 process of ELM

In the GA loop, selection strategy adopts the roulette wheel method. The variation on chromosomes in each generation is accomplished by mutation and crossover operators. Mutation operator randomly changes the elements of a chromosome with 100% probability. Crossover swaps the sections at same positions in two chromosomes. In each generation, 80% of population participates in mutation and 10% of population is in crossover.

Let *chr* be a chromosome of GA and m_i be the *i*-th model achieved by the corresponding ML technique. In a GA loop, the fitness function for evaluating a chromosome is denoted as (6).

$$f(chr) = \min\{ z(m_i) | BP|RBF|SVR \rightarrow \{m_i\} \} \quad (6)$$

Where $z(m_i)$ indicates the *MSE* of models $\{m_i\}$ built by BP|RBF|SVR calculated on validation set. This means the model trained by BP|RBF|SVR, who's *MSE* on validation set is smaller, will yield a higher fitness of the corresponding chromosome.

Fig3 shows the process of ELM. The *MSE* of BP/RBF/SVR models must be calculated before the fitness calculation of chromsomes. In each generation, the current best predictive model (i.e. SVR or BP or RBF built model) achieved is stored in a *pool*. A new group of BP/RBF/SVR is generated with the parameters decoded from the new chromosomes. When ELM stops, the final predictive model is the overall best one in the *pool*.

4 EXPERIMENT

In this section, we will test our proposed modeling method on the collected dataset, and take the ANNs (BP and RBF), SVR, linear regression models and currently popular predictive tool "IPWC" to compare with ELM. The performance of built predictive models is measured by three measurements, R^2 and *MSE*.

- R^2 is the squared correlation between the output of model and the actual data. R^2 illustrates that the correctness of predictive model;
- *MSE* is mean square error of model output and the desired output. *MSE* embodies an overall predictive accuracy of built model.

4.1 Data set

We have collected a dataset that contains 100 Chinese patients with warfarin treatment, which are provided by the department of cardiology in The First affiliated Hospital of Soochow University. This study was ethically permitted by the Health Authority Ethics Committee of the First Affiliated Hospital of Soochow University.

The dataset size is not large, but the dataset can be used for a preliminary study. We randomly select 60 training set, 20 verification set, and 20 test set from dataset. ELM runs five times for training on training set, and the final results are the performance of 3 best models of ten runs on training sets. The data for the clinical variables and genotypes are collected when patients' warfarin-dose requirement had remained constant for at least a minimum period of 3 months. The dataset have following features:

- 35% patients are in the range from 0 to 60 of age, 23% from 60 to 65, and 36% from 65 to 80*;*
- male constitutes 68% and female 32%*;*
- For CYP2C9 genotypes, 84% patients have AA and 16% ones have AC(a scarce genotype); For another genotype VKORC1, 78% patients have AA, 21% ones have AG, and only 1% are GG*;*
- 18% patients take amiodarone with warfarin, and 14% ones have on drinking.

4.2 Experimental Setting

Table 1: predictive models using clinical variables and genotypes

Models	variables and genotypes
M1	Age, Weight, CYP2C9, VKORC1, Amiodarone
M2	Age, Weight, Height, CYP2C9, VKORC1, Amiodarone
M3	Age, Weight, Height, Gender, Amiodarone, LA, drinking, ALT, SCr, INR

It must be noted that not all the clinical variables are bound to have influence on warfarin does-effect. Owing to the uncertainty of usefulness clinical variables, there can be variety of models with different variables. Each modeling method (ELM, SVR and ANNs) will be tested on three pharmacogenetic models listed in Table.1.

Table 2: parameters of ELM

Method		Parameters	Value
GA		*PopSize*	40
		G^{GA}	20
		P_c	0.1
		P_m	0.8
SVR	svr	Kernel function	'linear', 'poly', 'rbf'
		C(penalty factor)	1~5
		e(insensitive factor)	0.01~0.05
		Loss function	'quadratic', 'esenstivity'
RBF	newrb	Max neurons	3~20
		Goal	0.01
		Spread	1~5
BP	newff	S_{hidden}	4~10
		Epochs	5000~20000 (random)
		lr (Learning rate)	0.05
		Goal	0.01
		Transfer function	'logsig'

In order to observe the effect of clinical variables and genotypes on warfarin dose-effect, four models are given in Table.1. Model 1 is the most popularly used model including five clinical variables. Model 2 is changed from Model 1 by adding a variable "Height". Model 3, denoted by M3, includes ten clinical variables except the factors CYP2C9 and VKORC1. A predictive model without genetic factors will apply in these patients. The effect of the two genetic factors can be observed by the comparison of Model 2 with Model 3. By comparing M1, M2, and M3, we can observe the application effects of different complex models in the modeling of warfarin dose function.

The code implementation of ELM is written in the language *C* and java. The executable file of ELM runs on the server machine with CPU of 3.1GHz speed and cache of 4G. The setting on the parameters of ELM method is given in Table 2. *Popsize* indicates the population size, and G^{GA} denote the maximum generation of terminating GA. In order to reduce computational burden, population size is set to 40, and G^{GA} is set to 20. SVR, RBF and BP are implemented by using the functions of toolbox in Matlab. The functions of Matlab and their parameter setting for three methods are also listed in Table.2, the details of which can refer to the help of Matlab.

5 RESULTS AND DISCUSSION

ELM and the other machine-learning methods, RBF, BP, and SVR, will run ten times on each training set, and we present the value of the best solution obtained on training set in tables for M1, M2 and M3, respectively. We give the alternative models a fair chance through use of regularization on different parameters.

The R^2 of each discovered function over the training and testing sets and shown in Table.3. They are the best-trained solutions of ELM, RBF, BP, and SVR. Table.4 gives the *MSE* of different methods on training and test set. In the tables, "no-amiod" indicates the model without adding the variable 'amiodarone', and "amiod" indicates the model with variable 'amiodarone'.

5.1 Comparisons on R^2

It is observed in Table 3, the best R^2 is achieved by ELM when using M1 (i.e. R^2=48.1% to 72.7%), which is better than two regression models [4] and the model obtained by calculator of IWPC. M2 with variable "Height" performed a little poorer than M1 and a little better than M3. The calculator of IWPC used the same clinical variables and genotypes with M3. The R^2 of IWPC tested on our dataset ranges from R^2=52.35 to 57.02%. It must be noted that IWPC obtained the high accurate predictive model based on more than 4000 size of training dataset, which is much larger than the size of our dataset.

SVR got the best R^2 from 36.3% to 64.1% with M2, and SVR has performed very similar when using M1 and M2. BP got best R^2 from 46.7% to 74.8% with M1. When using M2, the solution in terms of R^2 is a little poorer than M1. These solutions are better than the R^2 from 47.18% to 47.88% obtained by the conventional regression models. It is also noted that SVR and BP both obtained the worst cases when using M3. RBF got a large R^2 at training set, but a very small value of R^2 at the test set in three models. This is because RBF needs a large size of dataset to build an accurate model.

By comparing the change in R^2 from training to testing set in Table.3, the generalization ability of different developed models can be demonstrated. The learning methods, ELM, SVR, and BP, present no decline of R^2 from training set to the testing set. This implies that ELM, SVR, and BP provided the models with good robustness. A large decline of R^2 provided RBF demonstrated that RBF has a drastic overfitting in these cases.

5.1 Comparisons on *MSE*

From Table.4, it can be seen that ELM obtained the overall best solutions in terms of *MSE*. It is observed that the *MSE*

obtained by ELM for M1 is better than that for M2 and M3. This illustrates the predictive accuracy cannot be improved by adding variable 'Height' into predictive model. The predictive accuracy of ELM with M2 is a little better than with M3. The change of *MSE* obtained by ELM from training to testing set is relatively small. This illustrates that the models developed by ELM can overcome the shortcomings, of using a single ML technique. BP obtained a little smaller *MSE* than SVR in some cases, but SVR performed more stable than BP. It is also noted that SVR has no decrease of *MSE* from training to test set. This demonstrates the good generalization ability of SVR in modeling. RBF had a very high accuracy on training data, but obtained too poor accuracy on test set. This demonstrates that RBF overfed the training data drastically.

Table 3: the R^2 of alternative modeling methods on training and test set

	ELM		**SVR**		**RBF**		**BP**		**Regression models**	**IWPC calculator**
	training	*test*	*training*	*test*	*training*	*test*	*training*	*test*	*training / test*	*training/ test*
M1	47.0	72.5	36.1	62.2	64.3	1.45	52.4	61.6		
	45.9	67.4	36.1	62.2	40.4	31.9	49.5	43.2		
	48.1	**72.7**	36.1	62.0	67.9	3.13	**46.7**	**74.8**		
M2	36.3	64.0	**36.3**	**64.1**	**21.8**	**31.9**	44.4	61.7	47.18/47.88 (no-amiod)	52.35
	45.6	62.7	36.3	64.1	52.8	1.4	43.0	58.1	47.36/47.58 (amiod)	/57.02
	45.2	64.2	33.2	61.9	59.1	1.4	44.1	59.7		
M3	28.7	41.5	27.8	38.8	39.8	2.1	43.0	11.3		
	32.0	32.9	27.0	38.6	62.1	2.1	55.0	14.0		
	26.9	41.0	26.8	41.6	58.2	0.001	38.3	18.4		

Table 4: the *MSE* of alternative modeling methods on training and test set

	ELM		**SVR**		**RBF**		**BP**		**Regression models**	**IWPC calculator**
	training (10^{-2})	*test (10^{-2})*	*training (10^{-2})*	*test (10^{-2})*	*training (10^{-2})*	*test (10^{-2})*	*training (10^{-2})*	*test (10^{-2})*	*training / test (10^{-2})*	*training/ test (10^{-2})*
M1	1.70	2.16	2.03	2.97	1.11	6.74	1.54	2.74		
	1.74	2.25	2.03	2.98	1.91	6.23	1.61	3.21		
	1.67	**1.97**	2.03	3.00	1.01	6.50	**1.71**	**2.29**		
M2	2.00	2.88	**2.01**	**2.88**	**2.47**	**6.08**	1.77	3.03	4.74/3.25 (no-amiod)	2.205
	1.73	2.99	2.01	2.89	1.49	6.39	1.68	3.20	3.145/2.87 (amiod)	/2.851
	1.74	2.90	2.18	3.43	1.32	6.50	1.88	3.10		
M3	2.27	3.29	2.31	4.58	1.91	6.23	1.83	6.13		
	2.16	3.87	2.31	4.60	1.20	6.59	1.44	5.29		
	2.32	4.45	2.31	4.42	1.32	6.49	1.96	5.23		

In general, learning algorithms, ELM, SVR and BP, outperform the conventional regression models and the model of IWPC in terms of three measurements. RBF is able to fit any complex function perfectly, but may seriously have over fitting when dataset size is small. This demonstrates that the machine learning techniques have great potential in practical medical applications such as warfarin dose prediction. According to results achieved by ELM and BP, the M3 and M2 including the variable 'Height' is overall poorer than model M1. This demonstrates that variable 'Height' has no influence on warfarin dose prediction under this dataset. As far as the results in this study, there is show that the variable 'CYP2C9' and 'VKORC1' in M1 and M2 can significantly improve the accuracy of warfarin dose prediction.

6 CONCLUSIONS

In this study, we have demonstrated an evolutionary learning modeling method, called ELM, to improve the warfarin-dose prediction in Chinese. ELM runs multiple-processes of ANNs and SVR to discover regression models. GA employed by ELM is to optimize the multiple-processes, and create diverse members. In the experiment, two machine-learning methods (i.e. ANNs and SVR), conventional regression models, and the seminal predictive method "IWPC" are used to compare with the ELM. The results show that the predictive models developed by ELM method present the highest accuracy on the prediction of warfarin dose. A small decrease in R^2 from training set to test set illustrates that the discovered models have better generalization ability than the other comparable methods. The conclusion is that diverse candidate models with can effectively improve the generalization ability of predictive model.

ACKNOWLEDGMENTS

This work is supported by Natural Science Found of Jiangsu province in China (Grant No. BK20140293) and National Natural Science Found of China (Grant No.61502327)

REFERENCES

[1] Klein T E, Altman R B, Eriksson N, et al. Klein T. E. et al. Estimation of warfarin dose with clinical and pharmacogenetic data. N. Engl. J. Med. 360, 753-764[J]. New England Journal of Medicine, 2009, 360(8):753-64.

[2] Jonas DE, McLeod HL. Genetic and clinical factors relating to warfarin dosing. Trends Pharmacol Sci 2009;30:375–86.

[3] Niclas E, Mia W. Prediction of warfarin dose: why, when and how?[J]. Pharmacogenomics, 2012, 13(4):429-440(12).

[4] Lai-San T, Boon-Cher G, Anne N, et al. A warfarin-dosing model in Asians that uses single-nucleotide polymorphisms in vitamin K epoxide reductase complex and cytochrome P450 2C9.[J]. Clinical Pharmacology & Therapeutics, 2006, 80(4):346-55.

[5] Wright D F B, Duffull S B. A Bayesian Dose-Individualization Method for Warfarin[J]. Clinical Pharmacokinetics, 2013, 52(1):59-68.

[6] Robert Treharne J, Mark S, David B. INRstar: computerised decision support software for anticoagulation management in primary care.[J]. Journal of the American College of Cardiology, 2011, 57(14):E1861.

[7] Idit S, Nitsan M, Gal C, et al. Applying an artificial neural network to warfarin maintenance dose prediction.[J]. Israel Medical Association Journal Imaj, 2004, 6(12):732-735.

[8] Saleh M I, Sameh A. Dosage individualization of warfarin using artificial neural networks.[J]. Molecular Diagnosis & Therapy, 2014, 18(3):371-9.

[9] Erdal C, Limdi N A, Duarte C W. High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans.[J]. Bioinformatics, 2011, 27(10):1384-1389.

[10] Hu Y H, Wu F, Lo C L, et al. Predicting warfarin dosage from clinical data: A supervised learning approach[J]. Artificial Intelligence in Medicine, 2012, 56(1):27-34.

[11] Smola A J, Schölkopf B. A tutorial on support vector regression[J]. Statistics & Computing, 2004, 14(3):199-222.

[12] Hanson LK, Salomon P. Neural network ensembles. IEEE Pattern Anal Machine Intell 1990;12(10):993–1001.

[13] Wall R, Walsh C P, Byrne S. Explaining the output of ensembles in medical decision support on a case by case basis[J]. Artificial Intelligence in Medicine, 2003, 28(2):191-206.

[14] Thomson M C, Doblas-Reyes F J, Mason S J, et al. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles.[J]. Nature, 2006, 439(7076):576-9.

[15] Caamaño P, Bellas F, Becerra J A, et al. Evolutionary algorithm characterization in real parameter optimization problems[J]. Applied Soft Computing, 2013, 13(4):1902-1921.

[16] Sharkey, Amanda J. Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems[M]. Springer-Verlag New York, Inc., 1999.

[17] Chen W C, Tseng L Y, Wu C S. A unified evolutionary training scheme for single and ensemble of feedforward neural network[J]. Neurocomputing, 2014, 143(143):347-361.