

doi:10.3233/SHTI210671

The Health Check-Up Data-Analysis for Risk Assessment of Chronic Kidney Disease (CKD) in Taiwan

Ming-Shu CHEN[a,1], Mao-Jhen JHOU[b], Chi-Jie LU[b,c] and Chung-Chih HUNG[d]
[a] *Department of Healthcare Administration and Department of Nursing, Asia Eastern University of Science and Technology (AEUST), New Taipei City, Taiwan*
[b] *Graduate Institute of Business Administration, Fu Jen Catholic University, Taiwan*
[c] *Department of Information Management, Fu Jen Catholic University, Taiwan*
[d] *Department of Laboratory Medicine, Taipei Hospital, Ministry of Health and Welfare, New Taipei City, Taiwan*

Abstract. Early detection of chronic kidney disease （CKD） for high-risk population adults is very important. It has a common risk factor and causal relationship with chronic diseases such as diabetes, hypertension and cardiovascular disease etc. The results of this study provide that for early high-risk factors detection in CKD healthy population can be used by home care to recommend adjuvant treatment.

Keywords. Chronic kidney disease, data analytics, deep learning, early detection

1. Introduction

With economic growth and changes in diet and lifestyle habits, people are increasingly being diagnosed with chronic kidney disease (CKD). Coupled with the improvement of medical standards, the prevalence of CKD in countries around the world has increased year by year. The prevalence rate of CKD worldwide is 13.4%, and the prevalence rate of women is higher than men (Hill et al., 2016) [1]. The CKD can pass the estimated glomerular filtration rate (e-GFR) is used for the diagnosis; It is associated with gender, age, and serum creatinine value [2]. CKD is the predecessor of kidney failure or end-stage kidney disease that causes the National Health Insurance to spend nearly 50 billion high medical resources each year in Taiwan. The USA (USRDS) 2016 Annual Report published the latest global uremia rankings. The data show that the prevalence of dialysis patients in Taiwan and the annual incidence of dialysis is the highest in the world. In Taiwan, the prevalence of CKD in become more and more important, because it is highly correlated with other diseases such as diabetes, hypertension and cardiovascular disease etc. [3]. Early high-risk factors detection in CKD healthy population can be used by home care to recommend adjuvant treatment.

[1] Corresponding Author, Min-Shu Chen, Department of Healthcare Administration and Department of Nursing, Asia Eastern University of Science and Technology (AEUST), No. 58, Sec. 2, Sichuan Road, Pan-Chiao District, New Taipei City 22061, Taiwan, ROC; E-mail: tree1013@gmail.com.

2. Methods

2.1. E-GFR

The formula of this study was as follows (1).

IDMS-MDRD equation for glomerular filtration rate estimation in Chinese patients: [2]

$$175 \times Cr^{-1.154} \times Age^{-0.203} \times (0.742, \text{ if female}) \times (1.212, \text{ if African American})$$

; [Cr.]: serum creatinine value. …(1)

2.2. Logistic Regression (LR) and Random Forests (RF)

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist [4]. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [5].

3. Results

This study analyzed datasets from the MJ Group [6]—a major health screening center in Taiwan—from 2010 to 2015. The MJ Health-Check-Up-Based Population Database (MJPD) collected from four MJ clinics provides periodic health examinations to the center's approximately 71,000 members and contains 201,087 cases. We keep the latest health examination data and discard other data just take 65,734 records for analysis, the analyze variables including Body Fat; Waist circumference; Blood pressure; Fasting serum glucose; GOT; GPT; r-GT; BUN; UA; Triglyceride; T-Cho; HDL; LDL etc. The main dependent variance is CKD classification which is calculated based on e-GFR, the descriptive statistics and data mining analysis results were as the following.

Table 1. The descriptive statistics for each variances of the MJ Health-Check-Up-data based.

Variances (N = 65734)	Mean [Xi]	± SD	Median [Xi]	Skew [Xi]
Body Fat	26.833	6.832	26.1	0.759
Waist circumference	78.601	10.436	78.0	0.458
Systolic Blood pressure (SBP)	115.524	17.246	114.0	0.716
Diastolic Blood pressure (DBP)	73.422	11.098	73.0	0.446
Serum glucose	102.411	19.11	99.0	5.658
GOT	24.094	13.526	22.0	15.013
GPT	28.661	25.239	22.0	9.769
r-GT	28.567	36.653	20.0	17.552
BUN	13.478	3.716	13.1	2.162
UA	5.733	1.536	5.6	0.504
Triglyceride	116.819	92.239	94.0	8.18
T-Cho	196.279	34.445	194.0	0.614
HDL	58.382	14.723	56.0	0.858
LDL	118.079	32.311	116.0	0.512

Table 2. The results of the SVMs method for the all variances of the CKD classification analysis.

Method	Accuracy	Sensitivity	Specificity	Auc	Precision	Recall	F1_Socre	Kappa
LR[1]	0.6405*	0.6279	0.6867	0.7113	0.8795	0.6279	0.7327	0.2282
RF[2]	0.6343*	0.6193	0.6888	0.7094	0.8788	0.6193	0.7266	0.2216

*Representative model accuracy: LR[1] (Logistic Regression); RF[2] (Random forest)

4. Conclusions

This study will use the annual health check database as the research material. By analyzing the huge amount of data for many years, we will use logistic regression, and random forest analysis tools. The common risk factors and major potential risk variables of the two are to realize the implementation and application of the adult sub-health management prediction system of CKD in the development of health data-driven model.

References

[1] Hill NR, Fatoba ST, Oke JL et al. (2016). Global prevalence of chronic kidney disease–a systematic review and meta-analysis. PloS one, 11(7), e0158765.

[2] Ma YC, Zuo L, Zhang CL, Wang M, Wang RF, Wang HY. (2006). Comparison of 99mTc-DTPA renal dynamic imaging with modified MDRD equation for glomerular filtration rate estimation in Chinese patients in different stages of chronic kidney disease. Nephrology Dialysis Transplantation, 22(2), 417-423.

[3] Naghibi M, Mojahedi MJ, Jarrahi L et al. (2015). Prevalence of chronic kidney disease and its risk factors in Gonabad, Iran. Iranian journal of kidney diseases, 9(6), 449.

[4] Tolles J, Meurer WJ. (2016). Logistic regression: Relating patient characteristics to outcomes. Jama, 316(5), 533-534.

[5] Barandiaran I. (1998). The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell, 20(8), 1-22.

[6] MJ Group: The health screening center in Taiwan. https://www.mjlife.com/index.aspx?lang=eng&fn=mj