Comparisons of forecasting for hepatitis in Guangxi province, China by using three neural networks models

Ruijing Gan [1], **Ni Chen** [1], **Daizheng Huang** Corresp. [1]

[1] School of Preclinical Medicine, Guangxi Medical University, Nanning, Guangxi, China

Corresponding Author: Daizheng Huang
Email address: daizheng-huang@qq.com

This study compares and evaluates the prediction of hepatitis in Guangxi Province, China by using back propagation neural networks based genetic algorithm (BPNN-GA), generalized regression neural networks (GRNN), and wavelet neural networks (WNN). In order to compare the results of forecasting, the data obtained from 2004 to 2013 and 2014 were used as modeling and forecasting samples, respectively. The results show that when the small data set of hepatitis has seasonal fluctuation, the prediction result by BPNN-GA will be better than the two other methods. The WNN method is suitable for predicting the large data set of hepatitis that has seasonal fluctuation and the same for the GRNN method when the data increases steadily.

Comparisons of Forecasting for Hepatitis in Guangxi Province, China by Three Neural Network Models

Ruijing Gan, Ni Chen, Daizheng Huang*

School of Preclinical Medicine, Guangxi Medical University, China

*Corresponding author: daizheng-huang@qq.com

Abstract: This study compares and evaluates the prediction of hepatitis in Guangxi Province, China by using back propagation neural networks based genetic algorithm (BPNN-GA), generalized regression neural networks (GRNN), and wavelet neural networks (WNN). In order to compare the results of forecasting, the data obtained from 2004 to 2013 and 2014 were used as modeling and forecasting samples, respectively. The results show that when the small data set of hepatitis has seasonal fluctuation, the prediction result by BPNN-GA will be better than the two other methods. The WNN method is suitable for predicting the large data set of hepatitis that has seasonal fluctuation and the same for the GRNN method when the data increases steadily.

Key words: hepatitis; forecasting; neural networks method; evaluation

Introduction

Hepatitis, which is an inflammation of the liver caused by a virus, is categorized into five different types: hepatitis A, B, C, D, and E. All of these viruses cause short term or acute infection; however, the hepatitis B, C, and D viruses can cause a long-term infection, called chronic hepatitis, which can lead to life-threatening complications such as cirrhosis (liver

scarring), liver failure, and liver cancer (1). Hepatitis causes an enormous amount of human suffering, particularly in Asia, sub-Saharan Africa, parts of the Arabian Peninsula, the South Pacific, tropical South America, and Arctic North America (2). Viral hepatitis kills 1.5 million people every year and over one-third of the world's population (more than 2 billion people) have been or are actively infected by the hepatitis B virus (HBV) (1, 3). It has been reported that the direct costs due to hepatitis B reach around 500 MM Yuan RMB (approximately 80MM US dollar) in China every year (4). Guangxi, officially known as Guangxi Zhuang Autonomous Region (GZAR), is a Chinese autonomous region in South Central China that is located in the southern part of the country and is bordered to Vietnam in the southwest and the Gulf of Tonkin

in the south ($20^0$54′–$26^0$24′ N, $104^0$26′–$112^0$04′ E). It occupies an area of 236,700 km^2 with a population of over 47 million people in 2014. The typical year-round climate is subtropical rainy, which consists of long, hot summers and short winters. The annual mean temperature and rainfall are 16°C to 23°C and 1080 mm to 2760 mm, respectively (5). Guangxi Province is a high-incidence area of viral hepatitis. Hepatitis B has been in the top three infectious diseases in Guangxi Province for the past ten years. Therefore, accurate incidence forecasting of hepatitis is critical for early prevention and for better strategic planning by the government.

Prediction of incidences of hepatitis diseases has been an ongoing effort and several complex statistical models have been offered. Zhang proposed a Nash nonlinear grey Bernoulli model termed PSO-NNGBM (1,1) to forecast the incidence of hepatitis B in Xinjiang, China(6). Ren proposed a combined mathematical model using an autoregressive integrated moving average model (ARIMA) and a back propagation neural network (BPNN) to forecast

the incidence of hepatitis E in Shanghai, China (7). Ture compared time series prediction capabilities of three artificial neural networks (ANN), algorithms (multi-layer perceptron (MLP), radial basis function (RBF), time delay neural networks (TDNN)), and an auto-regressive integrated moving average (ARIMA) model to HAV forecasting (8). Gan used a hybrid algorithm combining grey model and back propagation artificial neural network to forecast hepatitis B in China (9). A mathematical model of HBV transmission was used to predict future chronic hepatitis B (CHB) prevalence in the New Zealand Tongan population with different infection control strategies in literature (10). Other studies have been performed with supervised methods for predicting viruses and pathologies (11, 12, 13).

We note that nonlinear relationships may exist among the monthly incidences of hepatitis. While the ARIMA model can only extract linear relationships within the time series data and does not efficiently extract the full relationship hidden in the historical data. The artificial neural network (ANN) time series models can capture the historical information by nonlinear functions (5).

An artificial neural network employs nonlinear mathematical models to mimic the human problem-solving process by learning previously observations to build a system of "neurons" that makes new decisions, classifications, and forecasts (14, 15). The ANN model has been successfully used to predict hepatitis A (16).

The aim of this study was to use three neural networks methods, namely, back propagation neural networks based on genetic algorithm (BPNN-GA), wavelet neural networks (WNN), and generalized regression neural networks (GRNN) to forecast hepatitis in the Guangxi Province of China, and compare the performance of these three methods. This comparison may be

helpful for epidemiologists in choosing the most suitable methodology in a given situation.

Materials and Methods

Materials

The incidence of hepatitis data, including hepatitis A, B, C, and E, were collected on a monthly base from the Chinese National Surveillance System (17) and the Guangxi Health Information Network (18) from January 2004 to December 2014. These data composed the time series $X = \{x(0), x(1), ..., x(131)\}$. The information belongs to the government statistical data and is available to the public. Hepatitis D has not been considered because the data cannot be obtained from the Chinese National Surveillance System and the Guangxi Health Information Network. The incidence dataset between 2004 and 2013 was used as the training sample to fit the model, and the dataset in 2014 was used as the testing sample.

Methods

Three artificial neural networks methods: BPNN-GA, WNN, and GRNN, were used for prediction and their performances were compared.

BPNN-GA Model

BPNN is a multi-layered feed-forward neural network; the main features are that the signal transports forward, and the error transports backward. The input signal will be processed layer-by-layer from the input layer to the output layer. The next state of the neuron is only affected by the front state of the neuron in the layer. If the expected output was not received, the weights and the thresholds of the network will be adjusted by the error that transports backward.

Therefore, the desired output will be achieved in an iterative manner (19).

If the model of BPNN has *i* input nodes, *j* hidden nodes, and *k* output nodes, there will be weight variables of dimensionality $N = i \times j$ between the input layer and the hidden layer, $j \times 1$ threshold variables in hidden layer, weight variable of dimensionality $M = j \times k$ between the hidden layer and the output layer, and $k \times 1$ threshold variables in the output layer (20). The topology structure is shown in Figure 1.

Figure 1 Topology structure of BPNN.

GA is a search heuristic that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions to optimization and search problems (21). The initial weights and thresholds of BPNN are optimized by GA, which is called the BPNN-GA method. The algorithm of the BPNN-GA flow chart is shown in Figure 2.

Figure 2 Flow chart of the BPNN prediction algorithm optimized by GA.

WNN model

WNN is a kind of neural network with a structure that is established on the basis of BPNN, and the wavelet basis function is taken as the transfer function is in hidden layer nodes. The signal also transports forward and the error transports backward. The topology structure is shown in Figure 3. WNN includes two new variables, a scale factor, a displacement factor, which give it excellent functional approximation. The WNN method is composed of relatively less expensive terms that often has fast functional approximation abilities and good predicting precision (due to its ability to sift out the parameters). Compared to BPNN, the weight coefficient of WNN has the characteristics of linearity, and the objective function of learning has the feature of convexity. These properties will avoid being nonlinear in local optimization

when the network is trained (22, 23).

Figure 3 Topology structure of WNN.

The formula for calculating the hidden layer for an input signal sequence is $x_i(i=1,2,\ldots,k)$ is as follows:

$$h(j)=h_j\left[\frac{\sum_{i=1}^{k}\omega_{ij}x_i-b_j}{a_j}\right] \qquad j=1,2,\ldots,l$$

where k is the number of input signal; l is the number of nodes in the hidden layer; $h(j)$ is the output of the jth node in the hidden layer; h_j is the wavelet basis function; ω_{ij} is the weights between the input layer and hidden layer; a_j is the scale factor of h_j and b_j is the displacement factor of h_j .

The formula to calculate the output layer is as follows.

$$y(k)=\sum_{i=1}^{l}\omega_{ik}h(i) \qquad k=1,2,\ldots,m$$

where $h(i)$ is the output of the ith node in the hidden layer; l is the number of nodes in the

hidden layer; *m* is the number of nodes in the output layer; and ω_{ij} is the weight between the hidden layer and the output layer.

The weights of the network, the scale factor, and the displacement factor were estimated by the steepest descent method in WNN. The correction process of prediction used by WNN follows.

Step 1.Calculate error of prediction.

$$e=\sum_{k=1}^{m} \mathrm{yn}(k)-y(k)$$

where $\mathrm{yn}(k)$ is the expected output, namely the true value. $y(k)$ is the forecasting output.

Step 2. Correct the weight of the network and the coefficients of wavelet basis function according to the prediction error.

$$\omega_{n,k}^{(i+1)}=\omega_{n,k}^{i}+\Delta\omega_{n,k}^{(i+1)}$$

$$a_k^{(i+1)}=a_k^i+\Delta a_k^{(i+1)}$$

$$b_k^{(i+1)}=b_k^i+\Delta b_k^{(i+1)}$$

where

$$\Delta\omega_{n,k}^{(i+1)}=-\eta\frac{\partial e}{\partial \omega_{n,k}^{(i)}}$$

134 $$\Delta a_k^{(i+1)} = -\eta \frac{\partial e}{\partial a_k^{(i)}}$$

135 $$\Delta b_k^{(i+1)} = -\eta \frac{\partial e}{\partial b_k^{(i)}}$$

136 and η is the learning rate.

137 The algorithm of WNN flow chart is shown in Figure 4.

138 **Figure 4** Flow chart of the WNN prediction algorithm.

139 **GRNN Model**

140 GRNN is a memory-based network that provides estimates of continuous variables and
141 converges to the underlying (linear or nonlinear) regression surface (24). One advantage of it is
142 the simplicity. The adjustment of one parameter, namely, the spreading factor, is sufficient for
143 determining the network.

144 The topology structure of GRNN consists of four layers: the input layer, the pattern layer,
145 the summation layer, and the output layer. The topology structure is shown in Figure 5.

146 **Figure 5** Topology structure of GRNN.

147 The number of neurons in the input layer is equal to the dimension of the input vector of
148 the learning samples. Every neuron in the input layer is the simple distribution unit and directly
149 transmits the input variables to the pattern layer.

150 The neurons in the pattern layer and the neurons in the input layer have the same number
151 and every one of the neurons in the pattern layer corresponds to a different sample. The transfer
152 function of neurons in the pattern layer is as follows

153
$$P_i = e^{¿ \; -¿ \; ¿ \; X - X_i ¿^T (X - X_i)} \qquad i = 1,2,\ldots,n$$

154 where P_i is the output of neurons in the pattern layer; $X = \left[x_1, x_2, \cdots, x_n\right]^T$ is the input

155 vector; X_i is the learning samples of the *i*-th neurons; n is the number of input; i

156 is the number of neurons; and σ is the smoothness factor.

157 There are two kinds of summation for the neurons in the summation layer. The first one is

158 that the arithmetic sum is calculated for the output of neurons in the pattern layer. The weight

159 between the pattern layer and every neuron is 1. The transfer function is shown in formula as

160 follows.

161
$$S_D = \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} ❑ e^{¿ \; -¿ \; ¿ \; X - X_i ¿^T (X - X_i)} \qquad i = 1,2,\ldots,n$$

162 The second one is the weighted sum performed for the output of neurons in the pattern

163 layer. The weight between the *i*-th neuron in the pattern layer and the *j*-th summation neuron is

164 equal to the *j*-th element in the *i*-th output samples of Y_i . The transfer function is given

165 below:

$$
\begin{gathered}
X - X_i ¿^T (X - X_i) \\
¿ \\
¿ \\
-¿ \qquad\qquad j = 1,2,\ldots,k \\
S_{\mathrm{Nj}} = \sum_{i=1}^{n} Y_{\mathrm{ij}} P_i = \sum_{i=1}^{n} Y_i e^{¿}
\end{gathered}
$$

166

167 where k is the dimension of the output vector.

168 The number of neurons in the output layer is equal to the dimension of the input vector of

169 the learning samples. The output of the *j*-th neurons is shown in formula as follows.

$$
y_i = \frac{S_{\mathrm{Nj}}}{S_D} \qquad j = 1,2,\ldots,k
$$

170

171 Results

172 The incidence of hepatitis that took place in Guangxi Province from January 2004 to

173 November 2014 is considered as the original time series $X = \{x(0), x(1), ..., x(131)\}$ and is shown

174 in Figure 6.

175 **Figure 6** The main incidence of hepatitis in Guangxi Province, China from January 2004 to

176 December 2014.

177 The incidence dataset between 2004 and 2013 was used as the training sample to fit the

model, and the dataset in 2014 was used as the testing sample.

Of all three types of ANN, the optimal four layer neurons were experimentally selected and have average square error less than 0.01. The output layer only contains one neuron representing the forecast value of the incidence of the next month.

The hidden node n_2 and the input node n_1 in the three-layer BPNN-GA were related by $n_2 = 2n_1 + 1$ and a three-layer BPNN-GA model with 4 input nodes, 9 hidden nodes, and 1 output node (4-9-1) was obtained. The selection for parameters of BPNN and GA are based on the literature (25) and (26), respectively. S-tangent function tansig() and S-log function logsig() were used as transfer functions of the hidden layer neurons and the output layer neurons, respectively. The error between the training output and the expected output (actual output) was 0.001, learning rate was 0.9, momentum factor was 0.95, the training time was 1000 iterations, and the parameters of GA were as shown in Table 1.

Table 1 Parameters of the GA used to optimize the BPNN.

There were 4 input nodes, 6 hidden nodes, and 1 output node in (4-6-1) WNN. The weights of the network, the scale factor, and the displacement factor were estimated by the steepest descent method. The initial weight was 0.01, learning rate of parameter was 0.001, and the number of iterative learning was 100. The mother wavelet basis function of Morlet was used in the paper which is shown as follows.

$$y = \cos(1.75x)e^{-x^2/2}$$

4-fold cross validation was experimentally selected and has the best prediction, which is employed to train the GRNN model and the optimal spreading factor was calculated by looping

199 from 0.1 to 2 intervals 0.1. The transfer function of the summation layer neurons used in the
200 paper is shown as follows.

201
$$\begin{gathered} X - X_i ¿^T (X - X_i) \\ ¿ \\ ¿ \\ - ¿ \end{gathered} \qquad i = 1,2,\ldots,n$$
$$S_D = \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} ❑ e^{¿}$$

202 where P_i is the output of the pattern layer neurons; $X = \left[x_1, x_2, \cdots, x_n\right]^T$ is the input
203 vector; X_i is the learning samples of the *i*-th neurons; n is the number of input; i is
204 the number of neurons; and σ is the smoothness factor.

205 The contrast between the observed values and the predicted values obtained through the
206 three methods are shown in Figure 7.

207 **Figure 7** Contrast between observed values and predicted values using the three methods.

208 # Discussion

209 ### The Relationship Between Predictions and Seasonal Fluctuation Index

210 The seasonal fluctuation index of incidence is used to reveal the fluctuations of incidence
211 with seasons. The seasonal fluctuation index of the same month in eleven years from 2004 to
212 2014 can be calculated as:

213 $$\mathrm{SFI}1=\frac{\left|\overline{x}_{\mathrm{same}}-\overline{x}_{\mathrm{all}}\right|}{\overline{x}_{\mathrm{all}}}$$

214 where $\overline{x}_{\mathrm{same}}$ is the average incidence of the same month and $\overline{x}_{\mathrm{all}}$ is the average incidence
215 of all of the months from 2004 to 2014.

216 The seasonal fluctuation index of the every month in 2014 is calculated as:

217 $$\mathrm{SFI}2=\frac{\left|x_i-\overline{x}\right|}{\overline{x}},i=1,\ldots,12$$

218 where x is the incidence in each month and $\overline{x}$ is the average incidences of all of the
219 months in 2014.

220 Obviously, the greater the number that the seasonal fluctuation index is, the more
221 seasonal volatility of incidence is. That is to say, the index changes reflect the disease
222 variation in the different months. In order to compare the relationship between the seasonal
223 fluctuation index of incidence and the three prediction results, the relative error of prediction
224 is defined as:

225 $$RE_i=\frac{\left|\hat{y}_i-y_i\right|}{y_i},i=1,2,\ldots,n$$

226 where $\hat{y}_i$ is the predicted value and y_i are the observed values.

227 The seasonal fluctuation index of incidence and the relative error of the three prediction
228 results are shown in Figure 8.

Figure 8.The relationship between the seasonal fluctuation index and RE of the predictions by the three methods. (Histograms and curves represent RE of the predictions and the seasonal fluctuation index, respectively)

Looking at Figure 8, it can be seen that: 1) hepatitis A, B, and E have obvious seasonal characteristics. For Hepatitis B, in particular, the incidence which happens annually in January and February is relatively high with a rapid decline in March. April to September is relatively stable, but from October to December it began to rise significantly; 2) the greater the seasonal fluctuation index of the every month in 2014, the greater the relative error, especially in hepatitis C and E, which shows that the greater the disease fluctuations, the worse the prediction results; 3) the absolute error of the BPNN-GA is smaller than that of the other two methods when the incidence data is stable, such as from April to August for hepatitis A; the absolute error of the GRNN is smaller than that of the other two methods when the incidence data has great fluctuation, such as March, July, and December for hepatitis C and August, October, November, and December for in hepatitis E; and 4) the absolute error of the GRNN is larger than that of the other two methods when the incidence data is larger, and the absolute error of the WNN is larger than that of the other two methods when the incidence data is smaller. The size relationship of the average incidence is: B>C>E >A. When the incidence data is large, such as the data for hepatitis B, the size relationship of the absolute error of three methods is: GRNN>BPNN-GA >WNN.

Comparison of Evaluation Indexes

The mean square error (MSE), root mean square error (RMSE), mean average error (MAE), mean average percentage error (MAPE), and sum of squared error (SSE), have been

calculated and compared with the three methods. The performance indexes are defined as shown in the following.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| * 100$$

$$SSE = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

where $\hat{y}_i$ is the predicted value and y_i are the observed value.

The main evaluation indexes that were calculated by these three methods are listed in Table 2.

Table 2 Comparison of the evaluation indexes in the prediction results.

(best performers are in bold fonts)

From the definitions of the other evaluation indexes, including MSE, MAE, RMSE, SSE, and MAPE, we know that the smaller the values of these indexes are, the more accurate the prediction is. The BPNN-GA method had the smallest values of these evaluation indexes when it was used to predict hepatitis A and the WNN method had the smallest values of these indexes when it was used to predict hepatitis B; the same for the GRNN method when it was used to predict hepatitis E. It can be seen that the BPNN-GA and WNN methods were not superior to the others when they were used to predict hepatitis C, but they were all superior to GRNN method. According to Figure 8, we know that: 1) hepatitis A, B, and E have a strong seasonal volatility, but hepatitis C fluctuates up and down monthly and does not have seasonal volatility; and 2) the incidence data of hepatitis A and B are the smallest and the largest, respectively. Hepatitis E increased slowly from January to December (except for March). That is to say, these three prediction methods have their advantages when they are used to predict seasonal fluctuation data. The BPNN-GA and WNN methods are suitable for predicting small and large data, respectively, while GRNN is suitable for predicting data that increases steadily. The BPNN-GA and WNN methods were not superior to the others when they were used to predict the data that fluctuated up and down monthly and does not have seasonal volatility, and the GRNN method is not suitable for predicting these types of data.

Comparison of Statistical Significance Tests

Statistical significance of the obtained results was investigated using T-test; a p-value of<0.05 was considered significant. The results are listed in Table 3.

Table 3 Comparison of Statistical Significance Tests in the prediction results.

(R is correlation coefficient)

The correlation will be better when the correlation coefficient is close to 1, namely, the predicted value is closer to the observed value. From Table 3, it can be seen that the BPNN-GA method has the best correlation when it was used to predict hepatitis B and C. The same in regard to the GRNN and WNN method; they had the best correlation when they were used to predict hepatitis E and A, respectively.

$P<0.01$ are for all models from Table 3 which reveals that the difference is statistically significant between the predictive value and the original data.

Conclusion

This research compared and evaluated the prediction of hepatitis by the BPNN-GA, GRNN, and WNN methods. The prediction results will be affected by the data features. When the small data set has seasonal fluctuation, the prediction result by BPNN-GA will be better than the two other methods. The WNN method is suitable for predicting the large data set that has seasonal fluctuation and the same for the GRNN method when the data increases steadily. The results of all three methods show that the greater the disease fluctuations, the worse the prediction results.

The forecasting efficacies of three models are compared based on performances. GRNN is learns faster and converges to the optimal regression surface. Capturing the dynamic behavior of hepatitis incidence. Although the BPNN is easy to fall into the local optimum and has highly non-linear weight update and slow coverage rate, the accuracy of forecasting could be improved by optimizing the initial weights and thresholds. The advantage of the BPNN is that it is suited for prediction the small data set has seasonal fluctuation. Compared to BPNN-GA and GRNN, WNN has the best performance when it is used to predict large data set with seasonal fluctuation.

This study can be extended in different directions. First, only hepatitis incidence is predicted

in the paper. In order to ascertain performance of three models and possible factors that will impact on the model performance in practice, more infectious diseases should be considered. Finally, we limited the analysis to only three ANN methods, and in future studies more methods could be tested to predict incidence of important diseases, including hepatitis.

REFERENCE

1. **Available at:** http://www.who.int

2. **Eikenberry S, Hews S, Nagy JD, Kuang Y. 2009.** The dynamics of a delay model of HBV infection with logistic hepatocyte growth. *Mathematical Biosciences and Engineering*, **6(2)**:283-99.

3. **Gourley SA, Kuang Y, Nagy JD. 2008.** Dynamics of a delay differential equation model of hepatitis B virus infection. *Journal of Biological Dynamics*. **2(2):**140-153. DOI: 10.1080/17513750701769873.

4. **Chinese Society of Hepatology and Chinese Society of Infectious Diseases, Chinese Medical Association.2010.** The guideline of prevention and treatment for chronic hepatitis B.*Journal of Clinical Hepatology*: 2011-01.

5. **Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X. 2013.** Comparative Study of Four Time Series Methods in Forecasting Typhoid Fever Incidence in China. *PLoS One* **8(5):** e63116. DOI:10.1371/journal.pone.0063116.

6. **Zhang L, Zheng Y, Wang K. 2014.** An optimized Nash nonlinear grey Bernoulli model based on particle swarm optimization and its application in prediction for the incidence of Hepatitis B in Xinjiang, China. *Computers in Biology & Medicine*, **49C(1)**:67-73. DOI: 10.1016/j.compbiomed.2014.02.008.

7. **Ren H, Li J, Yuan ZA, Hu JY, Yu Y, Lu YH. 2013.** The development of a combined mathematical model to forecast the incidence of hepatitis e in shanghai, china. *Bmc*

Infectious Diseases, **13(1)**:1-6. DOI: 0.1186/1471-2334-13-421.

8.**Ture M, Kurt I. 2006.** Comparison of four different time series methods to forecast hepatitis A virus infection, *Expert systems with application*, **31(1)**:41-46. DOI: 10.1016/j.eswa.2005.09.002.

9. **Gan RJ, Chen XJ, Yan Y, Huang DZ. 2015**.Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis b in China. *Computational and Mathematical Methods in Medicine*, **Volume 2015**, Article ID 328273, 7 pages.http://dx.doi.org/10.1155/2015/328273.

10. **Thornley S, Bullen C, Roberts M. 2008.** Hepatitis b in a high prevalence new zealand population: a mathematical model applied to infection control policy. *Journal of Theoretical Biology*, **254(3):**599-603. DOI: 10.1016/j.jtbi.2008.06.022.

11. **Weitschek E, Lo Presti A, Drovandi G, Felici G, Ciccozzi M, Ciotti M, Bertolazzi P. 2012.** Human polyomaviruses identification by logic mining techniques. *Virology journal*, **9(1)**:1-6. DOI: 10.1186/1743-422X-9-58.

12. Weitschek E, Cunial F, Felici G. 2015. LAF: Logic Alignment Free and its application to bacterial genomes classification. *Biodata Mining*, **8(1)**:39. DOI: 10.1186/s13040-015-0073-1.eCollection 2015.

13. Polychronopoulos D, Weitschek E, Dimitrieva S, Bucher P, Felici G, Almirantis Y. 2014. Classification of selectively constrained DNA elements using feature vectors and rule-based classifiers. *Elsevier Genomics*, **104(2)**:79-86.DOI: 10.1016/j.ygeno.2014.07.004.

14. **Tang Z.H, Liu J, Zeng F, Li Z, Yu X , Zhou, L. 2013.** Comparison of prediction model for cardiovascular autonomic dysfunction using artificial neural network and logistic regression analysis. *Plos One*, **8(8):** e70571. DOI:10.1371/journal.pone. 0070571.

15. **Terrin N, Schmid CH, Griffith JL, D'Agostino RB, Selker HP. 2003.** External validity of

predictive models: a comparison of logistic regression, classification trees, and neural networks. *Journal of Clinical Epidemiology*, **56(8)**:721-9. DOI: 10.1016/S0895-4356(03)00120-3

16. **Guan P, Huang DS, Zhou BS. 2005.** Forecasting model for the incidence of hepatitis a based on artificial neural network. *World Journal of Gastroenterology*, **10(24):**3579-82.

17. **Available at:** http://www.phsciencedata.cn

18. **Available at:** http://www.gxws.gov.cn /cszc/jbkzc/ yiqingxinxi

19. **Ramesh BN, Arulmozhivarman, P. 2013.** Improving forecast accuracy of wind speed using wavelet transform and neural networks. *Journal of Electrical Engineering & Technology*, **8(3):** 559-563. DOI: 10.5370/JEET.2013.8.3.559.

20. **Huang DZ, Gong RX, Gong S. 2015.** Prediction of wind power by chaos and bp artificial neural networks approach based on genetic algorithm. *Journal of Electrical Engineering & Technology*, **10(1)** **:** 41-46. DOI **:** 10.5370/JEET.2015.10.1.041.

21. **Mitchell, Melanie. 1996.** An Introduction to Genetic Algorithms. *MA: MIT Press*, Cambridge,ISBN 9780585030944.

22. **Antonios K, Alexandridis, Achilleas D, Zapranis. 2014.** Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification. *John Wiley & Sons, New Jersey*, USA, ISBN 9781118592526.

23. **Alexandridis AK, Zapranis AD. 2013.**Wavelet neural networks: a practical guide.*Neural Network*. **42:**1-27. DOI: 10.1016/j.neunet.

24. **Specht D. 1991.** A general regression neural network. *IEEE Transactions on Neural Networks*, **2(6):**568-576.DOI: 10.1109/72.97934.

25. **Zhang Le, Suganthan PN. 2016.** A survey of randomized algorithms for training neural

networks. *Information Sciences*. **364**:146-155. DOI: 10.1016/j.ins.2016.01.039.

26. **Azadeh A, Ghaderi SF, Tarverdian S, Saberi M. 2007**. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematic and Computation*. **186(2)**:1731-1741.DOI: 10.1016/j.amc.2006.08.093.

Figure 1(on next page)

Figure 1 Topology structure of BPNN.

Figure 1

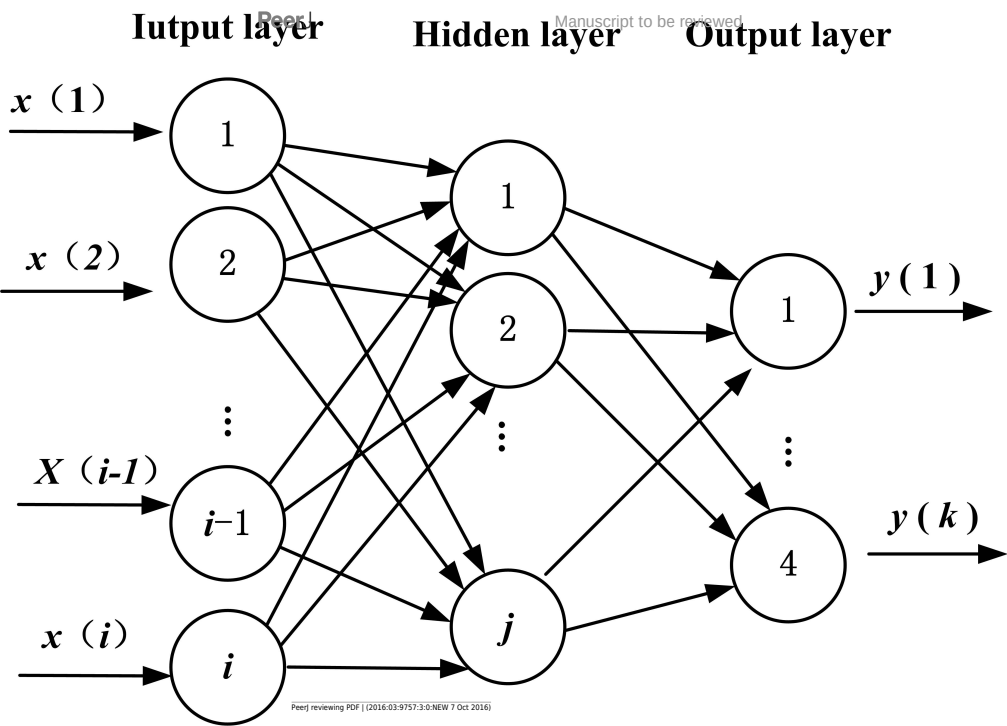

Figure 2(on next page)

Figure 2 Flow chart of the BPNN prediction algorithm optimized by GA .

Figure 2

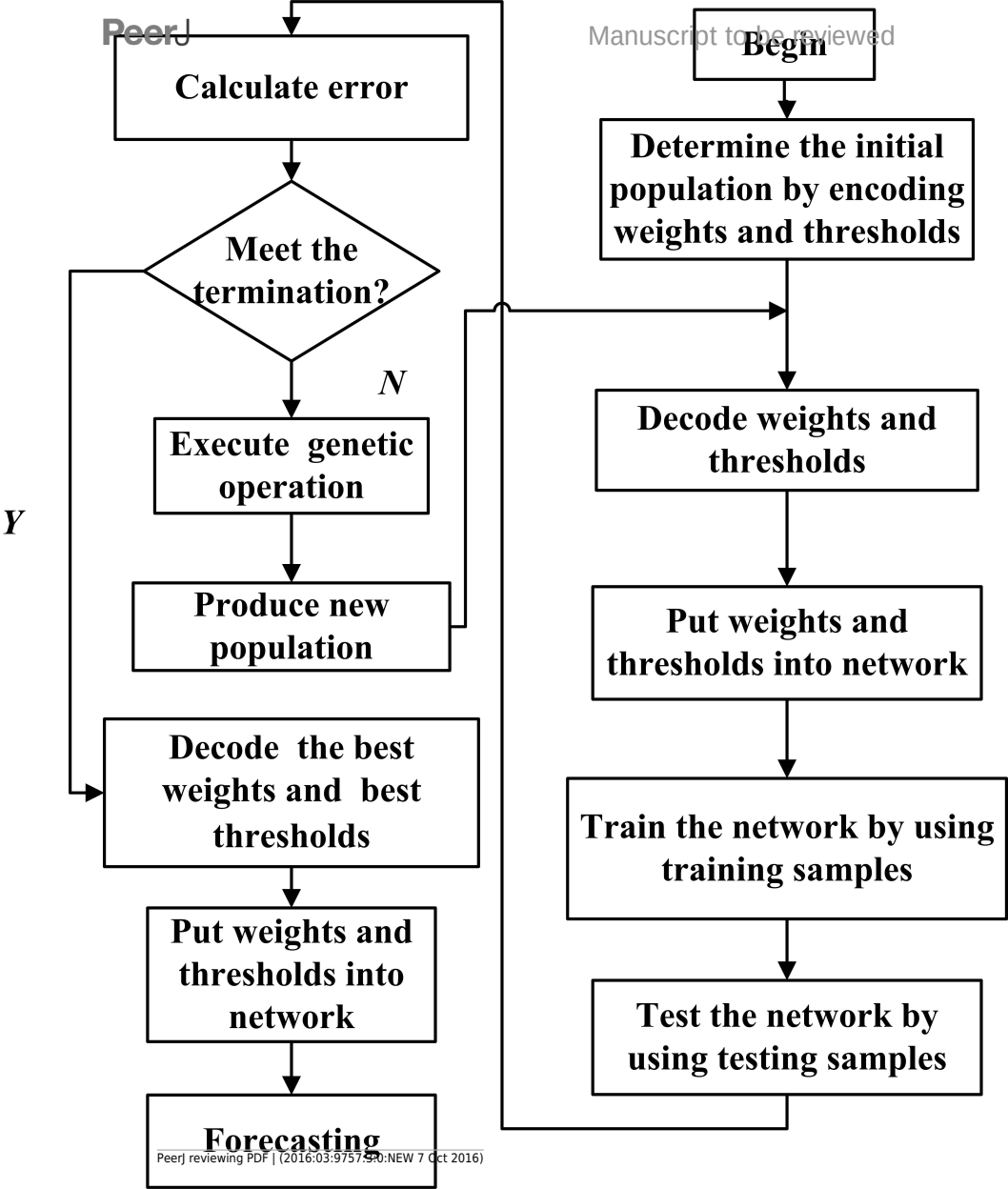

Figure 3(on next page)

Figure 3 Topology structure of WNN.

Figure 3

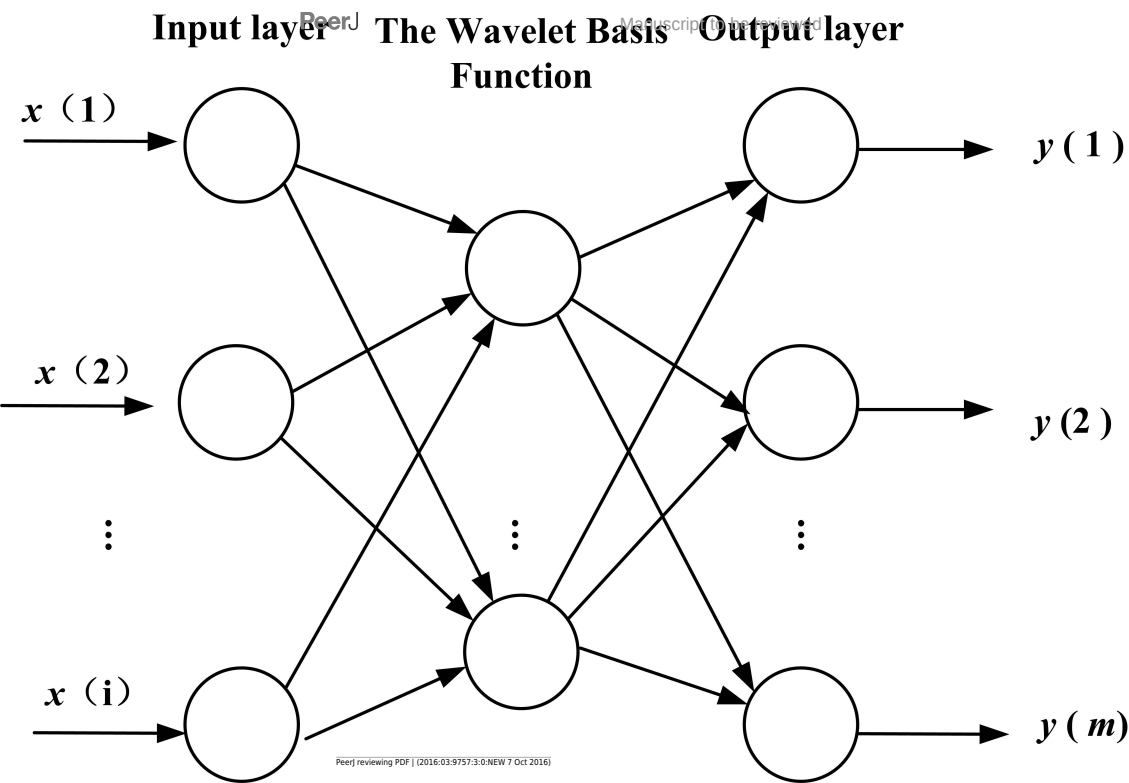

Figure 4(on next page)

Figure 4 Flow chart of the WNN prediction algorithm.

Figure 4

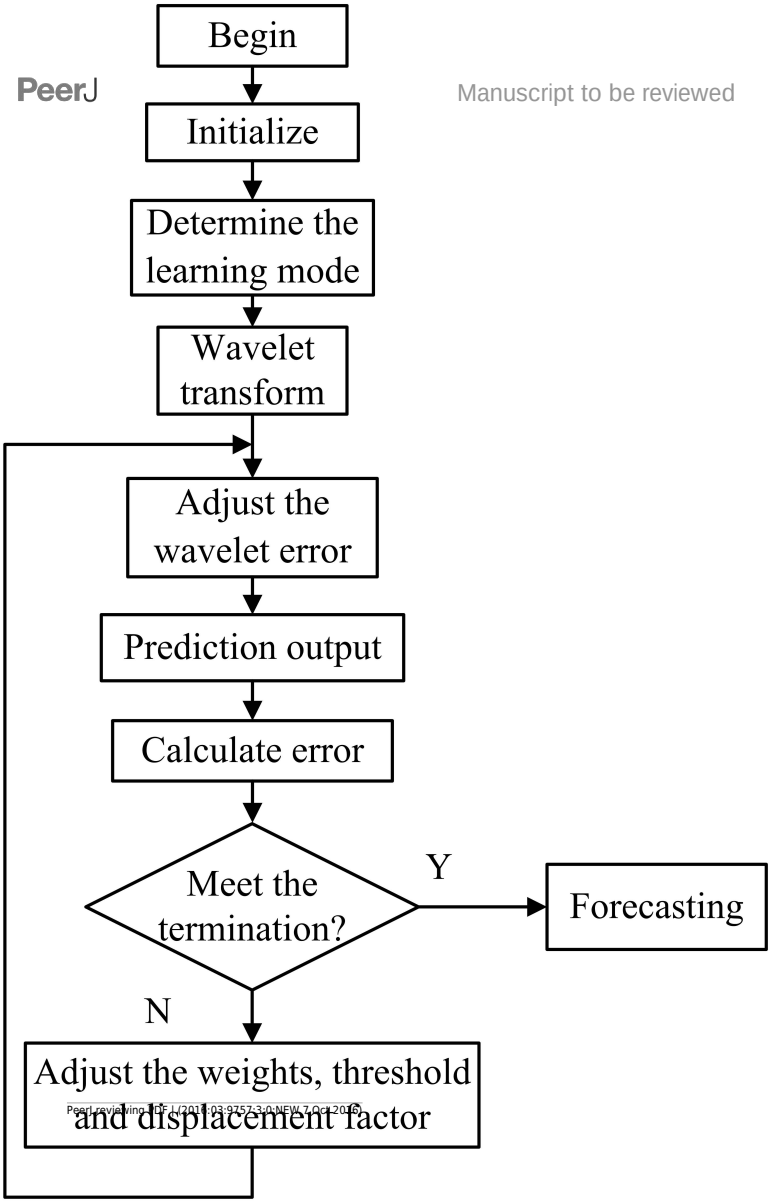

Figure 5(on next page)

Figure 5 Topology structure of GRNN.

Figure 5

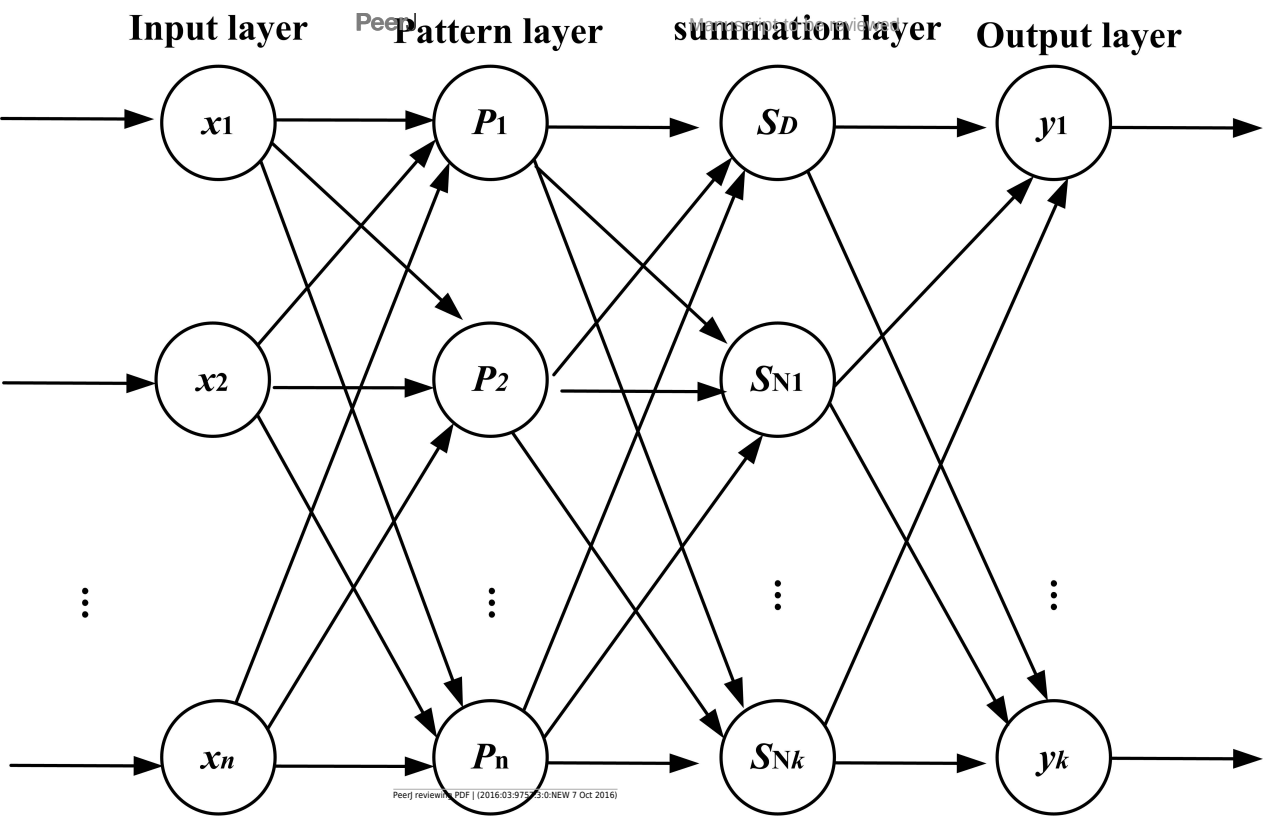

Figure 6(on next page)

Figure 6 The main Incidence of hepatitis in Guangxi Province , China from January 2004 to December 2014.

Figure 6

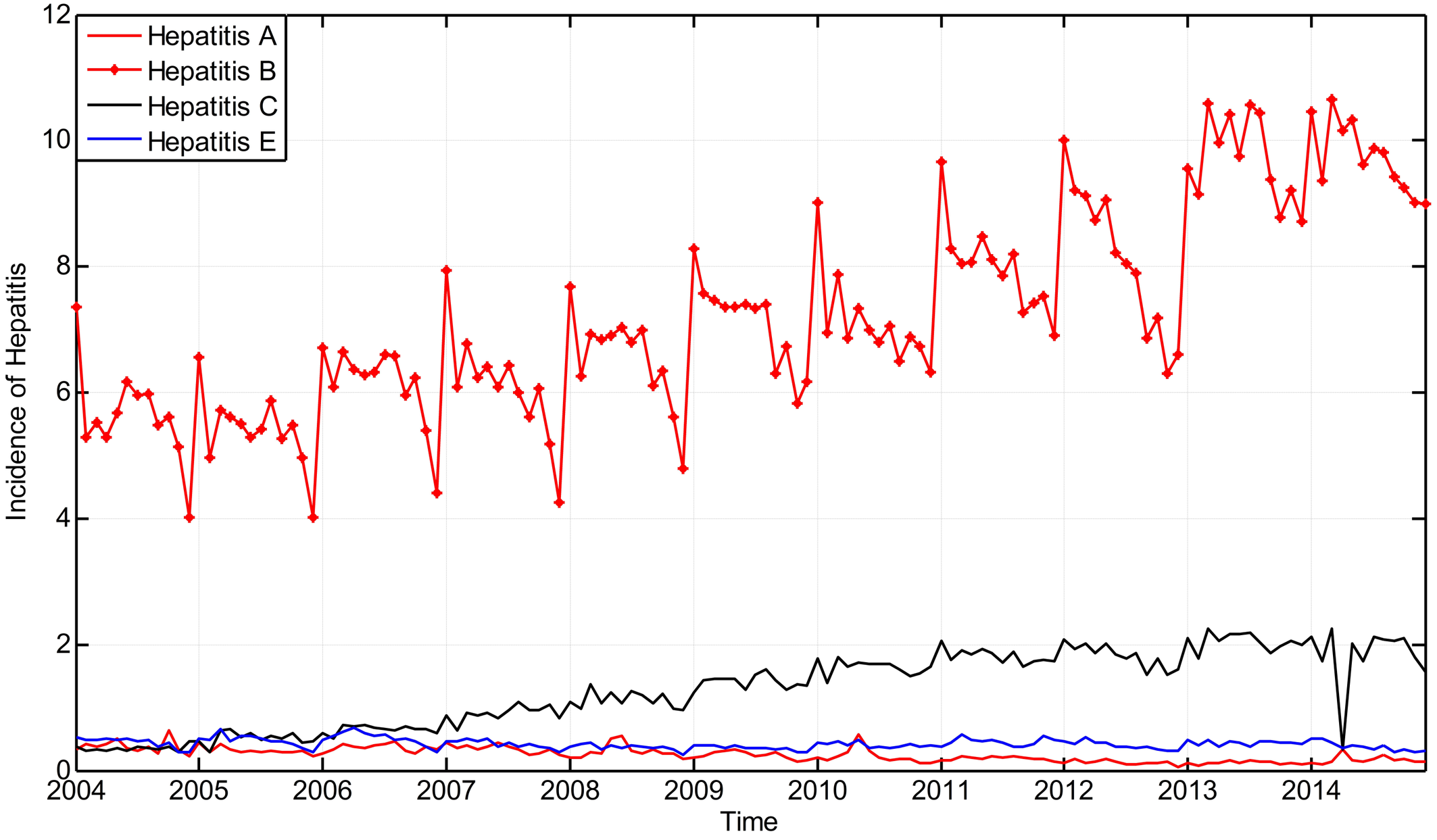
Hepatitis A
Hepatitis B
Hepatitis C
Hepatitis E
Incidence of Hepatitis
Time
0
2
4
6
8
10
12
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

Figure 7(on next page)

Figure 7 Contrast between observed values and predicted values using the three methods.

Figure 7

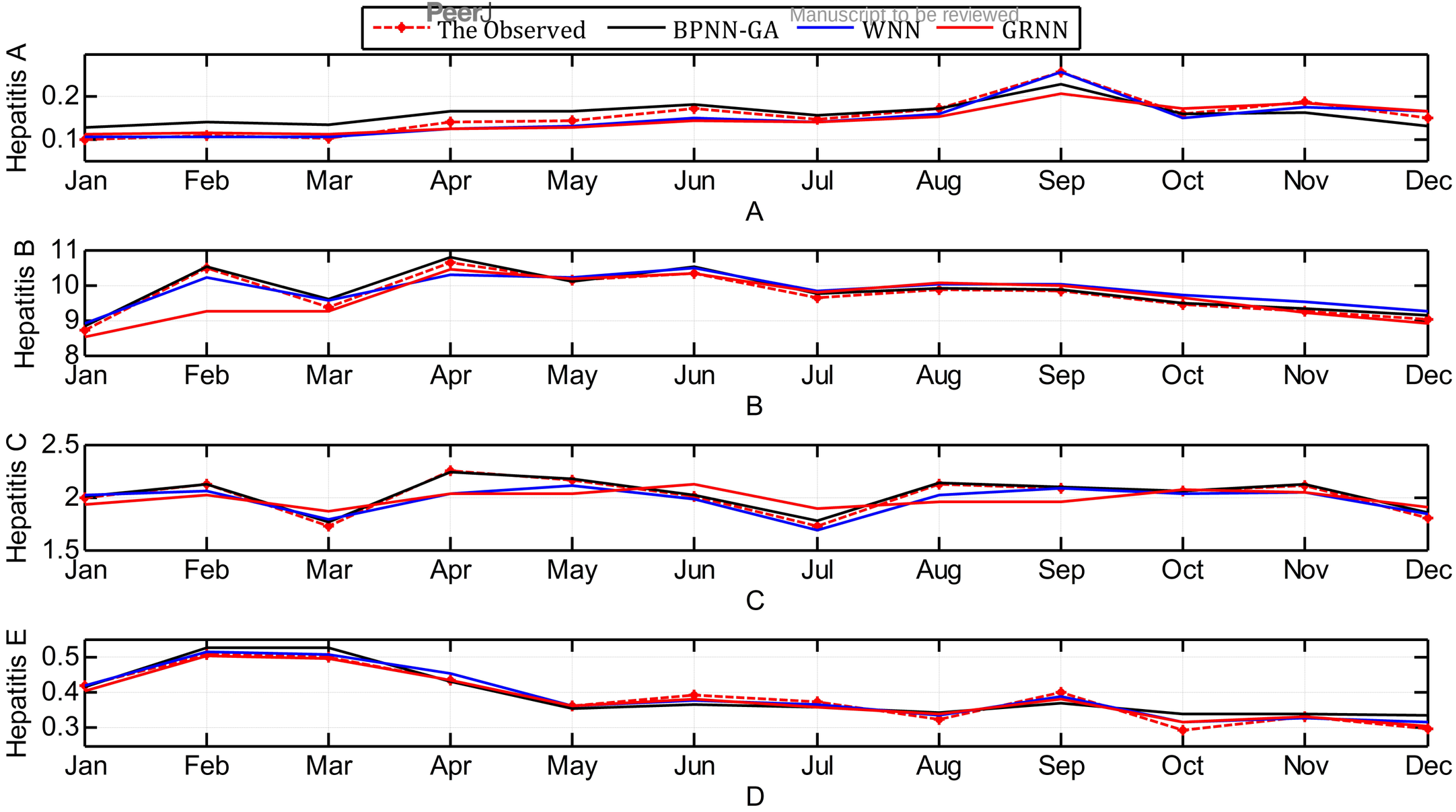
The Observed
BPNN-GA
WNN
GRNN
Hepatitis A
0.2
0.1
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
A
Hepatitis B
11
10
9
8
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
B
Hepatitis C
2.5
2
1.5
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
C
Hepatitis E
0.5
0.4
0.3
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
D

Figure 8(on next page)

Figure 8.The relationship between the seasonal fluctuation index and RE of the predictions by the three methods. (Histograms and curves represent RE of the predictions and the seasonal fluctuation index, respectively)

Figure 8

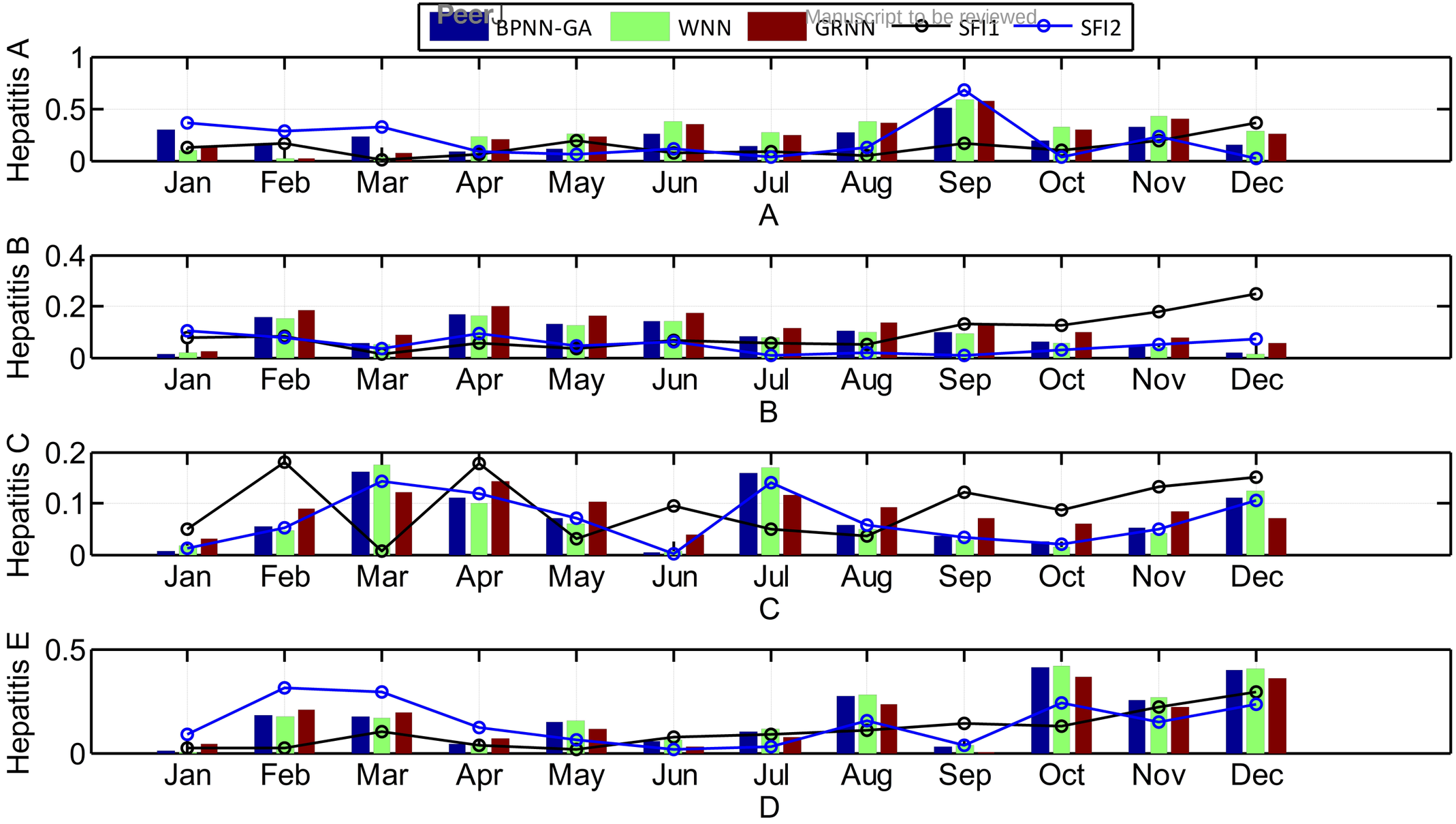

Table 1(on next page)

Table 1 Parameters of the GA used to optimize the BPNN.

Table 1

1 **Table 1** Parameters of GA

Population size	40
Algebra	50
Number of bits	10
Crossover probability	0.7
Mutation probability	0.01
Generation gap	0.95

2

Table 2(on next page)

Table 2 Comparison of the evaluation indexes in the prediction results.(best performers are in bold fonts)

Table 2

1 **Table 2** Comparison of Evaluation Indexes (best performers are in bold fonts)

Hepatitis	Method	MSE	MAE	RMSE	SSE	MAPE
A	BPNN-GA	**0.0024**	**0.0377**	**0.0488**	**0.0286**	**3.7743**
	WNN	0.0038	0.0480	0.0616	0.0455	4.7955
	GRNN	0.0034	0.0456	0.0587	0.0413	4.5566
B	BPNN-GA	1.1018	0.9008	1.0497	13.2217	90.0830
	WNN	**1.0285**	**0.8652**	**1.0141**	**12.3414**	**86.5163**
	GRNN	1.7907	1.2085	1.3382	21.4889	120.8490
C	BPNN-GA	**0.0273**	0.1376	**0.1651**	**0.3272**	13.7552
	WNN	**0.0273**	**0.1330**	0.1652	0.3274	**13.3042**
	GRNN	0.0338	0.1713	0.1839	0.4058	17.1327
E	BPNN-GA	0.0054	0.0617	0.0733	0.0645	6.1665
	WNN	0.0055	0.0626	0.0745	0.0665	6.2620
	GRNN	**0.0048**	**0.0577**	**0.0696**	**0.0582**	**5.7701**

2

Table 3(on next page)

Table 3 Comparison of Statistical Significance Tests in the prediction results.(R is correlationcoefficient)

Table 3

1 **Table 3** Comparison of Statistical Significance Tests(R is correlation coefficient)

Hepatitis	Statistic value	BPNN-GA	WNN	GRNN
A	R	0.8992	0.9686	0.9129
	p-value	0. 00006969	0.00000023	0. 00003383
B	R	0.9916	0.9575	0.8030
	p-value	0.00000000	0.00000102	0.00166221
C	R	0.9991	0.9141	0.6903
	p-value	0.00000000	0.00003198	0.01295323
E	R	0.9409	0.9835	0.9847
	p-value	0.00000510	0. 00000001	0.00000001

2