International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:9, No:5, 2015

A Medical Resource Forecasting Model for Emergency Room Patients with Acute Hepatitis

R. J. Kuo, W. C. Cheng, W. C. Lien, T. J. Yang

***Abstract*—**Taiwan is a hyper endemic area for the Hepatitis B virus (HBV). The estimated total number of HBsAg carriers in the general population who are more than 20 years old is more than 3 million. Therefore, a case record review is conducted from January 2003 to June 2007 for all patients with a diagnosis of acute hepatitis who were admitted to the Emergency Department (ED) of a well-known teaching hospital. The cost for the use of medical resources is defined as the total medical fee. In this study, principal component analysis (PCA) is firstly employed to reduce the number of dimensions. Support vector regression (SVR) and artificial neural network (ANN) are then used to develop the forecasting model. A total of 117 patients meet the inclusion criteria. 61% patients involved in this study are hepatitis B related. The computational result shows that the proposed PCA-SVR model has superior performance than other compared algorithms. In conclusion, the Child-Pugh score and echogram can both be used to predict the cost of medical resources for patients with acute hepatitis in the ED.

***Keywords*—**Acute hepatitis, Medical resource cost, Artificial neural network, Support vector regression.

I. Introduction

ACUTE hepatitis lasts less than 6 months, but chronic hepatitis lasts longer than 6 months. Acute hepatitis has several possible causes, such as infectious viral hepatitis (hepatitis A, B, C, D and E), other viral diseases (glandular fever and cytomegalovirus), severe bacterial infections, amoebic infections, medicines (acetaminophen and halothane) and toxins (alcohol and fungal toxins). The severity of illness in acute hepatitis ranges from asymptomatic to fulminant and fatal. Some patients are asymptomatic, with abnormalities noted only by laboratory studies, while other patients might have symptoms such as nausea, vomiting, fatigue, weight loss, abdominal pain, jaundice, fever, splenomegaly, or ascites.

Chronic hepatitis also has several different causes, such as infectious viral hepatitis (hepatitis B, C and D), reaction to drugs, alcohol, autoimmune hepatitis, Wilson's disease and hemochromatosis. According to an investigation by the World Health Organization (WHO), 2 billion people are infected with the hepatitis B virus (HBV) and more than 350 million have chronic (lifelong) infections [1]. Hepatitis B results in 600,000 deaths each year from cirrhosis and hepatocellular carcinoma [2]. Taiwan is a hyper-endemic area for the Hepatitis B virus (HBV). The estimated total number of HBsAg carriers in the general population who are more than 20 years old is 3,067,307. 61% patients involved in this study are hepatitis B related [3].

The typical signs of severe acute exacerbation in a patient with chronic hepatitis B are a short onset of jaundice and a very high ALT level, sometimes preceded by prodromal constitutional symptoms. It is often misdiagnosed as acute hepatitis, especially in those with no history of hepatitis or who have never been examined for hepatitis markers [4]. The symptoms of severe acute exacerbation of chronic hepatitis B can be very similar to those of acute hepatitis B. Therefore, severe acute exacerbation of chronic hepatitis B can be misdiagnosed as acute hepatitis B in some cases [5].

In countries with intermediate or high levels of HBV infection, the exacerbation of chronic hepatitis B may be the first presentation of HBV infection [6]. Although this exacerbation is usually transient and asymptomatic, 1%–2.4% of patients later develop hepatic decompensation [7]-[9]. The authors believe that a great majority of patients with acute hepatitis B suffer from acute exacerbation of chronic hepatitis B.

The Child-Pugh classification has been used for decades to measure the severity of chronic liver disease. Recent studies have shown that the model for end-stage liver disease (MELD) more accurately predicts the short and mid-term probability of survival for patients with cirrhosis than the CTP system. MELD, which has 3 parameters (serum bilirubin, creatinine and prothrombin time) that require logarithmic transformation, has the advantage of a wide-range continuous scale and is more objective and less variable [10]. Reports of predictors for acute hepatitis include the MELD Scoring System, the Discriminant Function (DF) and Multivariate analysis. However, there are limited reports of the use of an abdominal echogram and the Child-Pugh classification as a predictor for acute hepatitis.

Sheth et al. [11] used the MELD score and the Discriminant Function (DF) as a predictor of mortality for 34 patients who were hospitalized with alcoholic hepatitis. At 30 days, the MELD score predicts mortality as well as the DF. A MELD score of greater than 11, or the presence of both ascites and an elevated bilirubin level of greater than 8 mg/dL should prompt specific therapeutic intervention to reduce mortality.

Li et al. [12] investigated the prognostic factors for chronic severe hepatitis and constructed a prognostic model. The clinical and laboratory indices of 213 patients with chronic severe hepatitis were analyzed retrospectively within 24 hours after diagnosis. Death or survival was limited to within 3

R. J. Kuo and W. C. Cheng (corresponding author) are with the Department of Industrial Management, National Taiwan University of Science and Technology, Taipei City, Taiwan (phone: +886-2-2739-6328, +886-2-2733-3141#7103; fax: +886-2-2737-6344. e-mail: rjkuo@mail.ntust.edu.tw, jwz.cheng@gmail.com).

W. C. Lien is with the Department of Emergency Medicine, College of Medicine, National Taiwan University and National Taiwan University Hospital, Taipei City, Taiwan (e-mail: dtemer17@yahoo.com.tw).

T. J. Yang is with the Department of Emergency Medicine, Sijhih Cathay General Hospital, Taipei City, Taiwan (e-mail: yangteckjin@hotmail.com).

months after diagnosis. The mortality rate for all patients was 47.42%. Compared with the group that survived, the age, the basis of hepatocirrhosis, infection, the degree of hepatic encephalopathy (HE) and the levels of total bilirubin (TBil), total cholesterol (CHO), cholinesterase (CHE), blood urea nitrogen (BUN), blood creatinine (Cr), blood sodium ion (Na), peripheral blood leukocytes (WBC), alpha-fetoprotein (AFP), international normalized ratio (INR) of blood coagulation and prothrombin time (PT) were significantly different in the group who died. They concluded that Multivariate analysis in the prognosis for chronic severe hepatitis and the regression model was of significant value in the prognosis for this disease.

The prognosis for liver cirrhosis is assessed using the Child-Pugh score. Since abdominal ultrasound is also a commonly used tool for the evaluation and rapid diagnosis of acute hepatitis in the ED, this study determines whether the Child-Pugh score and abdominal ultrasound can be used together to predict the cost of medical resources for patients with acute hepatitis, but without liver cirrhosis, in the ED. In 2008, the consensus recommendation of the Asian Pacific Association for the study of the liver (APASL) indicated several prognostic scores for acute-on-chronic liver failure (ACLF) [13]. Liver-specific scoring systems (Mayo Risk Score, Combined Clinical and Laboratory Index) are adequate, but the Acute Physiology and Chronic Health Evaluation (APACHE) II and III have been proven to be more efficient, because they include additional physiologic parameters, so they also take into account additional complications that are associated with this liver disorder [14], [15]. An accurate prognosis for the patient allows the appropriate selection of a treatment program.

Data mining can extract hidden association rules from large databases to give useful knowledge. The advances in data mining techniques in recent years mean that scholars have begun to use them for metabolic syndrome research. Huang [16] used an efficient data mining algorithm, called DCIP (the data cutting and inner product method), to determine the association rules between the lifestyles of factory workers in Taiwan and the metabolic syndrome.

The remainder of this paper is organized as follows. The methodology and literature review are presented in Section II. Section III shows the computational results. Finally, the concluding remarks are made in Section IV.

II. Methodology

In order to develop the forecasting model, two novel methods: an artificial neural network (ANN) and support vector regression (SVR) are employed in this study. Before constructing the model, the data were collected from the emergency department of a Hospital in Taiwan. The proposed forecasting method has two stages: (1) feature selection and (2) forecasting. In the first stage, principal component analysis is used to select the important features. An ANN and SVR are then employed to determine the relationship between selected features and the output variable: cost.

A. Data Collection and Descriptive Statistics

An electronic search was performed on the medical record database for all patients admitted to the emergency department of a Hospital in Taiwan with an admission diagnosis of hepatitis (ICD-9 code 070 or 570), from January 2003 to June 2007. These patients were then further screened for the presence of liver enzyme (AST/ALT) above 400 and were excluded if there was liver cirrhosis. 117 patients were included in this study: 76 male and 41 female, from 21 to 77 years. The ratio of male to female is nearly 2:1. TABLE I shows the number of samples in terms of gender and age.

TABLE I
Descriptive Statistic of Samples

Features		N	Ratio
Gender	Female	41	35%
	Male	76	65%
Age	20-29	23	20%
	30-39	28	24%
	40-49	25	21%
	50-59	27	23%
	> 60	14	12%

B. Principal Component Analysis (PCA)

In multivariate statistical analysis, principal component analysis (PCA) is used to simplify the dataset. PCA is often used to reduce the dimensions of the dataset [17]. Its goal is to determine the important information from the table, in order to represent it as a set of new orthogonal variables, called principal components, and to display the pattern of similarity between the observations and the variables as points on a map. PCA can be generalized as canonical correlation analysis (CCA). CCA can be used to define the coordinate systems that optimally describe the cross-covariance between two datasets. Mathematically, PCA depends upon the eigen-decomposition of positive semi-definite matrices and upon the singular value decomposition (SVD) of rectangular matrices.

C. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a system that is derived from neurophysiology models. In general, it consists of a collection of simple nonlinear computing elements whose inputs and outputs are linked to form a network. There are three types of the ANN algorithm: supervised, unsupervised and hybrid learning. ANN's are nonparametric data driven ways to capture nonlinear data structures without prior assumption about the underlying relationships in a particular problem. ANNs are more general and flexible modeling and analysis tools for forecasting applications in that not only do they find nonlinear structures, but they can also model linear processes [18].

Many studies have attempted to use ANN's for time-series forecasting. However, their conclusions are often contradictory. Some studies conclude that ANN's are better than conventional methods [19], while others reach an opposite conclusion. The ANN approach is a leading statistical modeling approach. ANN has been found to be useful for modeling industrial problems [20], [21]. Zheng et al. [22] used an ANN system to predict the

3-month mortality risk for acute-on-chronic hepatitis B liver failure (ACHBLF) for individual patients.

There are several procedures that must be followed when this method is used.

Step 1. The parameters, including the learning rate (η), the momentum (α) and the number of training iterations must be established.

Step 2. The connecting weights, W^{xh} (input layer to hidden layer), W^{hy} (hidden layer to output layer) and the bias weights, θ^h (in the hidden layer) and θ^y (in the output layer) must be established randomly.

Step 3. The training data is input.

Step 4. The output, *Y*, is computed using (1) to (4).

$$net_k = \sum_i W_{ik}^{xh} \cdot X_i - \theta_k^h \quad (1)$$

$$H_k = f(net_k) = 1/1 + \exp(-net_k) \quad (2)$$

$$net_j = \sum_k W_{kj}^{hy} \cdot H_k - \theta_j^y \quad (3)$$

$$Y_j = f(net_j) = 1/1 + \exp(-net_j) \quad (4)$$

where $i = 1 \ldots N$, N is the number of inputs, $k = 1 \ldots M$, M is the number of hidden layer nodes and $j = 1 \ldots P$, P is the number of outputs.

Step 5. The amount of difference, δ, is computed using (5) and (6).

$$\delta_j = Y_j(1 - Y_j)(T_j - Y_j) \quad (5)$$

$$\delta_k = H_k(1 - H_k) \sum_j W_{kj}^{hy} \delta_j \quad (6)$$

where M is the number of hidden layer nodes and $j = 1 \ldots K$, K is the number of outputs.

Step 6. The updating amount of the connecting weight ΔW^{hy}, (hidden layer to output layer), ΔW^{xh} (input layer to hidden layer) and the amount of bias weight, $\Delta\theta^y$ (in output layer) and $\Delta\theta^h$ (in hidden layer) are computed using (7) to (10).

$$\Delta W_{kj}^{hy} = \eta \delta_j H_k + \alpha \cdot \Delta W_{kj}^{hy} \quad (7)$$

$$\Delta\theta_j^y = -\eta\delta_j + \alpha \cdot \Delta\theta_j^y \quad (8)$$

$$\Delta W_{ik}^{xh} = \eta \delta_k X_i + \alpha \cdot \Delta W_{ik}^{xh} \quad (9)$$

$$\Delta\theta_k^h = -\eta\delta_k + \alpha \cdot \Delta\theta_k^h \quad (10)$$

where $i = 1 \ldots N$, N is the number of inputs, $h = 1 \ldots M$, M is the number of hidden layer nodes and $j = 1 \ldots K$, K is the number of outputs.

Step 7. W^{xh}, W^{hy}, θ^h, and θ^y are updated using (11)-(14).

$$W_{kj}^{hy} = W_{kj}^{hy} + \Delta W_{kj}^{hy} \quad (11)$$

$$\theta_j^y = \theta_j^y + \Delta\theta_j^y \quad (12)$$

$$W_{ik}^{xh} = W_{ik}^{xh} + \Delta W_{ik}^{xh} \quad (13)$$

$$\theta_k^h = \theta_k^h + \Delta\theta_k^h \quad (14)$$

where $i = 1 \ldots N$, N is the number of inputs, $h = 1 \ldots M$, M is the number of hidden layer nodes and $j = 1 \ldots K$, K is the number of outputs.

Step 8. Steps 3 to 7 are repeated until the termination criterion is satisfied.

D. Support Vector Regression

SVR is a novel neural network algorithm technique that uses statistical learning theory, which has is increasingly used to solve nonlinear regression estimation problems. SVR is derived from the structural risk minimization principle, which estimates a function by minimizing an upper bound of the generalization error. It has been successfully applied in different time series prediction problems, such as production value forecasting, traffic flow prediction and financial time series forecasting [23]-[26]. The SVR model can be expressed as [27]:

$$f(x) = w^T \varphi(x) + b, \quad (15)$$

$f(x)$ denotes the forecasting values and the coefficient w($w \epsilon \Re^{n_k}$) and b ($b \epsilon \Re$) are adjustable. SVR method aims at minimizing the empirical risk as (16):

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} \Theta_\varepsilon(y_i, w^T \varphi(x_i) + b), \quad (16)$$

$\Theta_\varepsilon(y_i, w^T \varphi(x_i) + b)$ is the ε-insensitive loss function and defined as (17):

$$\Theta_\varepsilon(y_i, w^T \varphi(x_i) + b) = \begin{cases} |w^T \varphi(x_i) + b - y_i| - \varepsilon, if\ |w^T \varphi(x_i) + b - y_i| \geq \varepsilon \\ 0, otherwise. \end{cases} \quad (17)$$

In addition, $\Theta_\varepsilon(y_i, w^T \varphi(x_i) + b)$ is employed to determine an optimum hyper plane on the high dimensional feature space that maximizes the distances between the training data into two subsets. Therefore, the SVR determines the optimum hyper plane and minimizes the training error between the training data and the ε-insensitive loss function. The SVR then minimizes the overall errors as:

$$\text{Min}_{w,b,\xi^*,\xi}\, R_\varepsilon(w, \xi^*, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} (\xi_i^* + \xi_i) \quad (18)$$

Subject to:

$$y_i - w^T \varphi(x_i) - b \leq \varepsilon + \xi_i^*, i = 1,2, \ldots, N$$
$$-y_i + w^T \varphi(x_i) + b \leq \varepsilon + \xi_i^*, i = 1,2, \ldots, N$$
$$\xi_i^* \geq 0, i = 1,2, \ldots, N$$
$$\xi_i \geq 0, i = 1,2, \ldots, N$$

The first term of (18), which uses the concept of maximizing the distance of two separate training data, is used to regularize the weights, to penalize large weights and to maintain the flatness of the regression function. The second term penalizes training errors of f(x) and y using the ε-insensitive loss function. C is a parameter that trades off these two terms. Training errors above ε are denoted as ξ_i^* and training errors below εde are noted as ξ_i.

When the quadratic optimization problem with inequality on constraints is solved, the parameter vector, w, in (19) is obtained as:

$$w = \sum_{i=1}^{N}(\beta_i^* - \beta_i)\varphi(x_i), \quad (19)$$

where β_i^*, β_i are obtained by solving a quadratic program and are the Lagrangian multipliers. Finally, the SVR function is obtained as (20) in the dual space:

$$f(x) = \sum_{i=1}^{N}(\beta_i^* - \beta_i)\mathrm{k}(x_i, x_j) + b, \quad (20)$$

where $\mathrm{k}(x_i, x_j)$ is called the kernel function. The value of the kernel equals the inner product of two vectors, x_i and x_j, in the feature space, $\varphi(x_i)$ and, $\varphi(x_j)$ respectively. That is $\mathrm{K}(x_i, x_j) = \varphi(x_i)\ \varphi(x_j)$. Any function that meets Mercer's condition can be used as the kernel function. There are several types of kernel function. The most commonly used kernel functions are the Gaussian radial basis functions (RBF) with a width of σ: $\mathrm{K}(x_i, x_j) = \exp(-0.5\|x_i - x_j\|^2/\sigma^2)$ and the polynomial kernel with an order of d and constants a_1 and a_2 : $\mathrm{K}(x_i, x_j) = (a_1 x_i x_j + a_2)^d$. If the value of σ is very large, the RBF kernel approximates the use of a linear kernel (a polynomial with an order of 1). However, the Gaussian RBF kernel is not only easier to implement, but it can also nonlinearly map the training data into an infinite dimensional space, so it is suitable for nonlinear relationship problems. Therefore, the Gaussian RBF kernel function is specified in this study. It is known that the forecasting accuracy of an SVR model depends on good establishment of the hyper parameters, C, ε and the kernel parameters (σ). Therefore, the determination of all three parameters is also important.

III. Computational Results

This section presents the validation results for the two forecasting methods: the ANN and the SVR. The proposed method has two stages: (1) feature selection and (2) forecasting. The outcomes of these two stages are discussed as follows.

A. Features Selection

The original number of features is six, as follows:

- X_1 Total bilirubin,
- X_2 Prothrombin time,
- X_3 Ascites,
- X_4 Portal vein dilation,
- X_5 GB wall thickening,
- X_6 Splenomegaly,
- Y Total cost.

Features X_1 to X_6 are input variables and output Y is the total cost of the medical resources. After PCA, four features more appropriately represent the data characteristics. These four features are treated as input variables for the ANN and the SVR.

B. Construction of the Forecasting Model

In this study, an ANN and SVR are used to construct the forecasting models. The LIBSVM package developed by Chang and Lin [28] is used to construct the SVR model. The ANN is coded in C++ programming language. The forecasting performance is evaluated using the performance measures, the mean square error (MSE) and the mean absolute difference (MAD). The definitions of the measures are shown as (21) and (22):

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - T_i)^2 \quad (21)$$

$$\mathrm{MAD} = \frac{1}{N}\sum_{i=1}^{N}|Y_i - T_i|, \quad (22)$$

where T_i is the actual value, Y_i is the forecast value and n is the total number of data. The parameters for both algorithms are important, since they have a significant effect on the performance. The parameters for the PCA-ANN are learning rate (η) and momentum (α) and the parameters for the PCA-SVR are (C) and (ε). The objective is to minimize the MSE.

TABLE II shows the test results for the ANN and the results for the SVR are shown in

TABLE III. These results show that the parameters for the ANN (α, η) are (0.99, 0.01) and (C, ε) are (7, -5) for the SVR. The convergence curve for the ANN is shown in Fig. 1.

TABLE II
RESULT OF PCA-ANN PARAMETERS TESTING

α	η	MSE
0.99	0.10	0.0127
	0.05	0.0123
	0.01	**0.0122**
0.95	0.10	0.0129
	0.05	0.0130
	0.01	0.0129
0.90	0.10	0.0129
	0.05	0.0130
	0.01	0.0129

TABLE III
RESULT OF PCA-SVR PARAMETERS TESTING

C	ε	MSE
7	**-5**	**0.00111**
9	-5	0.00127
8	-5	0.00136
-7	4	0.00138
-7	6	0.00146

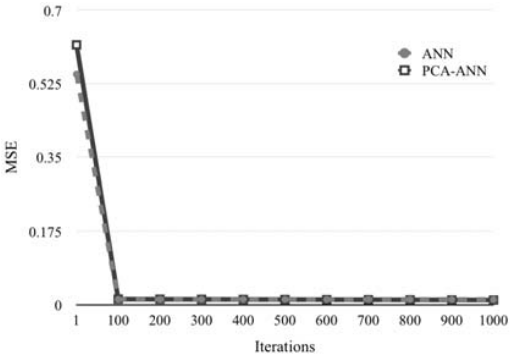

Fig. 1 Convergence curve of ANN

In order to compare the forecasting models' performance, *K*-fold cross-validation is used. In the experiment, the data are divided into 10 groups. 90% of the data is used for training and 10% for testing. The results for the PCA-ANN and the PCA-SVR are shown in

TABLE IV and

TABLE V, respectively. For the PCA-ANN, the average training MSE and the average test MSE are 0.0121 and 0.0033, respectively. For the PCA-SVR, the average training MSE and the average test MSE are 0.0103 and 0.0009. Therefore, in terms of the MSE criterion, the PCA-SVR provides a better forecast. For the MAD criterion, the results are the same. The computational results also show that the PCA-SVR is more stable than the PCA-ANN, because the PCA-SVR has a smaller standard deviation.

TABLE IV
10-FOLD CROSS-VALIDATION RESULT FOR PCA-ANN

	Training		Testing	
	MAD	MSE	MAD	MSE
Fold 1	0.0482	0.0126	0.0197	0.0041
Fold 2	0.0480	0.0125	0.0205	0.0005
Fold 3	0.0489	0.0127	0.0255	0.0007
Fold 4	0.0472	0.0118	0.0345	0.0015
Fold 5	0.4690	0.0127	0.0331	0.0012
Fold 6	0.0482	0.0119	0.0272	0.0009
Fold 7	0.0484	0.0127	0.0256	0.0008
Fold 8	0.0415	0.0106	0.0707	0.0149
Fold 9	0.0475	0.0119	0.0287	0.0012
Fold 10	0.0413	0.0118	0.0577	0.0070
avg.	0.0888	0.0121	0.0343	0.0033
std.	0.1336	0.0007	0.0167	0.0046

In order to determine the effect of PCA, an ANN and SVR were also used to learn from the data, without PCA. The computational results are shown in Tables VI and VII. The MSE for the training and testing results for the ANN are 0.0124 and 0.0138, respectively. For the SVR model, the average training and testing MSE values are 0.0116 and 0.009, respectively. These results show that the use of PCA for feature selection significantly improves the forecasting accuracy. The test results show that the PCA-SVR model outperforms the other models.

TABLE V
10-FOLD CROSS-VALIDATION RESULT FOR PCA-SVR

	Training		Testing	
	MAD	MSE	MAD	MSE
Fold 1	0.0554	0.0105	0.0253	0.0015
Fold 2	0.0567	0.0099	0.0146	0.0003
Fold 3	0.0556	0.0105	0.0153	0.0003
Fold 4	0.0563	0.0105	0.0179	0.0012
Fold 5	0.0557	0.0105	0.0171	0.0007
Fold 6	0.0561	0.0105	0.0156	0.0004
Fold 7	0.0657	0.0106	0.0096	0.0002
Fold 8	0.0587	0.0094	0.0176	0.0014
Fold 9	0.0602	0.0104	0.0252	0.0017
Fold 10	0.0577	0.0098	0.0156	0.0009
avg.	0.0578	0.0103	0.0174	0.0009
std.	0.0032	0.0004	0.0048	0.0006

TABLE VI
10-FOLD CROSS-VALIDATION RESULT FOR ANN

	Training		Testing	
	MAD	MSE	MAD	MSE
Fold 1	0.0856	0.0127	0.0542	0.0123
Fold 2	0.0849	0.0124	0.0639	0.0136
Fold 3	0.0812	0.0127	0.0965	0.0172
Fold 4	0.0849	0.0121	0.0645	0.0129
Fold 5	0.0781	0.0121	0.0758	0.0158
Fold 6	0.0836	0.0123	0.0698	0.0152
Fold 7	0.0887	0.0125	0.0632	0.0137
Fold 8	0.0823	0.0126	0.0811	0.0125
Fold 9	0.0849	0.0125	0.0625	0.0134
Fold 10	0.0894	0.0123	0.0587	0.0117
avg.	0.0844	0.0124	0.0690	0.0138
std.	0.0033	0.0002	0.0125	0.0017

TABLE VII
10-FOLD CROSS-VALIDATION RESULT FOR SVR

	Training		Testing	
	MAD	MSE	MAD	MSE
Fold 1	0.0898	0.0135	0.0151	0.0003
Fold 2	0.0895	0.0135	0.0161	0.0011
Fold 3	0.0906	0.0136	0.0210	0.0009
Fold 4	0.0900	0.0136	0.0147	0.0004
Fold 5	0.0893	0.0135	0.0094	0.0002
Fold 6	0.0884	0.0128	0.0672	0.0139
Fold 7	0.0908	0.0136	0.0434	0.0057
Fold 8	0.0334	0.0023	0.1650	0.0669
Fold 9	0.0748	0.0071	0.0177	0.0005
Fold 10	0.0868	0.0126	0.0137	0.0003
avg.	0.0823	0.0116	0.0383	0.0090
std.	0.0178	0.0038	0.0479	0.0208

IV. CONCLUSIONS

An echogram is a commonly used tool for the evaluation and rapid diagnosis of acute hepatitis in the ED. The parameters of the Child-Pugh classification are also easily accessed in the ED. This study constructs a cost forecast model that uses an echogram and the Child-Pugh classification data. The experimental results demonstrate that the PCA-SVR gives a better forecast than the PCA-ANN, in terms of the MSE and MAD. These findings show that an echogram and the Child-Pugh classification provide a practical tool for forecasting the cost for an acute hepatitis patient who is admitted to the ED. Future study might use more data, in order to improve the forecasting accuracy. Other learning algorithms could also be used to develop the forecasting model.

ACKNOWLEDGMENT

This study was financially supported by the National Science Council of the Taiwanese Government, under contract NSC102-2410-H-011-017-MY3. This support is appreciated.

REFERENCES

[1] W. M. Lee, "Hepatitis B virus infection," New England journal of medicine, vol. 337, pp. 1733-1745, 1997.

[2] S. R. Tujios and W. M. Lee, "New advances in chronic hepatitis B," Current Opinion in Gastroenterology, vol. 28, p. 193, 2012.
[3] C.-H. Chen, P.-M. Yang, G.-T. Huang, H.-S. Lee, J.-L. Sung, and J.-C. Sheu, "Estimation of Seroprevalence of Hepatitis B Virus and Hepatitis C Virus in Taiwan from a Large-scale Survey of Free Hepatitis Screening Participants," Journal of the Formosan Medical Association, vol. 106, pp. 148-155, // 2007.
[4] H. L. Y. Chan, S. W. C. Tsang, Y. Hui, N. W. Y. Leung, F. K. L. Chan, and J. J. Y. Sung, "The role of lamivudine and predictors of mortality in severe flare-up of chronic hepatitis B with jaundice," Journal of Viral Hepatitis, vol. 9, pp. 424-428, 2002.
[5] M. Kumar, S. Jain, B. Sharma, and S. Sarin, "Differentiating Acute Hepatitis B from the First Episode of Symptomatic Exacerbation of Chronic Hepatitis B," Digestive Diseases and Sciences, vol. 51, pp. 594-599, 2006/03/01 2006.
[6] E. Orenbuch-Harroch, L. Levy, and E. Ben-Chetrit, "Acute hepatitis B or exacerbation of chronic hepatitis B-that is the question," World journal of gastroenterology: WJG, vol. 14, p. 7133, 2008.
[7] G. L. Davis and J. H. Hoofnagle, "Reactivation of Chronic Type B Hepatitis Presenting as Acute Viral Hepatitis," Annals of Internal Medicine, vol. 102, pp. 762-765, 1985.
[8] G. Fattovich, L. Brollo, A. Alberti, G. Realdi, P. Pontisso, G. Giustina, et al., "Spontaneous reactivation of hepatitis B virus infection in patients with chronic type B hepatitis," Liver, vol. 10, pp. 141-146, 1990.
[9] I. S. Sheen, Y. F. Liaw, D. I. Tai, and C. M. Chu, "Hepatic decompensation associated with hepatitis B e antigen clearance in chronic type B hepatitis," Gastroenterology, vol. 89, pp. 732-735, 10/ 1985.
[10] P. S. Kamath, R. H. Wiesner, M. Malinchoc, W. Kremers, T. M. Therneau, C. L. Kosberg, et al., "A model to predict survival in patients with end-stage liver disease," Hepatology, vol. 33, pp. 464-470, 2001.
[11] M. Sheth, M. Riggs, and T. Patel, "Utility of the Mayo End-Stage Liver Disease (MELD) score in assessing prognosis of patients with alcoholic hepatitis," BMC Gastroenterology, vol. 2, p. 2, 2002.
[12] Q. Li, G.-Y. Yuan, K.-C. Tang, G.-W. Liu, R. Wang, and W.-K. Cao, "Prognostic factors for chronic severe hepatitis and construction of a prognostic model," Hepatobiliary Pancreat Dis Int, vol. 7, pp. 40-44, 2008.
[13] S. K. Sarin, A. Kumar, J. A. Almeida, Y. K. Chawla, S. T. Fan, H. Garg, et al., "Acute-on-chronic liver failure: consensus recommendations of the Asian Pacific Association for the study of the liver (APASL)," Hepatology international, vol. 3, pp. 269-282, 2009.
[14] S. Sarin, A. Kumar, and H. Garg, "Clinical profile of acute on chronic liver failure (ACLF) and predictors of mortality: a study of 64 patients," Hepatology, vol. 48, p. 450A, 2008.
[15] C. Zauner, R. Apsner, A. Kranz, L. Kramer, C. Madl, B. Schneider, et al., "Outcome prediction for patients with cirrhosis of the liver in a medical ICU: a comparison of the APACHE scores and liver-specific scoringsystems," Intensive care medicine, vol. 22, pp. 559-563, 1996.
[16] Y. C. Huang, "The application of data mining to explore association rules between metabolic syndrome and lifestyles," The HIM journal, 2013.
[17] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, pp. 433-459, 2010.
[18] G. P. Zhang, B. E. Patuwo, and M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-series forecasting," Computers & Operations Research, vol. 28, pp. 381-396, 4// 2001.
[19] A. S. Weigend and D. E. Rumelhart, Generalization through minimal networks with application to forecasting: Defense Technical Information Center, 1992.
[20] R. Fildes, K. Nikolopoulos, S. F. Crone, and A. A. Syntetos, "Forecasting and operational research: A review," Journal of the Operational Research Society, vol. 59, pp. 1150-1172, // 2008.
[21] C. J. Lu and Y. W. Wang, "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," International Journal of Production Economics, vol. 128, pp. 603-613, // 2010.
[22] M. H. Zheng, K. Q. Shi, X. F. Lin, D. D. Xiao, L. L. Chen, W. Y. Liu, et al., "A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network," Journal of viral hepatitis, 2012.
[23] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," Expert systems with applications, vol. 36, pp. 6164-6173, 2009.
[24] C.-J. Lu, T.-S. Lee, and C.-C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," Decision Support Systems, vol. 47, pp. 115-125, 2009.
[25] P.-F. Pai, S.-L. Yang, and P.-T. Chang, "Forecasting output of integrated circuit industry by support vector regression models with marriage honey-bees optimization algorithms," Expert Systems with Applications, vol. 36, pp. 10746-10751, 2009.
[26] L.-J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," Neural Networks, IEEE Transactions on, vol. 14, pp. 1506-1518, 2003.
[27] V. Vapnik, The nature of statistical learning theory: springer, 2000.
[28] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, pp. 1-27, 2011.

R. J. Kuo received the MS degree in Industrial and Manufacturing Systems Engineering from Iowa State University, Ames, IA, in 1990 and the Ph.D. degree in Industrial and Management Systems Engineering from the Pennsylvania State University, University Park, PA, in 1994. Currently, he is a Distinguished Professor in the Department of Industrial Management at National Taiwan University of Science and Technology, Taiwan. His research interests include architecture issues of computational intelligence and their applications to data mining, electronic business, customer relationship management, supply chain management and decision support systems.

W. C. Cheng received the M.B.A. degree in Business Administration from Fu Jen Catholic University, Taiwan, in 2010. Currently, he is the Ph.D. candidate in Industrial Management of National Taiwan University of Science and Technology, Taiwan. His research interests include architecture issue of computational intelligence and their applications to forecasting, production management and scheduling.

W. C. Lien got M.D. from National Taiwan University in 1998 and Ph.D. in Epidemiology and Preventive Medicine from National Taiwan University Hospital in 2012. Her research interest include emergency ultrasound, emergency medicine, epidemiology and biostatistics.

T. J. Yang received the degree of Master of Science from the College of Public Health, National Taiwan University in 2013. Currently, he is the Director and attending physician of emergency medicine in Sijhih Cathay General hospital. His research interests include emergency medicine and health management.