International Journal of Management, Economics and Social Sciences
Special Issue-International Conference on Medical and Health Informatics (ICMHI 2017)
2017, Vol. 6(S1), pp.293 – 306.
ISSN 2304 – 1366
http://www.ijmess.com

Data Mining Techniques for Forecasting the Medical Resource Consumption of Patients with Diabetic Nephropathy

Tian-Shyug Lee[1]
Wensheng Dai[2]
Bo-Lin Huang[1]
***Chi-Jie Lu[3]**

[1]Dept. of Business Administration, Fu Jen Catholic University, Taiwan
[2] International Monetary Institute and Finance, Policy Research Center, Renmin University of China, China
[3]Dept. of Industrial Engineering and Management, Chien Hsin University of Science and Technology, Taiwan

Diabetes has become an important public health issue in the twenty-first century, and dialysis treatment has become a large burden on the National Health Insurance of Taiwan. Diabetic nephropathy(DN) is the leading factor that determines whether patients with diabetes will require dialysis. Statistical data published by the Ministry of Health and Welfare in 2015 indicated that, second only to cancer, chronic kidney failure is the most prevalent disease treated by primary outpatient clinics. In addition, according to the National Health Insurance Administration Ministry of Health and Welfare, 6% of the national health insurance budget was spent to cover the dialysis treatment of ESRD patients. Therefore, in this study, we proposed and developed a forecasting model for the medical resource consumption of DN patients. We used multiple regression, stepwise regression, multivariate adaptive regression splines (MARS), support vector regression, and two-stage model (T-SVR). We used a combination of important variables screened out by stepwise regression and MARS to construct the T-SVR model. We screened out the important factors with a significant impact on medical consumption. We then identified the model with the best forecasting performance out of the five data mining techniques. Our results can aid the managers of medical institutions to properly and effectively allocate medical resources and control medical expenses.

Keywords: Medical resource consumption, diabetic nephropathy, data mining, multivariate adaptive regression splines, support vector regression

A report published by the World Health Organization in 2016 revealed that the number of patients with diabetes has tripled from that reported in1980 to 4.22 billion people in 2014. Furthermore, statistical data from the International Diabetes Association (IDF) in 2011 revealed that the total number of deaths from diabetes has increased to 4.8 million (Unwin et al., 2012). Approximately 20% to 40% of patients with type 1 diabetes mellitus and type 2 diabetes mellitus will develop diabetic nephropathy (DN) (Ito, 2010). The

Manuscript received May 20, 2017; revised September 8, 2017; accepted October 5, 2017.
*Corresponding author: jerrylu@uch.edu.tw

incidence and prevalence of dialysis in Taiwan have reached all-time highs: each year, more than 70,000 patients with major uremic injuries require dialysis treatment (Liu, 2015). Statistical data published by the Ministry of Health and Welfare in 2015 indicated that, second only to cancer, chronic kidney failure is the most prevalent disease treated by primary outpatient clinics. In addition, diabetes-induced uremiais the third most common reason for hospitalization.

In 2015, approximately 6 billion New Taiwan Dollars (NTD) was spent on inpatients, and approximately 34.2 billion was spent on outpatient dialysis (National Health Insurance Administration Ministry of Health and Welfare, 2016). However, according to Yang et al. (2001), each patient who requires dialysis treatment in Taiwan has spent an average of approximately 25,576 US dollars per year on medical care. Thus, patients who require dialysis treatment exert a significant burden on household expenses and national health insurance. Global data published by NCD-RisC, a non-communicable disease risk organization, revealed that over the past 35 years, the prevalence of male and female patients with diabetes in Taiwan has grown faster than that in Japan, Singapore, or Hong Kong. By 2025, the prevalence of male and female patients with diabetes in Taiwan is expected to become the third highest in East Asia, second only to that in China and South Korea. Chang et al. (2012) pointed out that the annual outpatient medical expenses and hospitalization costs associated with diabetes have increased by approximately 1.34 times over the past decade; in addition, the prevalence of patients with diabetes symptoms that manifest as vascular lesions annually increases at the rate of 9%-13%. Lin et al. (2005) reported that the medical resource consumption associated with diabetes treatment continues to increase as the number of patients with diabetes increases. The average cost of outpatient clinics in Taiwan is 1,485 NTD, which is higher than that of other national outpatient clinics, and the average cost of hospitalization for diabetes is 42,000 NTD. Taiwan's Central Health Insurance Board pointed out that the prevalence of diabetes in Taiwanese residents aged over 20 years old is 8%. Each year, approximately 1.4 million patients with diabetes exert a burden of 184 billion NTD on the national health insurance. (National Health Insurance Administration Ministry of Health and Welfare, 2011). Although the expenditure on diabetes accounts for only 4% of the national health insurance budget, it is associated with other medical expenses given the exclusion effect.

In recent years, Data mining techniques such as stepwise regression, multivariate adaptive regression splines (MARS), and support vector repression (SVR) have become popular data mining techniques for forecasting problems. Therefore, in this study, we use multiple regression, stepwise regression, MARS, and SVR to establish a prediction model for medical consumption by patients with DN. We also compared various data mining methodologies to determine the forecasting methodology with the best prediction accuracy. The selection of risk factors as predictors is a crucial issue for the construction of forecasting models because of the following reasons: An excessive number of prediction variables will extend the construction time and complexity of the forecasting model. However, a low number of prediction variables may not fully express the relationship between the target and predicted variables (Er& Beauchamp, 1988). Therefore, we used a combination of important variables screened out by stepwise regression and MARS to construct the model

The proposed model can enable the managers of medical institutions and central health bureaus to compare actual and predicted costs for the appropriate allocation of resources. The model can also help managers identify and analyze the factors that are significantly correlated with medical resource consumption. Finally, the model can help the Ministry of Health and Welfare, Central Health Insurance Agency to update its treatment equipment, treatment behavior, and organization to decrease the cost of medical care and improve the efficiency of health care and health insurance policies for patients with diabetes.

LITERATURE REVIEW

We used data mining technology to construct the forecasting model for the medical resource consumption of DN patients. We also compiled related domestic and international literature. We present the literature review in four sections: The first section presents a discussion of the literature on the resource consumption of patients with diabetes and receiving dialysis treatment. The second section provides a discussion of data mining technology. The third section presents relevant literature on MARS. Finally, the fourth section provides a brief review of the literature on SVR.

Medical Consumption of Diabetes and Dialysis

The 2016 IDF report showed that patients with diabetes consume approximately 22 trillion NTD of the global health care costs, accounting for 12% of the world's medical spending. Global costs associated with diabetes will reach 27 trillion NTD by 2040. Using a multiple regression model, Brandle et al. (2003) studied the direct medical costs of diabetes. They found that gender, ethnicity, hypertension, BMI index, hypoglycemic drug use, and epilepsy are associated with medical expenses for ischemic heart disease, cerebrovascular disease, and nephropathy.

According to the National Health Insurance Administration Ministry of Health and Welfare, 6% of the national health insurance budget, or 7.7% of total health insurance expenses, has been spent to cover the costs of dialysis treatment for ESRD patients at an average of approximately 0.62 million NTD per person per year. Smith et al. (2004) pointed out that in the US, patients with ESRD account for 0.1% of the total patient population but consume 6% of the total medical expenses. In Belgium, patients with ESRD contribute to 1.8% of the total cost of medical treatment in dialysis.

Data Mining

Regression analysis, exponential smoothing, and time series analysis are among the most commonly used data mining analytical tools. In recent years, however, an increasing number of studies have utilized quantitative tools for data mining and machine learning in artificial intelligence, such as MARS, clustering, artificial neural networks (ANN), genetic algorithms, SVM, and SVR (Song & Li, 2008; Lin et al, 2011; Lin & Lee, 2013).

For example, Kiran & Ravi (2008) utilized multiple regression, MARS, and BPN to investigate the accuracy and reliability of forecasting models. They reported that BPN and MARS have better accuracy and reliability than other models. Lee and Tang (2002) used MARS and BPN to integrate intellectual capital and financial ratio to construct an enterprise crisis diagnosis model. They reported that their model provided excellent classification results. Huang (2002) used association rules to explore the association between disease and microbes and reinforce potential disease relationships. Chen (2013) used ANN to construct a prediction model for thunderstorm in north Taiwan. The Canberra Airport has used an ANN-based model to predict cloud cover (Fabbian et al., 2006), and the New Chitose Airport used has multi-ANN to forecast fog (Nugroho et al., 2002).

MARS

MARS is a non-linear and non-parametric regression methodology that was first proposed by Friedman (1991). MARS is inspired by the recursive partitioning technique, governing classification, regression tree, and generalized additive modeling. MARS excels at finding optimal variable transformations and interactions and the complex data structure that often hides in high-dimensional data.

MARS essentially builds flexible models by fitting piecewise linear regressions; that is, the non-linearity of a model is approximated with separate linear regression slopes in distinct intervals of the independent variable space. Therefore, the slope of the regression line is allowed to change from one interval to the other as two 'knot' points are crossed. The variables to be used and the endpoints of the intervals for each variable are found through a fast but intensive search procedure. In addition to searching variables one by one, MARS also searches for interactions between variables, allowing any degree of interaction to be considered as long as the constructed model can fit the data well.

The general MARS function can be represented by following equation

$$f(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} [s_{km}(x_v(k,m) - t_{km})] \quad (1)$$

where a_0 and a_m are parameters, M is the number of basis functions, K_m is the number of knots, s_{km} takes the values of either 1 or − 1 and indicates the right/left sense of the associated step function, $x_v(k,m)$ is the label of the independent variable, and t_{km} indicates the knot location.

(Please refer to Friedman (1991) for more details regarding the complete model building process.)

SVR

SVR was developed for use in regression problems (Vapniket etal., 1997) and is mainly used to solve non-linear problems and for predicting various numerical fields (Yang and Liu, 1999; Koike and Takagi, 2004). SVR has a good generalization ability given that it mainly follows the principle of structural risk minimization.SVR parameters can be derived from the convex quadratic programming problem. Given that the only solution to SVR is the only solution, SVR avoids obtaining the solution from the minimal solution region (Chen, 2006).

Traditional regression obtains coefficients by minimizing the square error, which can be considered as empirical risk based on a loss function. Vapnik (1997) introduced the so-called ε -insensitivity loss function to SVR.

By simultaneously considering empirical and structural risks, the SVR model can be constructed to minimize the following programming:

$$\text{Min}: \frac{1}{2} z^T z + C \sum_i \left(\xi_i + \xi_i^* \right) \quad (2)$$

$$\text{Subject to} \begin{cases} y_i - z^T x_i - b \le \varepsilon + \xi_i \\ z^T x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

where i=1,…,n is the number of training data; $\left(\xi_i + \xi_i^* \right)$ is the empirical risk; $\frac{1}{2} w^T w$ is the structure risk preventing over-learning and lack of applied universality; and C is modifying coefficient representing the trade-off between empirical risk and structure risk. Equation (2) is a quadratic programming problem. After the selection of the appropriate modifying coefficient (C), width of band area (ε), and kernel function (K), the optimum of each parameter can be resolved though the Lagrange function. DeBoer et al. (2011) proposed that the radial basis function (RBF) is suitable for solving most forecasting problems.

Conceptual Framework

We used a retrospective cohort study design, and we retrieved data for the period of 2010 to the end of 2012 from the NHIRD. Figure 1 shows the conceptual framework of this study: The prediction variables of disease risk factors and medical resource consumption are introduced in step two. The dependent variable is the total cost of outpatient medical expenses, hospital medical expenses, and hospital days. First, we used multiple regression, stepwise regression, MARS, and SVR to construct the forecasting model. We then used stepwise regression and MARS to screen out significant variables.

Next, we combined the significant variables for input into SVR to construct the T-SVR prediction model. Finally, we compared the performances of the forecasting models we used in this study.

Dataset and forecasting variables

We used a retrospective generation study design. We monitored patients with DN who received dialysis

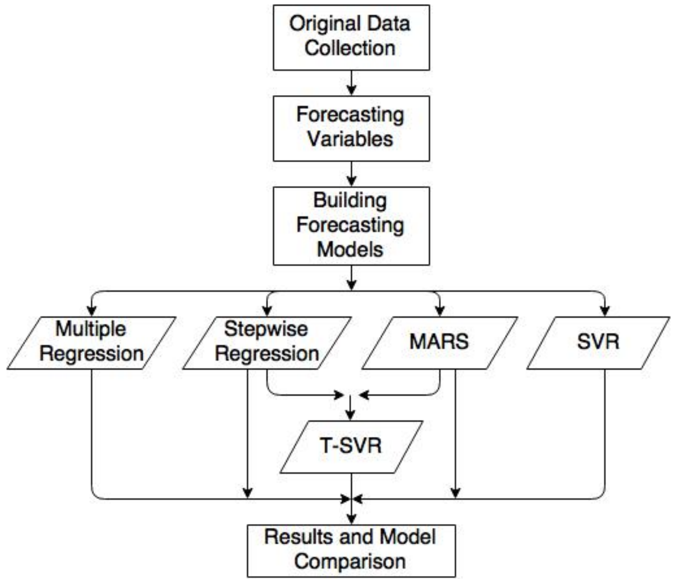

Figure 1. Conceptual Framework of the Forecasting Model

treatment in 2010 from the NHIRD. We continued to observe the patients and their collected medical expenses over a period of two years. We collected data in accordance with the definition of the International Classification of Diseases and Deaths (ICD-9-CM). Specifically, we collected data marked with a main diagnostic code or secondary diagnostic code containing the term“ 2504” .To ensure that the subjects were present throughout the entirety of the observation period, we excluded patients who had died, had been discharged, or had missing values.

We filtered four types of databases from the NHIRD: 1) Ambulatory care expenditures by visits (CD), 2) Details of ambulatory care orders (OO), 3) Inpatient expenditures by admissions (DD), and 4) Details of inpatient orders (DO). First, we collected cases that were reported in 2010 and had a main diagnostic code or secondary diagnostic code that contained “ 2504” from CD and DD. Next, we screened 519 kinds of hypoglycemic drugs from OO and DO the basis of their respective order code and drug number. Finally, we merged CD with OO and DD with DO.

For the next stage of our study, we collected cases that were reported over the period of2011 and 2012 in accordance with the methodology described above. We then individually merged the cases with those reported in 2010. Finally, to ensure that the subjects were present throughout the entirety of the study, we excluded patients who had died, were discharged, and who had missing values. We thus obtained our final research objects. The dependent variable is the total sum of outpatient medical expenses, hospital medical expenses, and hospital days.

We included 2597 subjects. We found that the total amount of money spent over the period of 2010 to 2012 totaled 285,304,619 NTD. The descriptive statistics of the data are presented in Table 1.

We then identified risk factors, such as patient characteristics and diabetic nephropathy, as prediction variables. Al-Rubeaan et al. (2014) showed that male gender and age are risk factors for the development and progression of DN. Zambrano-Galvánet et al. (2014) concluded that hypertension is a major predictive risk factor for DN.

Huang et al. (2014) reported that systolic blood pressure and DN are significantly and positively correlated. Mooradian (2009) found that hyperlipidemia is a major risk factor for the progression of DN. Moreover, they reported that excessive blood fat will increase the risk of renal vascular atherosclerosis and further cause the deterioration of renal function. (Toleren et al., 2014) found that the levels of high-density and low-density lipoproteins and fat are significant predictors of coronary heart disease and DN in patients with type 1 diabetes.

Almdal et al. (2004) and Jeerakathil et al. (2007) pointed out that patients with type II diabetes are two or three times at higher risk for stroke than the average person. Hägg et al. (2014) found that DN is a risk factor for ischemic and hemorrhagic stroke. Vijan and Hayward (2003) pointed out that if diabetic patients would strictly control their blood glucose levels, their risk for peripheral vascular disease or amputation and overall mortality rate within the next ten years would decrease by 0.5% and 0.8%, respectively. If patients would strictly control their blood pressure, their risk for peripheral-blood-vessel diseases or amputation would decrease by 4.3%. Thus, the overall mortality rate could be decreased by 4.3%.

The prediction variables used in this study are based on the above literatures and are provided in Table 2. The identified prediction factors are highly prevalent in patients with DN, indicating that patients with these factors are expected to consume more medical resources than those without.

The prediction results of the proposed MARS model were compared with those of the multiple regression and stepwise regression models without MARS. Prediction performance was evaluated with the following statistical metrics: root mean square error (RMSE), mean absolute difference (MAD), and mean absolute percentage error (MAPE).

Max	Min	Mean	SD
4203960	1778	109859	219373.7
Median	1st Qu.	3rd Qu.	Total(NTD)
68515	40139	108824	285304619

Table 1. Descriptive Statistics of the Subjects

Variables	Description(1=Yes, 0=No)
X1	Patient has hypertension
X2	Patient sex(1=Male,0=Female)
X3	Patient is over 65 years old
X4	Patient has dyslipidemia
X5	Patient has heart disease
X6	Patient has cerebrovascular disease
X7	Patient has peripheral vascular disease
X8	Patient has non-DN kidney disease

Table 2. Prediction Variables for the Construction of the Prediction Models

RESULTS

In this section, we report the empirical results obtained using the multiple regression, stepwise regression, MARS, SVR, and T-SVR models. In first section, we present the descriptive statistics of the research objects. We present the model results in the following section. In the last section, we compare the models on the basis of MAPE, RMSPE, MAD, and RMSE. The frequency of each prediction variables is listed in Table 3. In the table, a value of 1 indicates that the patient has the disease, whereas a value of 0 indicates that the object does not have the disease.

We ran all programming with R-studio software. We first input all the prediction variables for multiple regression. The results of multiple regression is listed in Table 4, which shows that the most significant prediction variables are kidney disease(X8) and dyslipidemia (X4).

Next, we developed the stepwise regression model. Its results are listed in Table 5, which shows that hypertension(X3), dyslipidemia (X4), cardiovascular disease(X5), and non-DN kidney disease (X8) are the variables with the highest relevance to DN incidence and medical resource consumption.

We used the R-package earth to develop the MARS model. The results and important predictor variables are listed in Table 6, which shows that the important predictor variables are non-DN kidney disease (X8), dyslipidemia (X4), hypertension disease(X1), and cerebrovascular disease (X6).

The parameters selected to develop the SVR model are $\varepsilon = 2^{-11}$ and $C = 2^{7}$, indicating that this combination is the best SVR parameter combination.

Finally, we constructed the T-SVR model on the basis of the variable selection results of stepwise regression and MARS. We used the stepwise regression model to screen out the most important variables: hypertension(X3), dyslipidemia (X4), cardiovascular disease(X5), and non-DN kidney disease (X8).The MARS model screened out hypertension (X1), dyslipidemia (X4), cerebrovascular disease(X6), and non-DN kidney disease(X8). Using the intersection concept, we found that dyslipidemia disease(X4) and non-DN kidney disease (X8) are two important variables. We then input these variables into the SVR model. The parameters chosen in T-SVR are $\varepsilon = 2^{-11}$ and $C = 2^{7}$.

The forecasting model results of the four models were computed and are listed in Table 7. Table 9 shows that out of all the models, the SVR model has the lowest MAE and MAPE, and T-SVR has the lowest RMSPE and RMSE. Thus, the SVR and T-SVR models may provide better forecasting results than the multiple regression, stepwise regression, and MARS models in terms of prediction error. To summarize, we identified dyslipidemia (X4) and non-DN kidney disease (X8) as important variables.

Variables	Yes(1)	No(0)
X1	1890	707
X2	1405	1192
X3	374	2223
X4	1637	960
X5	745	1852
X6	523	2074
X7	98	2499
X8	1019	1578

Table 3. Frequency of Each Prediction Variable

Coefficients	Estimate	Std. Error	*t*-value	Pr(>|t|)
(Intercept)	106710	12983	8.219	3.58e−16 ***
SEX (X1)	−2848	9340	−0.305	0.76047
AGE.65 (X2)	−19283	13686	−1.409	0.15901
Hypertension (X3)	−21409	10695	−2.002	0.04544 *
Dyslipidemia (X4)	−28971	9690	−2.990	0.00283 **
Cardiovascular disease (X5)	25903	10680	2.425	0.01538 *
Cerebrovascular disease (X6)	8408	11999	0.701	0.48356
Peripheral vascular disease (X7)	23924	25281	0.946	0.34410
Non-DN kidney disease (X8)	79613	9804	8.120	7.91e−16 ***

Note: *p<0.1;**p<0.05;***p<0.01

Table 4. Results of the Multiple Regression Model

Coefficients	Estimate	Std. Error	*t*-value	Pr(>|t|)
(Intercept)	105948	11846	8.944	< 2e−16 ***
Hypertension(X3)	−21768	10632	−2.047	0.04074 *
Dyslipidemia(X4)	−29348	9673	−3.034	0.00244 **
Cardiovascular disease (X5)	24964	10448	2.389	0.01697 *
Non-DN kidneydisease (X8)	78887	9663	8.164	5.56e−16 ***

Note: *p<0.1;**p<0.05;***p<0.01

Table 5. Results of the Stepwise Regression Model

Variable selection results	
Variable name	**Relative importance (%)**
Non-DN kidney disease(X8)	100.0
Dyslipidemia disease (X4)	39.6
Hypertension disease(X1)	20.2
Cerebrovascular disease (X6)	20.1
MARS prediction function: f(x) = 104230.62−20639.40* X1−25085.51* X4+24213.79 *X6+80497.86 *X8	

Table 6. Important Prediction Variables of the MARS Model

Methods	RMSPE(%)	MAE	RMSE	MAPE(%)
MARS	2.8236	84956	219791	1.477
Multiple regression	2.9548	85842	215736	1.504
Stepwise regression	2.9536	85726	215865	1.504
SVR	1.7984	**68745**	218676	**0.892**
T-SVR model	**1.7435**	69612	**214738**	0.901

Table 7. Comparison of Model Results

CONCLUSION

International governments are attempting to decrease the consumption of medical resources. However, patients with diabetes, a serious chronic disease associated with a high death rate, exert a substantial medical burden on their families and countries. Among other Asian countries, Taiwan has the highest prevalence of diabetic patients who require dialysis, which results in a considerable annual medical burden. To predict the medical resource consumption of patients with diabetes in Taiwan, we constructed a forecasting model using data retrieved from a health insurance research database. We also identified the factors that could increase medical expenses. We selected hypertension (X3), dyslipidemia (X4), cardiovascular disease (X5), cerebrovascular disease (X6), and non-DN kidney disease (X8) as important variables.

We then compared the results of the multiple regression, stepwise regression, MARS, SVR, and T-SVR models. The empirical results showed that the SVR and T-SVR models have the best performance among the tested models. This result is consistent with that of previous studies on data mining and machine learning, thus confirming that emerging data mining techniques have excellent predictive performance and provide various tools for data prediction. Dyslipidemia (X4) and non-DN kidney disease (X8) are important predictive variables for the medical resource consumption of patients with DN. We successfully developed a methodology for the construction of a forecasting model for medical resource consumption, compared data mining techniques, and identified the important variables related with DN medical expenses. We hope that our study will act as a reference for other researchers who wish to use similar methods or data from health insurance databases for predictive experiments.

REFERENCES

Abraham, A., Steinberg, D., & Philip, N. S. (2001). Rainfall forecasting using soft computing models and multivariate adaptive regression splines. IEEE SMC Transactions: Special, 1-12.

Brandle M, Zhou H.H., Smith, Barbara R.K.(2003). The direct medical cost of type 2 diabetes. Diabetes Care ; Aug; 26, 8

Chang C.C., & Lin, C.J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(27), 1-27.

Chang C.S., (2009). Evaluating the performance of different wavelet basis functions and levels in forecasting financial time series using MARS and SVR. Unpublished doctoral dissertation, National Taipei technology university Department of Business and management.

Chang T.J., Jiang Y.D, Chang C.H. (2012). Accountability, utilization and providers for diabetes management in Taiwan, 2000-2009: An analysis of the National Health Insurance database. Journal Formosa Medical Association. Nov;111(11): 605-16

Chen, T.M., & Tung M.S., (2013). Neural network analysis for hot thunderstorm occurrence probability and its application in northern Taiwan. Ministry of Transportation and Communication.

Christiansen, J. S., Gammelgaard, J., Tronier, B., Svendsen, P. A., &Parving, H. H. (1982). Kidney function and size in diabetics before and during initial insulin treatment. Kidney int, 21(5), 683-688.

DeBoer, I. H., Rue, T. C., Cleary, P. A., Lachin, J. M., Molitch, M. E., Steffes, M. W., Sun, W., Zinman, B., Brunzell, J. D., White, N. H., Danis, R. P., Davis, M. D., Hainsworth, D., Hubbard, L. D., & Nathan, D. M. (2011). Long -term renal outcomes of patients with type 1 diabetes mellitus and microalbuminuria: An analysis of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications cohort. Archives of internal medicine, 171(5), 412 -420.

Fabbian, D., de Dear, R., Lellyett, S.,(2006) : Application of Artificial Neural Network Forecast to Predict Fog at Canberra International Airport. Wea. Forecasting, 22, 372– 381.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). Annals of Statistics, 19, 1-141.

Friedman, J. H., and C.B. Roosen. (1995). An introduction to multivariate adaptive regression splines. Statistical Methods in Medical Research, 4, 197 -217.

Hamzacebi, C., Akay, D. & Kutay, F. (2009).Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert Systems with Applications,36, 3839-3844.

He, K., Lai, K. K. & Yen, J. (2010). A hybrid slantlet denoising least squares support vector regression model for exchange rate prediction , Procedia Computer Science, 1, 2397 -2405.

Hong, W. C., Dong, Y., Chen, L. Y. & Lai, C. Y. (2010).Taiwanese 3G mobile phone demand forecasting by SVR with hybrid evolutionary algorithms, Expert Systems with Applications, 37, 4452-4462.

Ilan, A., Min, Q., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. Journal of Retailing and Consumer Services, 8, 147-156.

Ito, S. (2010). Treatment strategies according to the stage of diabetic nephropathy. Nihon Rinsho, 68(suppl 9), 46 5-471.

Jim O., Jie, H., Ying Hong, P. &Qiushi, R. (2011).Electrical evoked potentials prediction model in visual prostheses based on support vector regression with multiple weights" , Applied Soft Computing In Press.

Johnson, R.J., Kivlighn, S.D.,Kim, Y.G., Suga, S., Fogo, A.B.(1999). Reappraisal of the pathogenesis and consequences of hyperuricemia in hypertension, cardiovascular disease, and renal disease. American Journal Kidney Disease 33: 225-234.

Kanas, A., & Yannopoulos, A. (2001). Comparing linear and nonlinear forecasts for stock returns. International Review of Economics and Finance, 10, 383-398.

Kang, D.H., & Nakagawa, T. (2005). Uric Acid and Chronic Renal Disease. Possible implication of hyperuricemia on progression of renal disease. Seminars in Nephropathy 25: 43-49 .

Kiran, N. J., & Ravi, V.(2008).Software reliability prediction by soft computing techniques. The Journal of Systems and Software, 81, 576 -583.

Koike, A., & Takagi, T. (2004). Prediction of protein -protein interaction sites using support vector machines. Protein Engineering Design & Selection,17(2),165 -173.

Lee T.S. & Tang H.C., (2004). Incorporating financial ratios and intellectual capital in business failure predictions using artificial neural networks and multivariate adaptive regression splines. Journal of Information Management. 11(2), P.161-189.

Leernders, M. R., Fearon, H. E., Flynn, A. E., & Johnson, P. F. (2002). Purchasing and supply management (12th ed.),New York: McGraw-Hill.

Leernders, M. R., Fearon, H. E., Flynn, A. E., & Johnson, P. F. (2002). Purchasing and supply management (12th ed.),New York: McGraw-Hill.

Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. Lancet, 379(9811), 165 -180.

Lin T, Chou P, Lai MS, Tsai ST, Tai TY.(2005). Direct costs-of-illness of patients with diabetes mellitus in Taiwan. Diabetes Research Clinical Practice; 54 supply 1:S43-6.

Lin, C. J., & Lee, T. S., (2013). Tourism demand forecasting: Econometric model based on multivariate adaptive regression splines, artificial neural network and support vector regression, Advances in Management and Applied Economics, 3(6), 1-18.

Lin, C. J., Chen, H. F., & Lee, T. S. (2011). Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines: Evidence from Taiwan. International Journal of Business Administration, 2(2), 14-24.

Liu, C. K.(2015). Applying Data Mining Techniques for Constructing Disease Risk Factor Analysis Model– The Case for Diabetic Nephropathy and Dialysis. Unpublished doctoral dissertation, Department of Business Administration Fu Jen Catholic University.

Mogensen, C. E., & Christensen, C. K. (1984). Predicting diabetic nephropathy in insulin-dependent patients. New England Journal of Medicine, 311(2), 89-93.

Mogensen, C. E., Christensen, C. K., & Vittinghus, E. (1983). The stages in diabetic renal disease: With emphasis on the stage of incipient diabetic nephropathy. Diabetes, 32(Suppl 2), 64-78.

Multistep-ahead time series prediction, Advances in Knowledge Discovery and Data Mining, Vol. 3918 of the series Lecture Notes in Computer Science, pp 765-774.

N Unwin, L Guariguata, D Whiting, C Weil. (2012). Complementary approaches to estimation of the global burden of diabetes. THE LENCET, V(379), No. 9825, p1487– 1488.

Nugroho, A. ,Kuroyanagi, S., Iwata, A. (2002) : A Solution for Imbalanced Training Sets Problem by CombNET-II and Its Application on Fog Forecasting. IEICE Trans InfSyst, E85-D.7, 1165-1174.

P. F. Pai & C. S. Lin,(2006) “ Using support vector machines in forecasting production values of machinery industry in Taiwan,” International Journal of Advanced Manufacturing Technology, vol. 27, pp.205-210.

Schölkopf, & A. J. Smola,(2002). Learning with kernel. Cambridge, MA：MIT Press.

Sharda, R. & Patil, R.B. J IntellManuf (1992), Connectionist approach to time series prediction: an empirical test. Journal of Intelligent Manufacturing, October 1992, Volume 3, Issue 5, pp 317– 323.

Shelbaya, S., Amer, H., Seddik, S., Allah, A. A., Sabry, I. M., Mohamed, T., & El-Mosely, M. (2012). Study of the role of interleukin-6 and highly sensitive C-reactive protein in diabetic nephropathy in type 1 diabetic patients. Eur Rev Med Pharmacol Sci, 16(2), 176-182.

Singhal, D. &Swarup, K. S. (2011).Electricity price forecasting using artificial neural networks. Electrical Power and Energy Systems, 33, 550 -555.

Smith DH, Gullion CM, Nichols G, Keith DS, Brown JB. (2004), Cost of medical care for chronic kidney disease and comorbidity among enrollees in a large HMO population. J Am Soc Nephrol;15:1300-6

Song, H., & Li, G., (2008). Tourism demand modeling and forecasting – A review of recent research. Tourism Management, 29, 203-220.

Steinberg, D., Bernard, B., Phillip, C., & Kerry, M.(1999). MARS user guide. San Diego, CA：Salford Systems.

Sung Y.H., (2013). Medical utilization and affecting factors of DM nephropathy and DM retinopathy. Unpublished doctoral dissertation, Kaohsiung medical university Department of Healthcare Administration and Medical Informatics,Kaohsiung.

Vapnik, V. N. (2000). The Nature of Statistical Learning Theory, Springer, New York, NY.

Vapnik, V.N., Golowich S. &Smola, A. (1997). Support vector method for function approximation, regression estimation and signal processing. In: Mozer, M., Jordan, M., and Petsche, T. (Eds.), Advance in Neural information processing system 9. Cambridge, MA: MIT Press, 281 -287.

Worku, D., Hamza, L., &Woldemichael, K. (2010). Patterns of diabetic complications at jimma university specialized hospital, southwest ethiopia. Ethiopian journal of health sciences, 20(1), 33-39.

Yang, W. C., Hwang, S. J., Chiang, S. S., Chen, H. F., & Tsai, S. T. (2001). The impact of diabetes on economic costs in dialysis patients: experiences in Taiwan. Diabetes research and clinical practice, 54(Suppl 1), 47-54.

Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, Berkeley, California, United States, 42-49.

E-source

Bureau of National Insurance (2011). Challenge 2025 Diabetes Medical International Forum. http://www.nhi.gov.tw/epaper/ItemDetail.aspx?DataID=2667&IsWebData=0&ItemTypeID=7&PapersID=227&PicID(2016/12/27)

Common Wealth (2016). Newly diabetic patients younger, take lifes every. http://wwzw.cw.com.tw/article/article.action?id=5076781 (2017/1/10).

Health benefits of chronic kidney disease in the early stage of health insurance, (2015), http://www.nhi.gov.tw/Resource/webdata/26254_1_Early%20CKD 計畫.pdf (2016/12/21)

National Health Insurance Administration Ministry of Health and Welfare (2016), https://www.nhi.gov.tw/Resource/webdata/26258_1_2015_發生率_盛行率_糖尿病透析_置放全球資訊網-1050714-2 版.pdf (2016/12/19).

National Health Insurance Administration Ministry of Health and Welfare (2016), Statistic abstract of citizen health insurance in 2015, http://dep.mohw.gov.tw/DOS/cp-2808-9853-113.html. (2017/1/21)

National Health Insurance Administration Ministry of Health and Welfare (2016), Citizen health insurance medical treatment given expense total consultation reference index- 2016. http://www.mohw.gov.tw/dl-3056-6e641965-4504-4033-b94a-a97ff905cef9.html(2017/1/21).