Applying the Temporal Abstraction Technique to the Prediction of Chronic Kidney Disease Progression

Li-Chen Cheng[1] · Ya-Han Hu[2] · Shr-Han Chiou[2]

Abstract Chronic kidney disease (CKD) has attracted considerable attention in the public health domain in recent years. Researchers have exerted considerable effort in attempting to identify critical factors that may affect the deterioration of CKD. In clinical practice, the physical conditions of CKD patients are regularly recorded. The data of CKD patients are recorded as a high-dimensional time-series. Therefore, how to analyze these time-series data for identifying the factors affecting CKD deterioration becomes an interesting topic. This study aims at developing prediction models for stage 4 CKD patients to determine whether their eGFR level decreased to less than 15 ml/min/1.73m^2 (end-stage renal disease, ESRD) 6 months after collecting their final laboratory test information by evaluating time-related features. A total of 463 CKD patients collected from January 2004 to December 2013 at one of the biggest dialysis centers in southern Taiwan were included in the experimental evaluation. We integrated the temporal abstraction (TA) technique with data mining methods to develop CKD progression prediction models. Specifically, the TA technique was used to extract vital features (TA-related features) from high-dimensional time-series data, after which several data mining techniques, including C4.5, classification and regression tree (CART), support vector machine, and adaptive boosting (AdaBoost), were applied to develop CKD progression prediction models. The results revealed that incorporating temporal information into the prediction models increased the efficiency of the models. The AdaBoost+CART model exhibited the most accurate prediction among the constructed models (Accuracy: 0.662, Sensitivity: 0.620, Specificity: 0.704, and AUC: 0.715). A number of TA-related features were found to be associated with the deterioration of renal function. These features can provide further clinical information to explain the progression of CKD. TA-related features extracted by long-term tracking of changes in laboratory test values can enable early diagnosis of ESRD. The developed models using these features can facilitate medical personnel in making clinical decisions to provide appropriate diagnoses and improved care quality to patients with CKD.

Introduction

Chronic kidney disease (CKD) is a disease in which patients progressively lose kidney function. CKD has drawn considerable attention in the public health domain in recent years. Approximately 10% of people worldwide suffer from this disease [1]. CKD causes considerable morbidity and markedly influences patients' quality of life; therefore, health professionals worldwide are attempting to derive means of preventing its occurrence [2, 3].

The early symptoms of CKD are typically nonspecific; they are similar to those of many other diseases. CKD is often not diagnosed until more serious symptoms such as edema and hematuria occur [4]. When the kidney function drops to

---

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Ya-Han Hu
yahan.hu@mis.ccu.edu.tw

[1] Department of Computer Science and Information Management, Soochow University, Taipei 100, Taiwan

[2] Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi 621, Taiwan

5% below normal, the body cannot maintain normal metabolism and the patient enters end-stage renal disease (ESRD). To maintain life and health, patients with ESRD need a regular lifelong dialysis support or a kidney transplant [5, 6].

A report by the United States Renal Data System (2012) revealed that the prevalence (2584 per million population) and incidence (361 per million population) of ESRD in Taiwan are among the highest in the world. According to the Taiwan Society of Nephrology (2009), the number of people currently undergoing regular treatment for ESRD in Taiwan is approximately 60 000, with an increase of 8000 people each year [7]. In the United States, the treatment and management costs for ESRD patients is nearly $26.8 billion, constituting 5.9% of all medical expenses. ESRD consumes a considerable portion of medical resources [8, 9].

Clinicians and researchers have exerted considerable effort in attempting to identify critical factors that may prevent the progression of CKD to ESRD [10]. However, the risk factors leading to CKD deterioration are generally highly complex and difficult to treat [4]. Previous studies have revealed that aging, a family history of kidney diseases, the use of analgesics, low socioeconomic status, and an unhealthy lifestyle are risk factors possibly resulting in the deterioration of kidney function [11, 12]. In addition, the progression of CKD may be expedited when patients have other chronic diseases (eg, hypertension, diabetes, high cholesterol, or cardiovascular disease) [5, 13, 14]. Because numerous potential risk factors can cause a decline in kidney function, addressing the question of how to accurately predict CKD deterioration based on all possible risk factors is challenging. A highly reliable prediction model can facilitate the prediction of the deterioration of patients with CKD, thus enabling medical staff to provide early treatment to inhibit or prevent the disease from progressing to a more severe stage (ie, ESRD).

In clinical practice, the physical conditions of patients with CKD are regularly recorded and managed. At each regular examination, medical personnel record detailed information relevant to the physical status of a patient with CKD such as biochemical blood test records, lifestyle and drug use habits, and patient education records. This information is recorded and stored in an electronic medical record (EMR) system and can be used to assist medical staff when evaluating a patient's condition [15–18]. Until a hospital closes a case or declares a patient with CKD deceased, the nephrology department continues to track the patient's status. Therefore, the medical records accumulate in the EMR system over time, resulting in high-dimensional time-series data.

Most studies have applied only cross-sectional data and statistical methods to construct prediction models of CKD progression. Although cross-sectional data are accurate predictors in prediction models of CKD progression, information regarding the variation of the variables over time is ignored. For example, blood pressure (BP) can be measured several times during a specific period (ie, a patient's BP can be recorded as time-series data). One of the easiest methods of presenting BP variations is through mean and standard deviation. However, BP variations also carry essential information regarding a patient's condition, and this information cannot be presented through mean and standard deviation. These pieces of information can reflect changes in the physical status of patients with CKD, which may also be major factors for predicting CKD progression. According to our review of the literature, no study has considered detailed information regarding the variation of the time-series data on the problem of predicting CKD deterioration. Therefore, the question of how to extract time-related features from high-dimensional time-series data has become a major obstacle to the prediction of CKD progression.

In the current study, effective CKD progression prediction models were developed by evaluating time-related features. Specifically, two main tasks had to be accomplished. The first task entailed extracting time-related features accurately reflecting the variation of variables for the time-series data of patients with CKD. The second task involved developing models for predicting CKD progression based on multiple machine learning techniques.

To extract features regarding the variations in the status of patients with CKD, the temporal abstraction (TA) technique was used to extract time-related variables from the EMRs of patients with CKD. The TA framework is a knowledge-based framework proposed by Shahar in 1997 [19]. TA can be used to define different change types in time-series data, such as trends, statuses, or other complex time-related attributes. By incorporating the clinical knowledge of experts, users can define their own TA variables. The values of these variables can be obtained by applying a series of data preprocessing tasks to patients' time-series data derived from the EMR system.

In this study, complete records of patients with CKD were collected from the EMR system in dialysis centers in southern Taiwan from January 2004 to December 2013. The data included the patient's basic information, medical history records, laboratory test records, and care and nutrition tracking records. The variable selection and extraction were performed by consulting the relevant literature and domain experts (ie, nephrologists). Multiple supervised learning techniques were used to develop the prediction models including C4.5, classification and regression tree (CART) [20], and support vector machine (SVM) [21]. Adaptive boosting (AdaBoost) was also used to improve the predictive performance of the models [22]. The results show that the inclusion of TA-related variables enhances the performance of the prediction models.

The rest of this paper is organized as follows. In "Background overview and literature review" section, we review related works. Subsequently, we define the problem of mining the records of patients with CKD and explain the development of the models. Next, experimental results are presented. Finally, we present our conclusions.

Background overview and literature review

Definition of CKD and its stages

The kidney, an essential organ for the body's metabolism, can be affected by various diseases or injuries including chronic diseases, recurrent infections, and urinary tract obstruction. CKD is characterized by a gradual loss of kidney function; specifically, this disease occurs when the kidney is damaged and unable to appropriately maintain its original function [2] for more than 3 months.

Early symptoms of CKD are generally difficult to detect; these symptoms do not appear until kidney function drops 30% below normal. Currently, the most commonly applied standards are those announced by the Kidney Disease Outcome Quality Initiative (K/DOQI), which uses the glomerular filtration rate (GFR) to determine the severity of the progression of CKD [23]. Patients who have a GFR of less than 60 mL/min/1.73 m2 for more than 3 months can be diagnosed with CKD. Such patients might also exhibit other signs of kidney damage including abnormal kidney ultrasound or biopsy results. Clinicians can provide specific treatment, such as offering psychological support and mental adjustment, according to a patient's detected stage.

Clinically, GFR cannot be determined through direct measurement; it is typically diagnosed through laboratory test indicators to determine an estimated GFR (eGFR). Experts worldwide have developed various equations for determining GFR, tailored to various countries and races. Most experts use the Serum Creatinine (Scr) measure along with other clinical options such as age, gender, and race. Modification of Diet in Renal Disease (MDRD), or the revised MDRD-S, is the most commonly used method, followed by the Cockcroft–Gault (CG) calculator. However, several scholars have found that for people with a high GFR, MDRD can cause systematic underestimation errors, resulting in reduced accuracy. To increase the accuracy of GFR, Levery et al. [24] developed a new GFR equation in 2009, called the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI). This equation can be effectively used to estimate normal kidney function. Currently, the National Kidney Foundation (NKF) in the United States has adopted the CKD-EPI equation.

Although medical staff can evaluate the decline in kidney function in patients by evaluating the eGFR, they cannot further understand related information regarding the progression of CKD if only the eGFR is clinically evaluated. Therefore, how to construct a reliable prediction model and identify CKD patients who are at potential risk of deterioration has become an essential question in medical research [8].

Factors and research affecting the progression of CKD

Studies have revealed that the factors leading to CKD progression are extremely complicated. According to the information from Taiwan Society of Nephrology (2007), three causes of dialysis exist: diabetes (43.2%), glomerulonephritis (25.1%), and hypertension (8.3%). Controlling these top three factors is critical.

Laboratory tests evaluate several items, such as eGFR, proteinuria, calcium, phosphorus, sodium, and albumin (ALB), which are used to create prediction models [25, 26]. Lorenzo et al. [25] determined that proteinuria is a risk factor leading to decreased renal function. Hallan et al. [26] revealed that gender, low amounts of exercise, diabetes, a higher body mass index (BMI) and waist circumference, high BP, the use of antihypertensive medications in therapies, low high-density lipoprotein cholesterol, triglycerides, and high blood sugar are all risk factors considerably increasing disease progression. Othman et al. [27] reported that in multiple regression analysis, changes in BMI and younger age were independently associated with the rate of decline in GFR. Go et al. [28] showed that a decrease in GFR led to a deterioration of kidney function in patients, which could increase the chance of cardiovascular-related diseases, thus necessitating hospitalization for treatment. In a previous study, indoxyl sulfate and p-cresylsulfate were added according to the suggestion of Wu and colleagues (2011) to provide further clinical information for predicting the risk of kidney function deterioration in patients with CKD [29]. Perotte et al. [30] used a time-series model (i.e., a variant of Kalman filter) to generate time-related features from both laboratory test results and patient notes. With the considerations to these features, the proposed risk stratification model for CKD progression prediction outperforms other baseline models.

Survival prediction models have been frequently used for analysis when predicting the risk of CKD progression, with two of the most common being the Kaplan–Meier model and the Cox model. However, these two prediction models are based on incomplete information, implying that one can obtain only limited information from the analysis results.

In this study, we mainly collected patients' demographic data, physical examination results, and information about complications and personal habits as well as other relevant information as study variables. Unlike considering only the feature set generated by the data mining algorithm in Perotte et al. [30], our approach was driven by the clinical experts; that is, the clinical experts can define their own TA variables from the time-series of each laboratory examination item.

Data mining in kidney research

In recent years, data mining (DM) has been used for analyzing CKD, ESRD, and hemodialysis (HD) information as well as

other related laboratory tests and clinical data [16, 31–33]. Chou et al. [34] proposed a method that entails incorporating rough set theory (RST) into the analysis process for evaluating and verifying a patient's blood dialysis data. In this method, RST is used for rule extraction from granulated data, and C4.5 and artificial neural networks (ANN) are subsequently used to compare the accuracy of prediction models before and after screening. The results revealed that the decision attribute of the Kt/V ratio can help doctors reduce diagnosis time.

Altintas et al. [35] used the DM technique to find the risk factors causing death in patients undergoing hemodialysis and also examined the interaction styles of these factors. Yeh et al. [6] combined the TA technique with a data mining model to analyze laboratory test data from hemodialysis patients. In addition to identifying the patterns of hospital stays as found in past studies, they observed clinically significant timing trends that had previously gone unnoticed. Belazzi et al. [36] developed a hemostatic system through Matlab by using information, such as BP and duration of dialysis, of patients subjected to dialysis treatments thrice a week. The results indicated that the developed systems can significantly improve the hemodialysis clearance rate, facilitated achieving the right amount of dialysis for patients, and improved the quality of dialysis care. The application of DM technology in kidney-related research demonstrates the importance of medically related data analysis and discussion with medical experts in related fields and has facilitated the process of ensuring the accuracy of the analysis results.

Temporal abstraction

TA focus on the extraction of qualitative aspects of time series based on rules that are defined by clinical experts [19, 33]. TA can be used to define different change types in time series data, such as trends, statuses, or other complex time-related attributes. Generally speaking, temporal abstraction can be classified as basic TA and complex TA [36]. Basic TA is often used to detect numerical or symbolic time series, and the episode found is presented through a qualitative approach [6, 36]. In particular, two types of TA were extracted: trend TAs, which capture the increasing, decreasing, or stationary patterns in a numerical time series, and state TAs, which detect qualitative patterns corresponding to low, high, and normal values in a numerical or symbolic time series. Complex TA is used to analyze the temporal relationship between interval series. The temporal relationships investigated correspond to the 13 temporal operators [19], which include BEFORE, FINISHES, OVERLAPS, MEETS, STARTS, DURING, their corresponding inverse relations, and EQUALS. Recently, several researches integrated TA with data mining techniques to develop a decision support system to predict hospitalization of hemodialysis (HD) patients in order to reducing hospitalization rate [6, 19, 36].

Method

The proposed CKD progression prediction models comprise the following modules (Fig. 1): data processing and filtering, TA, and classification modules.

Data collection and preprocessing

Data from 2066 patients with CKD were collected from January 2004 to December 2013 at one of the biggest kidney health center in a metropolitan hospital in southern Taiwan. The center provides services such as CKD health education, CKD case management and dialysis. The St. Martin De Porres Hospital Institutional Review Board approved the study protocol.

This study focused on finding TA-related variables to predict the progression of CKD. In practice, collecting the data from patients with CKD at the early stages is difficult because the symptoms regarding CKD normally do not appear until patients have reached at least stage 3 or 4 CKD. Therefore, most of patients with CKD at the early stages do not have routine body examinations and to extract the values of TA-related variables from these patients is infeasible.

Based on aforementioned reasons, the experimental data in this study included only patients with stage 4 CKD and having no dialysis treatment during 1-year study period (i.e., 538 patients). For each selected patients, one-year-long laboratory data were collected after stage 4 CKD onset to determine the values of TA-related variables. Because stage 4 CKD patients usually have a routine body examination every 3 months, a 1-year study period allows us to collect 4 body examination
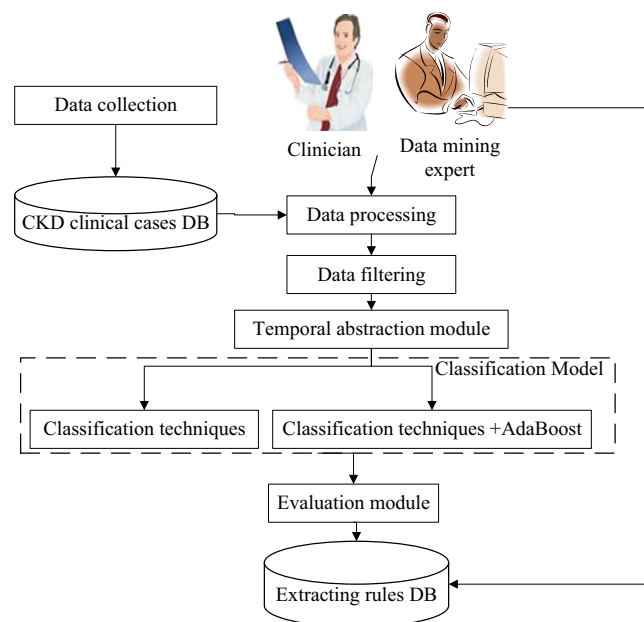
Fig. 1 The proposed framework

results for each patient. Patients with the number of body examinations less than 4 were deleted from the experimental data set. If the patients have more than 4 body examinations, we only consider the first 4 results in generating the values of TA-related variables. As a result, a total of 463 patients were included in the experimental evaluation. Among them, 132 cases deteriorated to ESRD 6 months after the last date of their one-year-long laboratory data and 331 cases remained at stage 4.

The dependent variable represented whether patients with stage 4 CKD exhibited signs of ESRD (ie, stage 5 CKD) 6 months after collecting their final laboratory test information. According to the National Kidney Foundation, ESRD is defined as the GFR decreases to less than 15 ml/min/1.73m^2. Because GFR cannot be determined by direct measurement and clinicians usually adopt biochemical test indicators to determine an estimated GFR (eGFR), this study determine CKD stage based on eGFR.

Table 1 lists all the selected patient information. In the EMR system, detailed information regarding each case, including demographic data, drug use, history, habits, laboratory tests, body checkup, and health education assessment, was recorded. Categorical variables included items such as gender, whether the patient smoked or used herbal medicine in medical history, and health education. These variables were constant over time and could be encoded directly. Continuous variables were primarily the results, such as body checkup and laboratory tests; the values of these variables were recorded each quarter year.

In addition, the clinicians determined the standard values of every laboratory test item. It is because most of the laboratory test values of CKD patients differ from those of a person without CKD. The identification of abnormal laboratory test values may be the important factors in determining the progression of CKD patients. In this study, two clinicians first examined the descriptive statistics of the patients with stage 4 CKD. Specifically, the mean, and both the one and the two standard deviations away from mean were considered as the thresholds of laboratory test values. After that, both two clinicians manually checked each threshold and altered the value according to their expertise.

Patients were divided into the following groups according to age: young (19–30 y), young adult (31–44 y), middle-aged (45–64 y), and elderly (older than 65 y). Most patients (333, or 71.92%) were aged 65 years or older. The elderly group contained 162 (48.65%) men and 171 (51.35%) women. In 88 (19.01%) cases, patients had taken traditional herbal medicine before; moreover, 25 cases (5.40%) had high cholesterol, 84 (18.14%) smoked, 34 (7.34%) drank, and 5 (1.08%) chewed betel nut.

TA module

Figure 2 illustrates the TA variables conversion process flow. First, the kidney and DM experts cooperatively defined the TA rules. After the data processing step, the variables were converted into TA format according to TA rules. Specifically, the patients' laboratory data were inputted into both the basic and complex TA algorithms to retrieve the values of TA variables. Finally, the records containing TA variables were stored in a database.

TA mechanisms can extract various types of changes, including state and trend, from time-series data. During this step, we consulted with kidney experts to verify the patient's laboratory test record and define the TA rules. We used two categories of TA variables: basic TA and complex TA variables [19, 36].

Basic TA variables can be identified from numerical or symbolic time-series. In particular, two types of TA variables were extracted: state TA variables, which detect qualitative patterns corresponding to low, high, and normal values in a numerical or symbolic time-series, and trend TA variables, which capture the increasing, decreasing, or stationary patterns in a numerical time-series.

Table 1 The selected variables for CKD patients

Categoty	Variable	Type
Demographic	Gender	Categorical
	Age	Numerical
Drug use	Chinese herbs, Analgesics	Categorical
History	Diabetes, Hypertension, Cardiovascular disease, High cholesterol, Anemia, Gout	Categorical
Habit	Smoke, Alcohol, Betel quid chewing, Exercise	Categorical
Laboratory	Creatinine, BUN, HB, Hct, WBC, RBC, Ca, P, Na, K, Cl, Mg, Uric Acid, Cholesterol, Triglyceride, Albumin, iPTH, Total Protein, Sugar [AC], Urine Protein, HDL-Cholesterol, LDL-Cholesterol, Fe, Ferritin, ACRatio, Ca × P	Numerical
Body checkup	Height, Weight, Waist, Hips, Systolic pressure (SBP), Diastolic pressure (DBP)	Numerical
Health education assessment	Health education, Cognitive appraisal, Behavior appraisal	Categorical
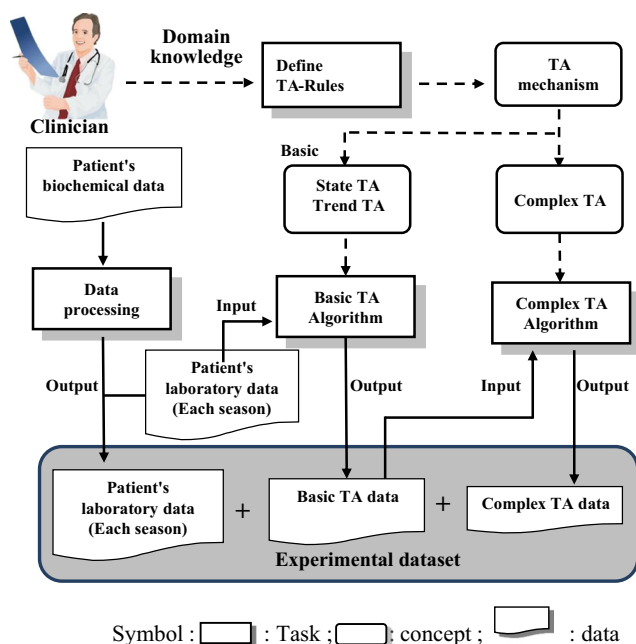
Fig. 2 The process of temporal abstraction

Each laboratory test item generated state TA variables according to the TA rules. However, the numbers for each type of test item differed. The laboratory test items in the TA rules are defined in Table 2. A total of seven possible types of status exist: extremely high (XH), high (H), higher than normal (N/H), normal (N), lower than normal (N/L), low (L), and extremely low (XL). For example, based on the threshold value in Table 2, the seven state TAs of Creatinine are XH (Creatinine ≥ 14.98), H ($14.98 >$ Creatinine ≥ 12.94), N ($12.94 >$ Creatinine ≥ 10.91), N/H ($10.91 >$ Creatinine ≥ 8.87), N/L ($8.87 >$ Creatinine ≥ 6.84), L ($6.84 >$ Creatinine ≥ 4.81), and XL (Creatinine < 4.81); the five state TAs of ALB are H (ALB ≥ 6.97), N/H ($6.97 >$ ALB ≥ 5.77), N ($5.77 >$ ALB ≥ 4.57), N/L ($4.57 >$ ALB ≥ 2.17), and L (ALB < 2.17).

To retrieve the value of a state TA variable, the average value of a laboratory test item at two adjacent time points need to be calculated and then mapped into the TA rules in Table 2. Trend TA variables were defined by checking the trend of the values between two adjacent time points (quarter year). The types of trend TA were divided into S (steady), I (increasing), and D (decreasing). Combing both the state and trend TA finally yields the value of the basic TA variable.

For example, Table 3 shows the examination data for ALB for each quarter-year. The average of ALB for seasons 1 and 2 is 5.54; therefore, the state TA of ALB for season 1–2 is N. Similarly, the average of ALB for season 2–3 is 4.24, indicating that the state TA of ALB for seasons 2–3 is N/L. In addition, because the ALB values of season 1–2 is 4.83 (N) and 6.25 (N/H), respectively, the value of the ALB from season 1 to season 2 is increasing (N → N/H), that is, the trend TA of ALB for season 1–2 is I. Therefore, the basic TA of ALB for season 1–2 is N-I.

Complex TA variable is used to analyze the temporal relationship between two adjacent basic TA variables. The complex TA was derived through basic TA information for every test item by using a time calculation approach. The symbol > was also used in the current study to connect two basic TA values and then combine them to determine the complex TA. For example, in Table 3, because the basic TAs of ALB for seasons 1–2, 2–3, and 3–4 are N-I, N/L-D, and L-D, respectively, three complex TAs can be obtained: N-I > N/L-D, N-I > L-D, and N/L-D > L-D.

Classification techniques

We adopted several common supervised learning techniques, including C4.5, CART, and SVM, to construct the prediction models [20, 21]. In addition, AdaBoost, proposed by Freund and Schapire [22], was integrated to enhance the prediction performance of the proposed models. AdaBoost can self-adjust the prediction models during the calculation process, thereby reducing the classification error rate by updating the weights of samples. AdaBoost combines multiple learning algorithms to construct a strong classifier with high accuracy, thereby improving the classification effectiveness [37–39].

Experimental evaluation

Evaluation design and performance measurement

We used Weka open-source machine learning software (www.cs.waikato.ac.nz/ml/weka/) to construct the prediction models based on C4.5, CART, and SVM. AdaBoost was further used to enhance effectiveness of the prediction models.

The data set containing the records of patients with CKD was imbalanced. Specifically, 132 cases deteriorated to ESRD 6 months later (Y) and 331 cases remained at stage 4 (N) in the experimental data set. We resampled the data set to avoid the class imbalance problem. A total of 132 cases from the 331 patients with stage 4 CKD were randomly selected and integrated with all 132 ESRD patients into one balanced data set (ie, the ratio of Y/N is 1:1). In addition, certain useful cases in the patients with stage 4 CKD may not have been selected during the resampling process, resulting in the loss of valuable information for classification. Therefore, a random resampling technique was applied 30 times to construct 30 data sets.

This study used creatinine as the baseline model. Specifically, creatinine model included the newest creatinine value in one-year-long laboratory data, age, and gender at stage 4 CKD onset, as independent variables. To determine whether the consideration of TA modules improved the accuracy of the prediction models, two experimental data sets, with and without the TA module, were generated for further comparison. We used 10-fold cross validation to verify the

Table 2 Threshold values of biochemical test items for TA transformation

Item	Unit	Number of threshold	Threshold values					
Creatinine	mg/dl	6	14.98(XH)	12.94(H)	10.91(N/H)	8.87(N/L)	6.84(L)	4.81(XL)
BUN	mg/dl	6	48.20(XH)	44.41(H)	42.51(N/H)	34.91(N/L)	33.02(L)	29.22(XL)
HB	g/dl	6	16.99(XH)	14.47(H)	13.22(N/H)	8.19(N/L)	6.93(L)	4.42(XL)
Hct	%	6	38.32(XH)	35.85(H)	34.61(N/H)	29.65(N/L)	28.42(L)	25.94(XL)
WBC	×1000/μL	6	13.17(XH)	11.67(H)	10.17(N/H)	4.16(N/L)	2.66(L)	1.16(XL)
RBC	×1000/μL	6	11.48(XH)	10.18(H)	8.87(N/H)	6.26(N/L)	4.95(L)	2.34(XL)
Ca	mg/dl	6	14.29(XH)	12.11(H)	11.02(N/H)	6.65(N/L)	5.56(L)	3.38(XL)
P	mg/dl	4	8.06(H)	6.77(N/H)	1.64(N/L)	0.35(L)		
Na	meq/l	6	143.89(XH)	141.78(H)	139.67(N/H)	135.44(N/L)	133.33(L)	131.21(XL)
K	mg/dl	6	9.11(XH)	7.90(H)	6.69(N/H)	5.48(N/L)	3.06(L)	1.86(XL)
Uric_Acid (M)	mg/dl	6	12.66(XH)	11.34(H)	10.01(N/H)	4.72(N/L)	3.39(L)	2.07(XL)
Uric_Acid (F)	mg/dl	6	12.30(XH)	10.98(H)	9.66(N/H)	4.40(N/L)	3.08(L)	1.76(XL)
Cholesterol	mg/dl	6	180.40(XH)	177.81(H)	176.51(N/H)	171.33(N/L)	170.04(L)	167.45(XL)
Triglyceride	mg/dl	6	136.90(XH)	133.31(H)	131.52(N/H)	124.34(N/L)	122.54(L)	118.96(XL)
ALB	gm/dl	5	6.97(H)	5.77(N/H)	4.57(N/L)	2.17(L)		
Urine_Protein	mg/dl	2	79.52(N/H)	64.09(N/L)				
iPTH	pg/ml	2	87.21(N/H)	76.33(N/L)				
Total_Protein	mg/dl	2	9.00(N/H)	4.22(N/L)				
Sugar	mg/dl	6	134.86(XH)	131.98(H)	130.54(N/H)	124.79(N/L)	123.35(L)	120.48(XL)
Fe	μg/dl	6	76.60(XH)	73.41(H)	71.81(N/H)	65.41(N/L)	63.82(L)	60.62(XL)
Ferritin	ng/dl	2	229.40(N/H)	219.30(N/L)				
HDL_Cholesterol	mg/dl	6	48.88(XH)	45.94(H)	44.46(N/H)	38.57(N/L)	37.10(L)	34.15(XL)
LDL_Cholesterol	mg/dl	6	99.68(XH)	96.76(H)	95.30(N/H)	89.46(N/L)	88.00(L)	85.09(XL)
ACRatio	μg/mg	2	109.46(N/H)	56.20(N/L)				
Ca × P	mg^2/dl^2	6	43.67(XH)	41.13(H)	39.85(N/H)	34.76(N/L)	33.49(L)	30.94(XL)
BMI	kg/m^2	5	35(SO)	30(MDO)	27(MIO)	24(OW)	18.5(UW)	
SBP	mmHg	5	180(XH)	160(VH)	140(H)	120(N/H)	90(L)	
DBP	mmHg	5	110(XH)	100(VH)	90(H)	80(N/H)	60(L)	

reliability of the proposed prediction models. The predictive performance was measured by evaluating the accuracy, sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic curves of each classification model [40].

Assume that CKD_{TP} represents the number of patients who entered ESRD and are accurately identified as ESRD patients; CKD_{TN} represents the number of patients who did not enter ESRD and are accurately identified as stage 4 CKD patients; CKD_{FP} represents the number of patients who did not enter

Table 3 Example of TA transformation

		Season 1	Season 2	Season 3	Season 4
	Albumin (ALB)	ALB_1: 4.83 (N)	ALB_2: 6.25 (N/H)	ALB_3: 2.23 (N/L)	ALB_4: 1.2 (L)
	Avg. value of two adjacent items	5.54 (N)	4.24 (N/L)	1.715 (L)	
	State TA	N	N/L	L	
	Trend TA	I (N → N/H)	D (N/H → N/L)	D (N/L → L)	
	Basic TA	ALB_BasicTA_12: N-I			
		ALB_BasicTA_23: N/L-D			
		ALB_BasicTA_34: L-D			
	Complex TA	ALB_ComplexTA_1: N-I > N/L-D			
		ALB_ComplexTA _2: N-I > L-D			
		ALB_ComplexTA _3: N/L-D > L-D			

ESRD but are inaccurately identified as ESRD patients; CKD_{FN} represents the number of patients who entered ESRD but are inaccurately identified as stage 4 CKD patients. The definitions are formulated as follows:

$$CKD_{\text{Sensitivity}} = \frac{CKD_{TP}}{(CKD_{TP} + CKD_{FN})}, \qquad (1)$$

$$CKD_{\text{Specificity}} = \frac{CKD_{TN}}{(CKD_{TN} + CKD_{FP})}. \qquad (2)$$

The predictive ability of each model was evaluated by calculating the AUC. In general, if the predictive accuracy of the proposed model is perfect, its AUC is nearly 1. If the AUC is between 0.8 and 0.9, then the model has high predictive accuracy. If its AUC is between 0.7 and 0.8, then the proposed model is acceptable. We compared the pros and cons of each prediction model according to accuracy, sensitivity, specificity, and AUC and then selected the most appropriate model for predicting the course of disease progression in patients with CKD.

Results

Table 4 lists the experimental results. Here, we report only the mean and standard deviation of the results of the 30 generated data sets to simplify the explanation. Regarding the single classifiers, the CART model demonstrated the highest AUC, followed by SVM; C4.5 registered the lowest AUC. The AUC of the CART model was between 0.6 and 0.7, indicating that an acceptable predictive ability. After the integration of

AdaBoost, the AUC and specificity of the CART model were still the highest among the classifiers. However, SVM had the highest accuracy and C4.5 demonstrated the highest sensitivity. After the integration of AdaBoost, the predictive power of CART improved significantly. The CART model exhibited the highest performance among all single classifiers, and the integration of AdaBoost significantly enhanced the performance of this model. Overall, the CART model integrated with AdaBoost (AdaBoost + CART) demonstrated the most accurate prediction regarding the deterioration of patients with CKD. In addition, considering TA-related features in the constructions of CART and SVM classifiers (both in single classifier form and in the integration with AdaBoost) increased the prediction performance, which indicates that TA-related features greatly affect the deterioration of CKD. In addition, we further compared the prediction performance between the best proposed model (i.e., AdaBoost + CART techniques with considerations of TA-related features) and the baseline model. As shown in Table 5, the results showed that our model outperform the creatinine model. The AUC of the creatinine model is close to 0.5, meaning that its prediction ability is almost similar to random guessing.

We adopted four feature selection modules provided by WEKA to determine the factors influencing the deterioration of CKD. Specifically, the average of the rank results obtained from Gini, ChiSquared, InfoGain, and GainRatio modules were used to rank the importance of factors. The top 25 factors found in this study are summarized in Table 6.

According to the assessment results, gender was the most critical factor affecting the deterioration of CKD among the first 25 variables that exerted the greatest impact. Age was also a major factor at stage 3. Gender and aging are related

Table 4 Evaluation results of the prediction models

(Mean ± std. dev.)			C4.5	CART	SVM
With TA	Single	Accuracy	0.500 ± 0.015	0.606 ± 0.041	0.652 ± 0.037
		Sensitivity	0.738 ± 0.100	0.603 ± 0.052	0.614 ± 0.042
		Specificity	0.250 ± 0.084	0.619 ± 0.063	0.692 ± 0.047
		AUC	0.490 ± 0.075	0.653 ± 0.033	0.652 ± 0.037
	Adaboost	Accuracy	0.502 ± 0.018	0.662 ± 0.031	0.667 ± 0.035
		Sensitivity	0.742 ± 0.105	0.620 ± 0.038	0.630 ± 0.037
		Specificity	0.261 ± 0.119	0.704 ± 0.036	0.700 ± 0.041
		AUC	0.495 ± 0.062	0.715 ± 0.035	0.707 ± 0.036
Without TA	Single	Accuracy	0.563 ± 0.032	0.604 ± 0.036	0.634 ± 0.037
		Sensitivity	0.555 ± 0.103	0.574 ± 0.057	0.606 ± 0.039
		Specificity	0.571 ± 0.110	0.636 ± 0.052	0.662 ± 0.044
		AUC	0.580 ± 0.043	0.635 ± 0.038	0.634 ± 0.037
	Adaboost	Accuracy	0.601 ± 0.037	0.657 ± 0.030	0.643 ± 0.032
		Sensitivity	0.584 ± 0.043	0.608 ± 0.041	0.600 ± 0.039
		Specificity	0.618 ± 0.054	0.706 ± 0.034	0.687 ± 0.043
		AUC	0.642 ± 0.039	0.705 ± 0.036	0.672 ± 0.031

Table 5 Evaluation results of the proposed and the baseline models using AdaBoost + CART technique

Method	Accuracy (mean ± std. dev.)	Sensitivity (mean ± std. dev.)	Specificity (mean ± std. dev.)	AUC (mean ± std. dev.)
With TA	0.662 ± 0.031	0.620 ± 0.038	0.704 ± 0.036	0.715 ± 0.035
Creatinine (baseline)	0.500 ± 0.032	0.482 ± 0.064	0.518 ± 0.085	0.502 ± 0.031

to a decline in kidney function [25, 26, 29, 41–44]. For the CKD patients, the variables exerting the greatest impact are consistent with those reported in previous studies, indicating that renal function [3, 26, 45, 46], BP [25, 26], and blood Hct [47, 48], are all vital indicators.

According to the Taiwan Society of Nephrology's Dialysis Information (2007), diabetes (43.2%), glomerulonephritis (25.1%), and hypertension (8.3%) are the main reasons for patients eventually requiring dialysis. This report illustrated the importance of controlling hyperglycemia, hyperlipidemia, and hypertension in preventing CKD. This is consistent with the features we have found in Table 6.

This study found that kidney function-related factors have considerable impact on CKD progression. In clinical examinations of patients with CKD, two evaluation indices, Scr and

Table 6 Average Results of the importance of study variables obtained from Gini, ChiSquared, InfoGain, and GainRatio modules

Rank	Variables
1	Gender
2	Scr_4
3	Diabetes
4	Age (in years)
5	BUN_4
6	**BUN_BasicTA_34**
7	Hct_4
8	Betel Nut
9	Hct_3
10	Scr_3
11	cognitive assessment
12	**BUN_BasicTA_23**
13	BUN_3
14	Hypertension
15	Hyperlipidemia
16	**BUN_ComplexTA_3**
17	Current drinking
18	**SBP_ComplexTA_1**
19	**Hct_BasicTA_34**
20	Heart disease
21	Current smoking
22	**BMI_BasicTA_23**
23	SBP_3
24	**SBP_ComplexTA_3**
25	**BUN_ComplexTA_1**

Bold font indicates TA-related variables

blood urea nitrogen (BUN), are periodically tested. Calculating the patients' eGFR through the Scr can provide a clearer understanding of the disease progression and deterioration. Recording changes in kidney function over a long period can facilitate the process of determining the severity and progression of CKD (eg, Scr_4, BUN_4, BUN_Basic_34, Scr_3, BUN_Basic_23, BUN_3, BUN_Complex_1). We also found factors related to chronic disease or lifestyle oriented including diabetes, hypertension, high cholesterol, cardiovascular disease, chewing betel nut, drinking, and smoking. Previous studies have also indicated that a significant correlation exists between gout and a decline in kidney function, rendering it a potential risk factor for the progression of the disease [49, 50].

Most of the relevant Hct factors have time information variables. Showing changes in Hct by recording it over a long period can identify the severity and progression of the disease. Previous studies have proven that anemia may also exacerbate the symptoms of kidney disease [51–53]. Hct, HB, and other indicators have also been reported to be correlated with the progression of CKD [47, 48]. The experimental result showed that the cognitive assessment value was also essential. Strengthening patient's awareness of CKD through measures such as health education, lifestyle changes, diet, and dialysis can improve the condition of patients and their quality of life [10, 54, 55]. Therefore, health education should be included as a reference factor when determining CKD progression.

Discussion

Patients with stage 4 CKD have a low understanding of their disease status, indicating that medical institutions must enhance patient care and improve health education at each stage [10, 54, 55]. Patients should reach a certain level of awareness and proactively improve their self-care, which can help them to control the disease and mitigate the deterioration of kidney function. Such health education can be provided to a patient or someone close to said patient, such as a spouse, child, relative, or friend. Educational courses regarding diet, medication, regular exercise, and dialysis treatment can also be arranged. At the fourth stage, most patients begin dialysis treatment, and a kidney transplant will be required if no attention is paid to the disease progression.

Because the CART model exhibited the highest performance among all prediction models, we consulted with experts and identified the meaningful rules from the CART model. Tables 7 shows the occurrences and accuracy rate of each

Table 7   Clinically meaningful rules for the progression of CKD to ESRD

No	Rules	Occurrence	Accuracy
1.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)| (N-S > N-S)| (N-D > N-I)| (H-S > H-S)| (H-D > N-D)| (N-I > N/H-S)| (XL-S > L-I) **AND** Hct_4 = (Normal) **AND** BMI_4 = (Normal)	5	93%
2.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)| (N-S > N-S)| (N-D > N-I)| (H-S > H-S)| (H-D > N-D)| (N-I > N/H-S)| (XL-S > L-I) **AND** Albumin_4 = (Normal) **AND** Diabetes = YES	9	86%
3.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)|(N-S > N-S)|(N-D > N-I)|(H-S > H-S)| (H-D > N-D)|(N-I > N/H-S)|(XL-S > L-I) **AND** Regular exercise = NO **AND** Smoking = YES	8	84%
4.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)|(N-S > N-S)|(N-D > N-I)|(H-S > H-S)| (H-D > N-D)|(N-I > N/H-S)|(XL-S > L-I) **AND** Regular exercise = NO **AND** Diabetes = YES **AND** Hypertension = YES	2	100%
5.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)|(N-S > N-S)|(N-D > N-I)|(H-S > H-S)| (H-D > N-D)|(N-I > N/H-S)|(XL-S > L-I) **AND** Regular exercise = NO **AND** Diabetes = YES **AND** Hypertension = YES	13	89%
6.	Age = (elderly) **AND** BUN_ComplexTA_3 = (XH-D > XH-I)|(N-S > N-S)|(N-D > N-I)|(H-S > H-S)| (H-D > N-D)|(N-I > N/H-S)|(XL-S > L-I) **AND** Regular exercise = NO **AND** BUN_BasicTA_23 = (XH-D)|(XH-I)	11	100%
7.	Age = (elderly) **AND** Heart disease = YES **AND** CKD awareness = (total, partial)	3	75%
8.	Age = (elderly) **AND** Heart disease = YES **AND** CKD awareness = (total) **AND** Smoking = YES	4	100%

rule. These rules can be divided into IF and THEN categories. Each rule results in the progression of stage 4 CKD patients to the ESRD after 6 months. As shown in Table 7, the TA-related variables exert a major effect; almost every major rule contains complex TA patterns, implying that long-term tracking of changes in laboratory test values can enable early diagnosis and facilitate the process of mitigating the deterioration of CKD.

Rules 1–6 show that for elderly patients with stage 4 CKD, no deterioration in BUN indicators occurs among the laboratory tests. However, if the patient has a poor lifestyle, does not exercise, smokes, or has complications or other illnesses such as heart disease or diabetes, deterioration to ESRD occurs easily.

Conclusion

CKD has become a major concern and has attracted the attention of experts in the public health domain. To determine the progression of CKD, several studies have established prediction models to discover the potential factors affecting patient deterioration. Therefore, how to construct a reliable prediction model; identify patients that have potential risks of deteriorating to ESRD; and execute early detection, prevention, and disposal are critical topics in medical research.

This study proposes effective CKD progression prediction models that involve considering time-related features. The aim of this study was to establish a reliable prediction model for patients at different stages of CKD to enable medical personnel to understand and forecast the progression of CKD symptoms. This study also demonstrates that time-related features are clinically meaningful and can be used to mitigate the deterioration of CKD.

According to the experimental results, the accuracy of the prediction models can be improved effectively when integrated with the TA module. The prediction model established using the AdaBoost + CART model has a higher prediction efficiency compared with the other models. To determine the relevant factors that inhibit the deterioration of CKD in patients, this study evaluated the importance of the variables. Factors that may affect the progression of the disease are gender, age, other chronic diseases (ie, diabetes, hypertension, high cholesterol, cardiovascular disease), renal function (Scr, BUN), BP, blood (Hct), BMI, and health education (awareness evaluation); These factors have already been proven to correlate with the deterioration of renal function in past studies. In addition, the time-related features detected through consultation with the clinical experts can provide deeper knowledge to explain the progression of CKD.

According to the experimental results, we can create novel decision support systems to help medical personnel to understand the patient status in the future. Moreover, to prevent deterioration to the next stage after 6 months, doctors and health education staff can gather additional information regarding the variations in the status of CKD patients for developing a more appropriate treatment program.

This study has several limitations. The experimental data collected the records from a single hospital in southern Taiwan. Only one-year-long laboratory data (i.e., 4 times a year) were considered in developing prediction models. Therefore, the generalizability of the results to other CKD patients from different regions or countries is limited. Future studies should consider expanding sample size for confirming the validity of the model, such as including clinical cases from other hospitals in experimental evaluations. In addition, this study considered only a few relevant variables to construct CKD progression prediction models, but the proposed models

obtain fair prediction accuracy only. In practice, the factors affecting CKD deterioration are extremely complex; therefore, to improve the prediction accuracy, we suggest that future studies could analyze other potential factors. Possibilities include environmental factors, education level, socioeconomic status, diet, medication, nutrition assessment projects, and geographic region.

Compliance with Ethical Standards

Conflicts of Interest   Authors declare that they have no conflict of interest.

Ethical Approval   All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent   Written consent from the study was unavailable because the dataset comprises only de-identified secondary data for research purposes, and the Institutional Review Board of St. Martin de Porres Hospital issued a formal written waiver of the need for consent and approved the study.

References

1. Chen, N., Hsu, C.C., Yamagata, K., and Langham, R., Challenging chronic kidney disease: Experience from chronic kidney disease prevention programs in Shanghai, Japan. *Taiwan and Australia. Nephrology.* 15(s2):31–36, 2010.
2. National Kidney Foundation, About chronic kidney disease. Available: http://www.kidney.org/kidneydisease/aboutckd.cfm, 2015.
3. Thomas, R., Kanso, A., and Sedor, J.R., Chronic kidney disease and its complications. *Prim. Care.* 35(2):329–344, 2008.
4. Kronenberg, F., Emerging risk factors and markers of chronic kidney disease progression. *Nat. Rev. Nephrol.* 5(12):677–689, 2009.
5. Abbasi, M. A., Chertow, G. M., and Hall, Y. N., End-stage renal disease. *Clin. Evid.*, 2010.
6. Yeh, J.Y., Wu, T.H., and Tsao, C.W., Using data mining techniques to predict hospitalization of hemodialysis patients. *Decis. Support. Syst.* 50(2):439–448, 2011.
7. Taiwan Society of Nephrology, Available: http://www.tsn.org.tw/, 2015.
8. Collins, G.S., Omar, O., Shanyinde, M., and Yu, L.M., A systematic review finds prediction models for chronic kidney were poorly reported and often developed using inappropriate methods. *J. Clin. Epidemiol.* 66(3):263–266, 2013.
9. Yang, W.C., and Hwang, S.J., Incidence, prevalence and mortality trends of dialysis end-stage renal disease in Taiwan from 1990 to 2001: The impact of national health insurance. *Nephrol. Dial. Transplant.* 23(12):3977–3982, 2008.
10. Hsu, C.C., Hwang, S.J., Wen, C.P., Chang, H.Y., Chen, T., Shiu, R.S., et al., High prevalence and low awareness of CKD in Taiwan: A study on the relationship between serum creatinine and awareness from a nationally representative survey. *Am. J. Kidney Dis.* 48(5):727–738, 2006.
11. Haroun, M.K., Jaar, B.G., Hoffman, S.C., Comstock, G.W., Klag, M.J., and Coresh, J., Risk factors for chronic kidney disease: A prospective study of 23,534 men and women in Washington County. *Maryland. Journal of the American Society of Nephrology.* 14(11):2934–2941, 2003.
12. Stengel, B., Tarver-Carr, M.E., Powe, N.R., Eberhardt, M.S., and Brancati, F.L., Lifestyle factors, obesity and the risk of chronic kidney disease. *Epidemiology.* 14(4):479–487, 2003.
13. Peralta, C.A., Shlipak, M.G., Fan, D., Ordonez, J., Lash, J.P., Chertow, G.M., et al., Risks for end-stage renal disease, cardiovascular events, and death in Hispanic versus non-Hispanic white adults with chronic kidney disease. *J. Am. Soc. Nephrol.* 17(10): 2892–2899, 2006.
14. Schaeffner, E.S., Kurth, T., Curhan, G.C., Glynn, R.J., Rexrode, K.M., Baigent, C., et al., Cholesterol and the risk of renal dysfunction in apparently healthy men. *J. Am. Soc. Nephrol.* 14(8):2084–2091, 2003.
15. Zolbanin, H.M., Delen, D., and Zadeh, A.H., Predicting overall survivability in comorbidity of cancers: A data mining approach. *Decis. Support. Syst.* 74:150–161, 2015.
16. Oztekin, A., Delen, D., and Kong, Z.J., Predicting the graft survival for heart–lung transplantation patients: An integrated data mining methodology. *Int. J. Med. Inform.* 78(12):e84–e96, 2009.
17. Yang, C.-S., Wei, C.-P., Yuan, C.-C., and Schoung, J.-Y., Predicting the length of hospital stay of burn patients: Comparisons of prediction accuracy among different clinical stages. *Decis. Support. Syst.* 50(1):325–335, 2010.
18. Chen, L., Li, X., Yang, Y., Kurniawati, H., Sheng, Q.Z., Hu, H.Y., and Huang, N., Personal health indexing based on medical examinations: A data mining approach. *Decis. Support. Syst.* 81:54–65, 2016.
19. Shahar, Y., A framework for knowledge-based temporal abstraction. *Artif. Intell.* 90(1):79–133, 1997.
20. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and regression trees.* Wadsworth & Brooks, Monterey, 1984.
21. Cortes, C., and Vapnik, V., Support-vector networks. *Mach. Learn.* 20(3):273–297, 1995.
22. Freund, Y., and Schapire, R.E., A short introduction to boosting introduction to AdaBoost. *Journal of Japanese Society for Artificial Intelligence.* 14:771–780, 1999.
23. Levey, A.S., Eckardt, K.U., Tsukamoto, Y., Levin, A., Coresh, J., Rossert, J., et al., Definition and classification of chronic kidney disease: A position statement from kidney disease: Improving global outcomes (KDIGO). *Kidney Int.* 67(6):2089–2100, 2005.
24. Levey, A.S., Stevens, L.A., Schmid, C.H., Zhang, Y.L., Castro 3rd, A.F., Feldman, H.I., et al., A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* 150(9):604–612, 2009.
25. Lorenzo, V., Saracho, R., Zamora, J., Rufino, M., and Torres, A., Similar renal decline in diabetic and non-diabetic patients with comparable levels of albuminuria. *Nephrol. Dial. Transplant.* 25: 835–841, 2010.
26. Hallan, S.I., Ritz, E., Lydersen, S., Romundstad, S., Kvenild, K., and Orth, S.R., Combining GFR and albuminuria to classify CKD improves prediction of ESRD. *J. Am. Soc. Nephrol.* 20(15):1069–1077, 2009.
27. Othman, M., Kawar, B., and El Nahas, A.M., Influence of obesity on progression of non-diabetic chronic kidney disease: A retrospective cohort study. *Nephron Clinical Practice.* 113(1):16–23, 2009.
28. Go, A.S., Chertow, G.M., Fan, D.J., McCulloch, C.E., and Hsu, C.Y., Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N. Engl. J. Med.* 351(13):1296–1305, 2004.
29. Wu, I.W., Hsu, K.H., Lee, C.C., Sun, C.Y., Hsu, H.J., Tsai, C.J., et al., p-Cresyl sulphate and indoxyl sulphate predict progression of chronic kidney disease. *Nephrol. Dial. Transplant.* 26(3):938–947, 2011.
30. Perotte, A., Ranganath, R., Hirsch, J.S., Blei, D., and Elhadad, N., Risk prediction for chronic kidney disease progression using

heterogeneous electronic health record data and time series analysis. *J. Am. Med. Inform. Assoc.* 22(4):872–880, 2015.

31. Bala, S., and Kumar, K., A literature review on kidney disease prediction using data mining classification technique. *International Journal of Computer Science and Mobile Computing.* 3:960–967, 2014.

32. Vijayarani, S., and Dhayanand, M.S., Data mining classification algorithms for kidney disease prediction. *Int. J. Cybern. Inf.* 4(4), 2015.

33. Kurbalija, V., Radovanović, M., Ivanović, M., Schmidt, D., von Trzebiatowski, G.L., Burkhard, H.-D., et al., Time-series analysis in the medical domain: A study of Tacrolimus administration and influence on kidney graft function. *Comput. Biol. Med.* 50:19–31, 2014.

34. Chou, H.L., Wang, S.H., and Cheng, C.H., Discovering knowledge of hemodialysis (HD) quality using granularity-based rough set theory. *Arch. Gerontol. Geriatr.* 54:232–237, 2012.

35. Altintas, Y.Y., Gokcen, H., Ulgu, M., and Demirel, N., Analysing interactions of risk factors according to risk levels for hemodialysis patients in Turkey: A data mining application. *Gazi University Journal of Science.* 24(4):829–839, 2011.

36. Belazzi, R., Larizza, C., Magni, P., and Bellazzi, R., Temporal data mining for the quality assessment of hemodialysis services. *Artif. Intell. Med.* 34(1):25–39, 2005.

37. Tan, C., Chen, H., and Xia, C.Y., Early prediction of lung cancer based on the combination of trace element analysis in urine and an Adaboost algorithm. *J. Pharm. Biomed. Anal.* 49(3):746–752, 2009.

38. Tan, C., Li, M.L., and Qin, X., Study of the feasibility of distinguishing cigarettes of different brands using an Adaboost algorithm and near-infrared spectroscopy. *Anal. Bioanal. Chem.* 389(2):667–674, 2007.

39. Zhang, M.H., Xu, Q.S., Daeyaert, F., Lewi, P.J., and Massart, D.L., Application of boosting to classification problems in chemometrics. *Anal. Chim. Acta.* 544(1):167–176, 2005.

40. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., and Herrera, F., A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 42(4):463–484, 2012.

41. Levin, A., Djurdjev, O., Beaulieu, M., and Er, L., Variability and risk factors for kidney disease progression and death following attainment of stage 4 CKD in a referred cohort. *Am. J. Kidney Dis.* 52(4):661–671, 2008.

42. Chiu, Y.L., Chien, K.L., Lin, S.L., Chen, Y.M., Tsai, T.J., and Wu, K.D., Outcomes of stage 3-5 chronic kidney disease before end-stage renal disease at a single center in Taiwan. *Nephron Clinical Practice.* 109(3):109–118, 2008.

43. Jones, C., Roderick, P., Harris, S., and Rogerson, M., Decline in kidney function before and after nephrology referral and the effect on survival in moderate to advanced chronic kidney disease. *Nephrol. Dial. Transplant.* 21(8):2133–2143, 2006.

44. Yoshida, T., Takei, T., Shirota, S., Tsukada, M., Sugiura, H., Itabashi, M., et al., Risk factors for progression in patients with early-stage chronic kidney disease in the Japanese population. *Intern. Med.* 47(21):1859–1864, 2008.

45. Tsai, S.Y., Tseng, H.F., Tan, H.F., Chien, Y.S., and Chang, C.C., End-stage renal disease in Taiwan: A case-control study. *Journal of epidemiology.* 19(4):169–176, 2009.

46. Chang, C.L., and Chen, C.H., Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Syst. Appl.* 36(2):4035–4041, 2009.

47. Ishani, A., Grandits, G.A., Grimm, R.H., Svendsen, K.H., Collins, A.J., Prineas, R.J., et al., Association of single measurements of dipstick proteinuria, estimated glomerular filtration rate, and hematocrit with 25-year incidence of end-stage renal disease in the multiple risk factor intervention trial. *J. Am. Soc. Nephrol.* 17(5):1444–1452, 2006.

48. Mohanram, A., Zhang, Z., Shahinfar, S., Keane, W.F., Brenner, B.M., and Toto, R.D., Anemia and end-stage renal disease in patients with type 2 diabetes and nephropathy. *Kidney Int.* 66(3): 1131–1138, 2004.

49. Ohno, I., Hosoya, T., Gomi, H., Ichida, K., Okabe, H., and Hikita, M., Serum uric acid and renal prognosis in patients with IgA nephropathy. *Nephron.* 87(4):333–339, 2001.

50. Sanchez-Lozada, L.G., Tapia, E., Santamaria, J., Avila-Casado, C., Soto, V., Nepomuceno, T., et al., Mild hyperuricemia induces vasoconstriction and maintains glomerular hypertension in normal and remnant kidney rats. *Kidney Int.* 67(1):237–247, 2005.

51. Keane, W.F., Zhang, Z.X., Lyle, P.A., Cooper, M.E., de Zeeuw, D., Grunfeld, J.P., et al., Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: The RENAAL study. *Clin. J. Am. Soc. Nephrol.* 1:761–767, 2006.

52. Ravani, P., Tripepi, G., Malberti, F., Testa, S., Mallamaci, F., and Zoccali, C., Asymmetrical dimethylarginine predicts progression to dialysis and death in patients with chronic kidney disease: A competing risks modeling approach. *J. Am. Soc. Nephrol.* 16:2449–2455, 2009.

53. Taal, M.W., and Brenner, B.M., Predicting initiation and progression of chronic kidney disease: Developing renal risk scores. *Kidney Int.* 70:1694–1705, 2006.

54. Narva, A.S., and Briggs, M., The National Kidney Disease Education Program: Improving understanding, detection, and management of CKD. *Am. J. Kidney Dis.* 53:S115–S120, 2009.

55. Young, H.N., Chan, M.R., Yevzlin, A.S., and Becker, B.N., The rationale, implementation, and effect of the Medicare CKD education benefit. *Am. J. Kidney Dis.* 57:381–386, 2011.