Journal of Biomedical Informatics xxx (2019) xxx-xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com

Predicting the function of transplanted kidney in long-term care processes: Application of a hybrid model

Parviz Rashidi Khazaee[a, b], Jamshid Bagherzadeh M.[a], Zahra Niazkhani[c], Habibollah Pirnejad[d, e, *, 1]

[a] Electrical and Computer Engineering Department, Urmia University, Urmia, Iran
[b] Urmia University of Technology, Urmia, Iran
[c] Nephrology and Kidney Transplant Research Center, Urmia University of Medical Sciences, Urmia, Iran
[d] Patient Safety Research Center, Urmia University of Medical Sciences, Urmia, Iran
[e] Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Prediction model
Artificial neural network
Kidney transplantation
Glomerular filtration rate
Hybrid model
Post-transplant care

ABSTRACT

Background: A tool that can predict the estimated glomerular filtration rate (eGFR) in routine daily care can help clinicians to make better decisions for kidney transplant patients and to improve transplantation outcome. In this paper, we proposed a hybrid prediction model for predicting a future value for eGFR during long-term care processes.

Methods: Longitudinal, historical data of 942 transplant patients who received a kidney between 2001 and 2016 at Urmia kidney transplant center was used to develop a hybrid model. The model was based on three primary models: multi-layer perceptron (MLP), linear regression (LR), and a model that predicted a smoothed value of eGFR. The hybrid model used at-hand, longitudinal data of physical examinations and laboratory test values available at each visit. Two different datasets, a generalized dataset (GData) and a personalized dataset (PData), were created. Then, in both datasets, two data subsets of development and validation were created. For prediction, all records related to the fourth to tenth previous visits of patients in time order from the target date, i.e., window size (WS) = 4–10, were used. The performance of the models was evaluated using Mean Square Error (MSE) and Mean Absolute Error (MAE). The differences between the models were evaluated with the F-test and the Akaike Information Criterion (AIC).

Results: The datasets contained 35,066 records, totally. The GData contained 26,210 and 8856 records and the PData had 24,079 and 9103 records in the development and validation datasets, respectively. In the hybrid model, the MSE and MAE were 153 and 8.9 in the GData, and 113 and 7.5 in the PData, respectively. The model performance improved using a wider WS of historical records (from 4 to 10). When the WS of ten was used the MSE and MAE declined to 141 and 8.5 in the GData and to 91 and 6.9 in the PData, respectively. In both datasets, the F-test showed that the hybrid model was significantly different from other models. The AIC showed that the hybrid model had a better performance than that of others.

Conclusions: The hybrid model can predict a reliable future value for eGFR. Our results showed that longitudinal covariates help the models to produce better results. Smoothing eGFR values and using a personalized dataset to develop the models also improved the models' performances. They can be considered as a step forward towards personalized medicine.

1. Introduction

Patients who received a kidney transplant require long-term follow-ups based on the function of their transplanted kidney (TK). However, there are few tools that can help clinicians predict the function of TKs [1,2]. Glomerular filtration rate (GFR) is an optimal index to measure kidney function/failure[2]. It is the volume of fluid that human kidneys can filter per unit time [3]. Predicting functional status of a TK in future time via predicting future values of GFR allows clinicians to be aware of their patients' future health condition, provide better clinical support if indicated, and avoid consequences of graft failure by making early and best possible decisions [4].

Different prediction models have been proposed for predicting the function of TKs following transplantation [2,5,6–12], and also for predicting the progress of kidney failure in patients with chronic kidney disease (CKD) [13–16]. Those models are based on different methodologies including cox proportional hazard ratio [2,6,7,9–11,13–17], analysis of covariance (ANOVA) [5], linear regression [7,8,18–20], artificial neural networks (ANN)

* Corresponding author at: Patient Safety Research Center, Urmia University of Medical Sciences, Urmia, Iran.
Email address: pirnejad.h@umsu.ac.ir, pirnejad@eshpm.eur.nl (H. Pirnejad)
[1] Postal address: Urmia University of Medical Sciences, Emergency Alley, Resalat Blvd., Postal code: 5714783734, Urmia, Iran.

https://doi.org/10.1016/j.jbi.2019.103116
Available online xxx
1532-0464/

[11,12,19–21], support vector machine (SVM) [19,20], random forest [20], logistic regression [11,12], and nomogram[3] [7]. They are applied to gauge a current estimated GFR (eGFR), or to predict future values of eGFR [5,7,20], survival/failure of TKs [3,6–12,15,16], or the progress of renal failure in CKD [13–14].

In the context of longitudinal care, it is often desired to incorporate patients' recent clinical histories in predictive models for predicting risk or a specific critical value. These include the rate of change and volatility of biomarkers, medication history, hospitalization episodes, and quality-of-life improvements [22]. Most of the currently proposed prediction models have used static models for risk prediction and only a few of them used some types of dynamic ones [8,14,15,16,21]. Dynamic models use time-varying, longitudinal data with varying distributions within the at-risk population [15]. Studies have shown that time-dependent covariates contain valuable information and their inclusion can improve the performance of prediction models [14–16,21,23]. Greene et al. have noted that moving toward dynamic prediction by including longitudinal data in the prediction models can make the models more efficient [23]. Tangri et al. [13] and de Bruijne et al. [16] accurately underlined the improvements in the prediction models in their studies. Studies have also shown that hybrid or ensemble models can work better and predict more precisely, especially in ambiguous conditions or when there is a large amount of data [19,24,25].

Many of the proposed prediction models could not be used in long-term daily care processes for transplant patients [22], mainly because they used only baseline variables (e.g. discharge data) [5,6–10,12,16,20] or variables at a fixed period of time (e.g. the 7th day, the 6th month, or the first year after transplantation) [5,6–10,16] in order to predict the risk or a target value for fixed future intervals (e.g. 6 months, 1 year, or 5 years after transplantation) [5,6,7,9,12]. Therefore, these models cannot be used in long-term care processes when a prediction has to be done for a patient who is, for example, in his/her 3rd year of transplantation or later. Moreover, to use the results of predictive models in clinical practice, it is required not only to classify a patient into a stage of the kidney function/failure but also to calculate the strength of association with that stage [18]. This study was designed to address the aforementioned issues. Our objective was to propose and evaluate a hybrid, dynamic model to predict the value of eGFR for KT patients' upcoming visits using their longitudinally measured values of time-dependent covariates. Furthermore, in a step toward personalized medicine, we also evaluated the effect of using a so-called "personalized dataset" on the performance of our proposed model.

2. Data preparation

2.1. Study population

The datasets used for developing our prediction models included historical values of the time-dependent covariates of 942 adult KT patients (age > 18 years) in the kidney transplant center of Urmia University of Medical Sciences (UMSU), Urmia, Iran. KT patients routinely refer to an outpatient clinic after transplantation for their follow up care. Physical examinations, physicians' clinical notes and laboratory test results (biomarkers) of each clinic visit were mainly recorded in paper-based medical records. The longitudinal data of 942 adult patients who received their transplant between 2001 and 07-24 and 2016-03-06 and received their follow up care at our center between 2001 and 08-22 and 2017-04-17 (~16 years) was collected. After digitizing the data, the eGFR of each patient for each recorded visit was calculated using the Cockcroft-Gault formula, as one of the reliable eGFR estimation methods in transplantation [27]. Transplantation age and patient age at each visit time were also calculated. As suggested by Greene et al., for predictive models, the baseline time point should normally be designated as time 0, which is the starting point for data collection [23]. In our study, the date of hospital discharge was considered as the time 0. The institutional review board of UMSU reviewed and approved the study. We used anonymized data records for this study.

2.2. Datasets

Two different datasets, a generalized dataset (GData) and a personalized dataset (PData), were created before preparing and validating the cohorts. This allowed us to check the effect of model personalization on its performance.

2.2.1. Generalized dataset (GData)

The model would learn from the relationships among all members of the general population of KT patients and then use it to predict the future eGFRs value of patients for their upcoming visits. Within this dataset, we randomly selected 75% of patients and used their historical values of time-dependent covariates as the development dataset. All records of the remaining 25% of patients were used as an external unseen validation (test) dataset.

2.2.2. Personalized dataset (PData)

Taking into account the notion of personalized medicine, and similar to what clinicians do in real practice, a patient's history of biomarkers and her/his current status should be considered before decision-making. Therefore, an ideal prediction model should learn from each patient's history of TK functions and use it to predict a future function indicator of that patient's TK. To this end, the first 75% of records of each patient's historical data, in time order from the discharge day, were selected non-randomly for developing the model and the remaining 25% of that patient's records were used for evaluating the model's predictive performance.

2.3. Data cleaning and correction

Only one percent missing data existed in our dataset, which were imputed using custom imputation. For this purpose, missing values of each covariate were substituted with the moving average of its three previous values in the time order. Also, in order to have a generalizable model, outlier records were removed using 3-scaled median absolute deviation (MAD) away from the median of the dataset [28]. In total, 108 and 25outlier records were removed from the development and validation datasets in the GData, respectively; and 214 and 49 outlier records were removed from the development and validation datasets in the PData, respectively.

2.4. Moving average target (MAT)

Using MAT in our research came from the weighted exponential moving average method, which was successfully used for time series analysis [29]. The eGFR is a 'noisy' variable with a great deal of fluctuation in a general population; thus, considering it as a solitary indicator for declining TK function at the individual patient level may not be very informative [30]. MAT presented a smoothed yet similar value to a real target eGFR and was calculated by formulas (1) to (6).

$$t_1 = \frac{(eGFR(-4) + eGFR(-3))}{2} \quad (1)$$

$$t_2 = \frac{(t_1 + eGFR(-2))}{2} \quad (2)$$

$$t_3 = \frac{\left(t_2 + eGFR(-1)\right)}{2} \quad (3)$$

[3] According to the national kidney foundation (NKF), the normal GFR in young adults is approximately 120 ml/min/1.73 m^2. A decrease in GFR is an excellent index of decreasing kidney function and precedes the onset of kidney failure. There are 5 stages for kidney failure: Stage 1 (GFR ≥ 90), stage 2 (GFR between 60 and 89), stage3a (GFR between 45 and 59), stage3b (GFR between 30 and 45), stage 4 (GFR between 15 and 29), and stage

$$t_5 = \left(\frac{\left(t_4 + eGFR\,(Target)\right)}{2} + eGFR\,(Target) \right) /2 \quad (5)$$

$$t_5 = eGFR\,(MAT) \quad (6)$$

In formulas (1) to (3), the -*n* indicates the eGFR in the n^{th} previous record. The "*eGFR (Current)*" is the calculated eGFR at the time of a current visit and "*eGFR (Target)*" is the target value of eGFR in an upcoming visit. In construction of our hybrid model, the target eGFR was also smoothed by formulas (1) to (6) in model (2) (please see Section 3.1). The main reason for using "eGFR (Target)" two-times in the formula (5) was that it helped the target MAT in the prediction model to have a smoothed yet similar behavior to the real future eGFR.

2.5. Dynamic prediction and history window size

For predicting risk of clinical events in the context of long-term care where patients' clinical conditions and biomarkers may change over time, a dynamic prediction that can adapt to a patient's most recent longitudinal data is required. Using patients' longitudinal data to construct dynamic prediction models has recently got much attention [13–15,22,20]. For example, Li et al. used history windows of longitudinal data as indicators of the amount of past history that, they believed, was relevant to the prediction of the future renal failure [15]. In our study, the history window size (WS) varied from four to ten and it was considered as the size of a sliding window, containing values of prognostic biomarkers from patients' previous visits (i.e. WS = 4 means using biomarker values of the current and three previous visits). The main reason for starting WS from four was that the moving average target (MAT) was calculated based on its four previous records. The WS = 10 was set as the upper limit since the improvements in the model performance beyond WS = 10 were trivial. Moreover, using a WS greater than 10 led to an increase in the dimension of records and as a result, the number of training and validation records decreased.

3. Statistical analysis: Prediction models

The prediction of the eGFR, as an optimal measure of patients' TK function, was considered as the main goal in the current study.

3.1. Construction of the prediction models

We developed and constructed a simple but efficient hybrid prediction model based on the well-known prediction models of the MLP [12,32,19] and the LR [18–20]. The noise robust, better discrimination power, different training algorithms, parallel nature and the ability to detect nonlinear complex relationships of the MLP have made it a popular tool in outcome prediction [31,32]. On the other hand, the white-box structure of the LR made it possible to interpret its results and showed the importance of covariates [33]. In predicting clinical outcomes, the ANN-based prediction models have had comparable results with the regression-based prediction models such as the linear and logistic regression or the Cox proportional hazard ratio models [12,19,33]. Fig. 1 shows the structure of our proposed hybrid model that aims to use the prominent features of both ANN and LR models to bring out more reliable results.

The hybrid model was constructed using three primary models. Model (1) named as "direct prediction" (DP) model, tries to predict a future value for eGFR via the MLP directly from all input covariates. The primary model (2), named as "moving average prediction" (MAP) model, uses all input covariates and predicts the MAT via the MLP. The model (3), a "linear regression" (LR) model, tries to predict a future value for eGFR via the LR. Model (4) is our hybrid (HB) model and predicts the eGFR (Target) via the MLP using the outputs of the three primary models mentioned above.

In the MLP based models, the number of nodes in the hidden layer varied from one to ten and the best run was selected as the final model. The DP, MAP, and LR used the historical (lag) values of 12 covariates (see Section 3.2) with different WSs varying from four to ten of previous visits. The WS worked as lag-period; in other words, the WS of five meant using all 12 covariates from the current visit plus all 12 * 4 covariates from the four previous visits comprising 60 covariates. Therefore, in selecting WS of ten (recent 10 visits), the input set consisted of all 120 covariates from current and nine previous visits of all 12 covariates. Fig. 2 shows the internal structure of the models (1) and (2).

3.2. Candidate predictors

In this study, 12 time-dependent candidate variables were selected considering data availability and published literatures [2–11]. The covariates were demographic variables of the recipients (current age (C-Age) and transplantation age (T-Age) both at the time of each visit), physical examinations (i.e., systolic (Sys-BP) and diastolic blood pressures (Dias-BP) and patient weight at each visit), laboratory test values (i.e., fasting blood sugar (FBS), serum creatinine (Scr), blood urea nitrogen (BUN), total number of white blood cells (WBC), hemoglobin (Hgb)), and current eGFR. The patients are visited at our transplant clinic based on a local protocol. They have more scheduled visits at the early months after transplantation, and then the number of visits decreases to one visit every six-month period after two years. Patients also have irregularities in their scheduled clinic visits. To recognize this variation in the visit schedules of patients, in addition to the aforementioned covariates, we also defined a time-dependent variable, namely "time difference" (TD). It indicated the time differences between the current visit date and the date of a future visit for which the target eGFR was being predicted.

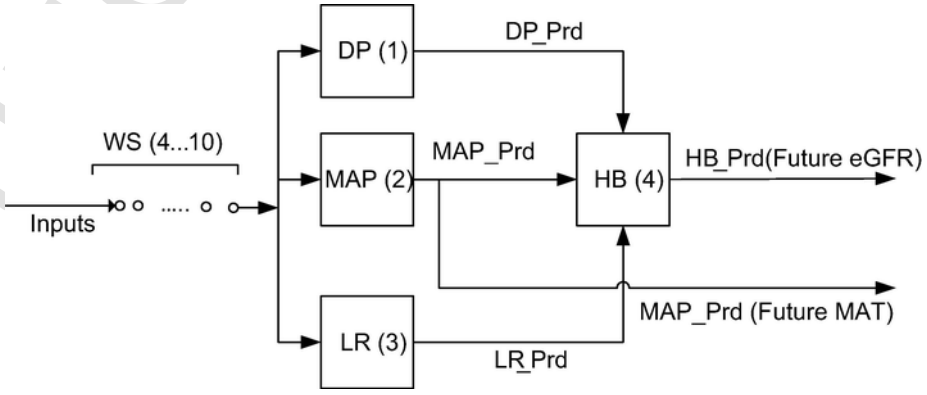

Fig. 1. The structure of our hybrid prediction model. WS = Window Size; DP = Direct Prediction Model; eGFR = estimated Glomerular Filtration Rate; MAT = Moving Average Target; MAP = moving Average Prediction Model; LR = Linear Regression Model, Prd = Predicted value; HB = Hybrid Prediction Model.

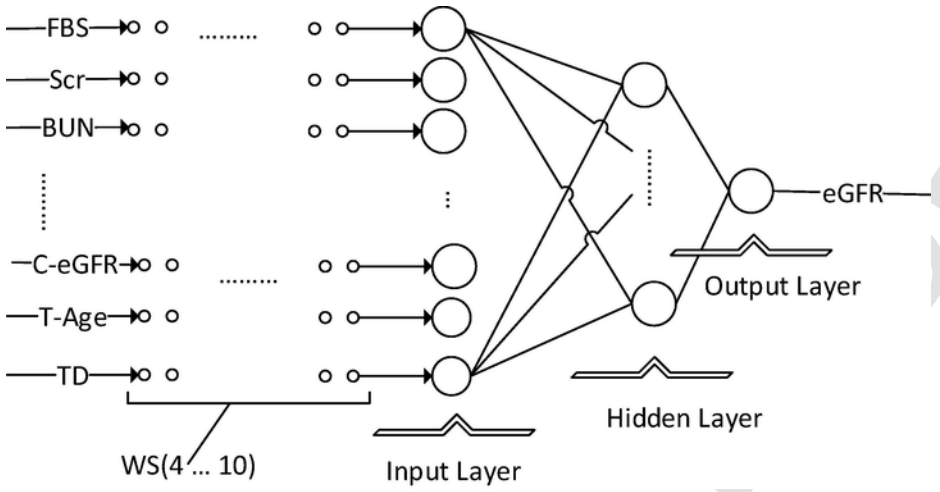

Fig. 2. The internal structure of the models (1), (2) and (4). FBS = Fasting Blood Sugar; Scr = Serum Creatinine; BUN = Blood Urea Nitrogen; C-eGFR = Current estimated Glomerular Filtration Rate; T-Age = Transplantation Age (the age of transplanted kidney); TD = Time difference between current (data available) and target day; WS = Window Size.

3.3. Primary outcome

The primary goal of this study was using the longitudinal, time-dependent covariates available at the current visit time to predict a recipient's eGFR at her/his upcoming visit.

3.4. Prediction model performance evaluation

Our models predicted a continuous variable. Therefore, two common MLP evaluation measures i.e., the Mean Square Error (MSE) and the Mean Absolute Error (MAE) were used to evaluate the performance of our models [31,18,20]. Error was defined as the difference between a predicted eGFR obtained from the models and the real eGFR at a next referring time. The criterion, MSE, is the average of squares of prediction errors and the MAE is the average of absolute errors. MSE and MAE are calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y'_i - y_i)^2, MAE = \frac{1}{n}\sum_{i=1}^{n}|y'_i - y_i|$$

where n is the number of observations, y'_i is the output (next predicted eGFR) of the prediction model for the i^{th} record and y_i is the real value (the next real eGFR) for the i^{th} record.

To evaluate the significance of the differences between the models, we performed the F-test for variances [34] and the Akaike Information Criterion (AIC) to estimate the relative quality of proposed models [35]. The AIC was calculated as:

$$AIC = n * \log\left(\frac{SSE}{n}\right) + 2 * p$$

where n is the number of samples, SSE is the Sum of Squared Errors and p is the number of model parameters.

The AIC scores are reported as ΔAIC scores. The ΔAIC is the relative difference between the best model (which has a ΔAIC of zero) and each of the other models. The ΔAIC was calculated as:

$$'AIC = AICi - minAIC$$

where $AICi$ is the score for the particular model i, and $minAIC$ is the score for the best model.

4. Results

4.1. Study participants

The historical data of lab-test values and physical examinations of 942 kidney recipients contained 35,066 records and were suitable for inclusion in our prediction models. In the GData, the mean duration of follow-ups in the development and validation cohorts were 2136 (±1289) and 2135 (±1210) days, respectively. Median number of visits per patient (visiting record history) was 35 (ranging 10–136), with an average of two months between follow-up visits. Average eGFR at the time of transplantation discharge was 68.4 ml/min/1.73 m^2. In average, there were 37 records of laboratory test results, physical examinations, and eGFRs for each patient. Detailed baseline characteristics of these two cohorts in the GData are shown in Table 1.

Table 1
Characteristics of the development and validation cohorts in Generalized Data (GData).

Characteristics	Development	Validation	p-value
Number of patients	706	236	NA
Number of records*	26,210	8856	NA
Follow-up duration in day, mean (±SD)	2136(±1289)	2135(±1210)	NA
Donor-age in year, mean (±SD)	28(±6)	28(±7)	<0.001
Donor-gender, Male (%)	667(94%)	227(96%)	0.98
Recipient-gender, Male (%)	401(57%)	154(65%)	<0.001
Time-dependent covariates (Recipient)			
Laboratory values, mean (±SD)			
FBS, mg/dl	95.71(±38.5)	96.4(±43.7)	0.27
Scr, mg/dl	1.3(±0.5)	1.31(±0.44)	<0.001
BUN, mg/dl	37.5(±20.3)	37.3(±18.8)	<0.001
WBC, count/per mcL	8.0 (±2.7)	7.7(±2.9)	0.014
Hgb, g/dl	13.1 (±2.7)	13.2(±2.2)	<0.001
Physical Examination covariates (Recipient) mean (±SD)			
Weight, kg	69.1(±13.8)	70.5(±13.9)	<0.001
Systolic BP, mm Hg	127.6(±184.9)	126.6(±16.6)	0.72
Diastolic BP, mm Hg	80.2(±9.1)	80.3(±8.9)	<0.001
Current eGFR, ml/min/1.73 m^2	70.3(±23.6)	72.8(±24.1)	<0.001
The age of transplanted kidney (in day)	1194(±1112)	1141(±1058)	<0.001
Current age (in year)	42.9(±13.9)	42.7(±12.5)	<0.001
Time difference#, (in day)	59.2(±62.3)	58.1(±61.7)	<0.001
Outcome, mean (±SD)			
Predicted eGFR, ml/min/1.73 m^2	70.2(±23.4)	72.7(±24.2)	NA

* Total number of visit records of all patients following transplant surgery in outpatient clinic.

The time elapsed between the current visiting day and a future predicted eGFR's visiting day.

Patients in the validation cohort of the PData were the same as patients included in the development cohort of this dataset. Mean age of recipients at the time of transplantation was 40 years and about less than one percent of them were older than 65 years. Male gender was dominant in both recipient and donor populations (59% and 95%, respectively). Detailed baseline characteristics of these two cohorts in the PData are shown in Table 2.

Table 2
Characteristics of development and validation cohorts in Personalized Data (PData).

Characteristics	Development	Validation	p-value
Number of patients	942	The same	NA
Number of records*	24,079	9103	NA
Follow-up duration in day, mean (±SD)	1304(±881)	749(±528)	NA
Donor-age in year, mean (±SD)	28(±7)	The same	<0.001
Donor-gender, Male (%)	894(95%)	The same	0.98
Recipient-gender, Male (%)	555(59%)	The same	<0.001
Time-dependent covariates (Recipient)			
Laboratory values, mean (±SD)			
FBS, mg/dl	93.7(±39.7)	99.4(±40)	0.27
Scr, mg/dl	1.27(±0.4)	1.47(±0.6)	<0.001
BUN, mg/dl	35.9(±18.4)	41.3(±23.1)	<0.001
WBC, count/per mcL	7.7(±2.6)	7.5(±2.8)	0.015
Hgb, g/dl	13.2(±2.3)	12.9(±1.9)	<0.001
Physical Examination covariates (Recipient) mean (±SD)			
Weight, kg	68.4(±13.7)	71.9(±13.8)	<0.001
Systolic BP, mm Hg	128.2(±190.1)	125.5(±19.2)	0.72
Diastolic BP, mm Hg	80.4(±9.5)	79.6(±7.9)	<0.001
Current eGFR, ml/min/1.73 m^2	72.7(±23.1)	66.1(±23.8)	<0.001
The age of transplanted kidney, in day	712.1(±731.4)	2188.2(±1134.7)	<0.001
Current age, in year	41.6(±13.4)	45.6(±13.5)	<0.001
Time difference#, in day	48.2(±46.0)	85.5(±83.7)	<0.001
Outcome, mean (±SD)			
Predicted eGFR, ml/min/1.73 m^2	72.9(±23.2)	65.2(±23.9)	NA

* Total number of visit records of all patients following transplant surgery in outpatient clinic.

The time elapsed between the current visiting day and a future predicted eGFR's visiting day.

Table3
Comparing Moving Average Target (MAT) Characteristics with eGFR in the two datasets.

		Development			Validation		
GData		eGFR (Y)	MAT	| Y – MAT |*	eGFR (Y)	MAT	| Y – MAT |*
	Mean	70.2	71.2	4.5	72.7	68.2	4.6
	SD	23.4	21.1	4.4	24.2	22.1	4.7
PData	Mean	72.9	72.9	4.7	65.2	63.8	3.8
	SD	23.2	20.7	4.6	23.9	22.1	3.8

* Indicates absolute difference value.

By predicting the MAT via the MAP model, the model tried to predict a smoothed target eGFR, which had similar statistical behavior to that of the real eGFR as shown in Table 3. The average absolute difference between the MAT and the real eGFR was 4.5 (±4.4) and 4.6 (±4.7) in the development and validation cohorts of the GData, respectively. This difference was 4.7 (±4.6) and 3.8 (±3.8) in the development and validation cohorts of the PData, respectively. In spite of the fact that the MAT is a smoothed eGFR, it cannot be considered as an exact indicator of the real eGFR.

Because eGFR (Target) was included in the MAT formula two-times, the MAT had a smoothed yet similar behavior to the real eGFR (Target). Fig. 3 shows the difference between the real eGFR and the MAT using a random patient from the sample population, as an example.

4.2. Prediction model performance in the GData

Prediction performances for our proposed models in the GData are shown in Table 4 and Fig. 4. The results showed that the HB-model performed well in both training and validation datasets. By increasing the WS, the performance of the models got better and error measures declined. The MSE and MAE for the HB-model in the validation data were 153 and 8.9 when the WS was four and they declined to 141 and 8.5 when the WS was ten. As shown in Fig. 4, the MAP model outperformed other models. However, the main distinction was that this model predicted a MAT and not a real eGFR.

4.3. Prediction model performance in the PData

Prediction performances for the proposed models in the personalized dataset are shown in Table 5 and Fig. 5. The results showed that the HB-Model performed well in both training and validation datasets by increasing the WS to ten. The MSE and MAE in the validation dataset were 113 and 7.5 when the WS was four and they became 91 and 6.9 when the WS was ten. Prediction results in the PData showed that the performance of the MAP model was better than those of other models.

4.4. Differences between the models

The proposed models were compared with respect to the MSE and MAE. All covariates of the prediction models had normal distribution. As it can be seen in Table 6, the MAP and LR models significantly differed from the HB model (F-Score > F Critical and p-value < 0.05) in the GData. However, only

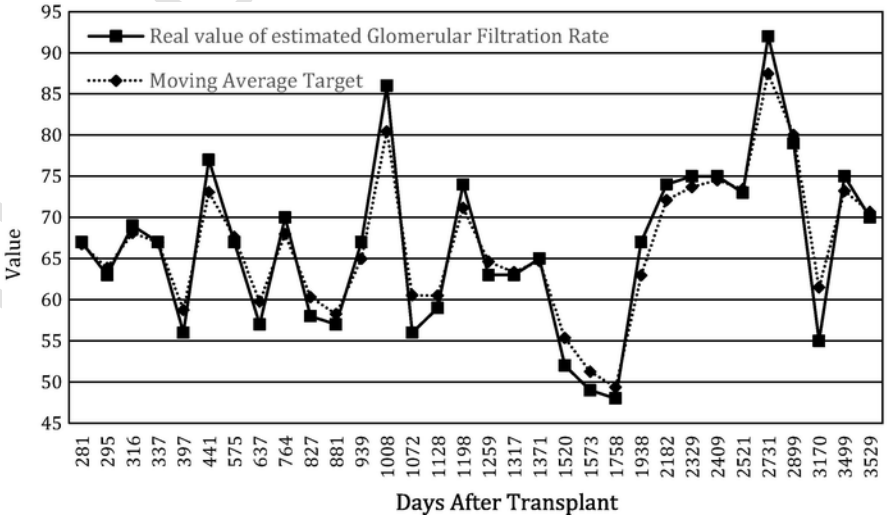

Fig. 3. Comparing real target eGFR and MAT in an example patient.

Table 4
Prediction performance of proposed models in Generalized Data (GData).

Window Size									
Models			4	5	6	7	8	9	10
Development (Train)	DP (1)	MAE	9	8.9	8.9	8.8	8.8	8.8	8.7
	MAP (2)		6.7	6.7	6.7	6.6	6.6	6.5	6.5
	LR (3)		9.2	9.1	8.7	8.7	8.7	8.7	8.7
	HB (4)		8.9	8.8	8.8	8.8	8.7	8.7	8.6
	DP (1)	MSE	157	155	154	153	151	151	147
	MAP (2)		88	87	86	85	85	84	83
	LR (3)		170	168	148	148	148	148	148
	HB (4)		156	152	152	150	149	149	145
Validation (Test)	DP (1)	MAE	8.9	8.9	8.9	8.8	8.8	8.7	8.6
	MAP (2)		6.7	6.6	6.7	6.6	6.6	6.6	6.5
	LR (3)		9.2	9.2	9	9	9	9	9
	HB (4)		8.9	8.8	8.8	8.8	8.7	8.6	8.5
	DP (1)	MSE	154	153	152	148	148	150	142
	MAP (2)		86	85	86	83	85	83	81
	LR (3)		186	165	158	158	158	158	158
	HB (4)		153	152	151	146	145	145	141

MAE = Mean Absolute Error, MSE = Mean Square Error.

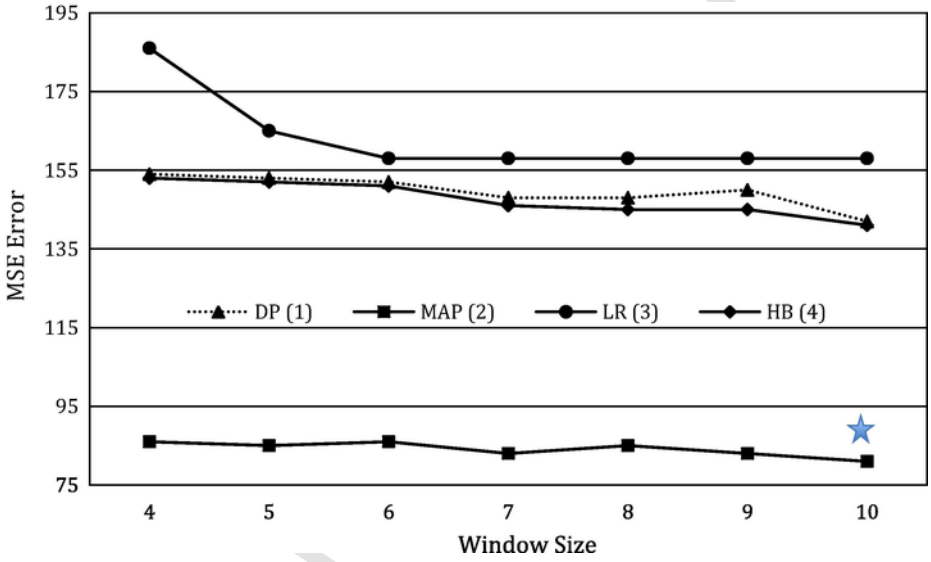

Fig. 4. Validation prediction performance (MSE) in GData. MSE = Mean Square Error, ★ Please note that although MAP model performed better than other models, as noted in the text, it predicts MAT value of eGFR and not real value of eGFR.

the LR model differed significantly from the HB model in the PData. There were significant differences between the DP and the HB models in both the GData and the PData. In both datasets, the relative quality of the DP model with respect to the optimal model (the HB model) was supported mostly by the results of the AIC. This followed by the relative quality of the MAP and the LR models.

5. Discussion

In this study, we proposed three models to predict a future value for eGFR, as the main indicator of TK function evaluation metric. We evaluated these model performances as shown in Fig. 4 and Table 4. All the DP, MAP, and LR models had more or less similar performances in prediction of eGFR in different WSs. Thus, each model can individually be used for predicting eGFR. However, as a clinical tool, the model requires to be as reliable and robust as possible. To improve the performance of these models, we used the outputs of those three primary models as the inputs for the HB model. The results of HB model's performance confirmed that its performance was better than that of the primary models. In the WS of ten, the HB model predicted a future value for eGFR with the MSE and MAE of 141 and 8.5 for the GData and 91 and 6.9 for the PData, respectively. This finding is in line with previously published articles concerning the efficiency of hybrid models [19,24,25]. The MLP as a prediction tool in the MAP and the HB models offered better performance than the LR (Table 4 and 5). Among the three primary models, the MAP had the best performance. However, this model predicts a smoothed value of eGFR instead of its real value. Hence, it cannot be considered as a good clinical tool. Figs. 4 and 5 show that including the MAP in the HB model improved its performance. Therefore, the smoothed value of the MAT could produce a better prediction for TK function. A population-based study showed that GFR is a noisy variable with a very fluctuating trend in the general population [30]. Therefore, smoothing eGFR based on each patient's own records with the MAT can help predict the trend of TK's function with less noise and fluctuation. This can be considered as a step forward towards personalized medicine [30]. Also, the AIC showed that the HB model had the best performance, thus, its selection helped minimize information loss. In spite of the better performance of the HB model, other models had more or less similar behavior. Therefore, to check the importance of the differences between our results, the F-test was done to check the equality of the variances. The results showed that there were significant dif-

Table 5
Prediction performance of proposed models in Personalized Data (PData).

Window Size									
Models			4	5	6	7	8	9	10
Development (Train)	DP (1)	MAE	9.2	9.2	9.3	9.2	9.2	9.1	9.1
	MAP (2)		6.9	6.9	6.9	6.9	6.9	6.8	6.8
	LR (3)		9.8	9.4	9.3	9.2	9.2	9.2	9.1
	HB (4)		9.2	9.2	9.2	9.1	9.1	9	9
	DP (1)	MSE	162	163	163	161	160	156	157
	MAP (2)		91	91	92	90	90	89	88
	LR (3)		165	164	164	162	161	159	158
	HB (4)		160	161	161	160	158	154	153
Validation (Test)	DP (1)	MAE	7.6	7.5	7.4	7.3	7.1	7.2	6.9
	MAP (2)		5.8	5.6	5.5	5.5	5.4	5.3	5.3
	LR (3)		7.9	7.6	7.5	7.3	7.2	7.2	7.1
	HB (4)		7.5	7.5	7.3	7.1	7	7	6.9
	DP (1)	MSE	115	110	107	103	97	97	92
	MAP (2)		65	61	60	58	55	54	52
	LR (3)		117	112	108	105	99	98	94
	HB (4)		113	109	105	101	96	95	91

MAE = Mean Absolute Error; MSE = Mean Square Error

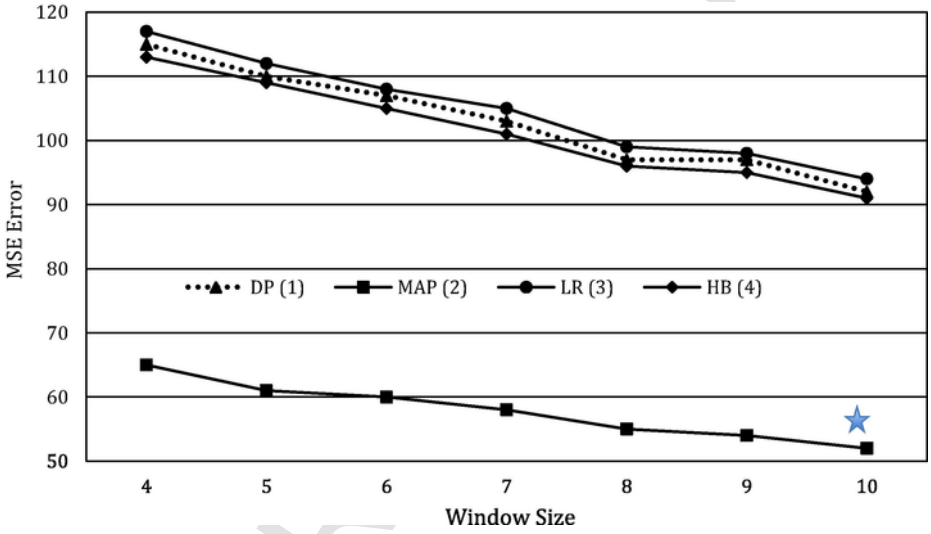

Fig. 5. Validation prediction performance (MSE) in PData. MSE = Mean Squre Error. ★ Please note MAP model, as noted in the text, predicts MAT value of eGFR and not real value of eGFR.

Table 6
The significant of the Differences between the models on GData and PData.

Comparison With HB Model					Comparing quality of the models to HB model	
	Models	F-Score	P-Value	F Critical	AIC#	ΔAIC
GData	DP	1.013	0.290	1.038	20446.38	69.5
	MAP	1.040	0.044	1.038	20481.24	104.4
	LR	1.076	0.001	1.038	20598.78	221.9
	HB				20376.88	0.0
PData	DP	1.01	0.46	1.07	5960.65	4.9
	MAP	1.02	0.34	1.07	5975.97	20.2
	LR	1.06	0.09	1.07	6004.24	48.5
	HB				5955.75	0.0

Abbreviation: F-Score = calculated by F-Test. AIC = Akaike Information Criterion.
The AIC was calculated based on sum of squared errors between model outputs and real target eGFRs.

ferences between the HB and the MAP and also significant differences existed between the HB and the LR models. Also, the results of the F-test showed that there was no significant difference between the HB and the DP models. The main reason was the fact that both models used the MLP as a core prediction tool and there were small differences concerning the MSE and MAE between them in comparison to other models.

One important observation of our study was that the models worked better when they were built on the PData (MSE of 113 in the PData *vs.*, MSE of 153 in the GData). In addition, increasing the window size improved the performance of the models. Prediction results within the validation cohort (Table 5) showed that the MSE declined around 20% (113 in WS = 4 vs. 91 in WS = 10) in the PData. This finding showed the importance of using patients' own data in building prediction models and indicated another step forward towards personalized medicine [4]. When compared with similar research, the results of our work were better than those of Pape et al. (MSE = 153 when WS = 1 vs. 199 in model 3 of their work) [18]. However, it should be noted that our study was performed using a completely different dataset.

Predicting the future trajectory of TK function is an important yet challenging issue for long-term follow-ups of transplant patients. It becomes even more important when patients' visits are scheduled for longer intervals and they get back when it is probably too late. Thus, in order to use in routine daily care processes, we needed to develop a tool that can use data that is available at patients' ordinary visits. This would enable us to use the model in daily basis of long-term care processes and make the best possible decisions for transplant

patients. Different prediction models have been proposed to help transplant care from different aspects. To the best of our knowledge, however, this is the first study that aims to help transplant care by predicting an eGFR value for the upcoming visit of a patient based on historical values of multiple longitudinal, at-hand data.

The eGFR prediction model proposed by Lasserre et al. used donor-recipient characteristics available at the time of transplantation for predicting recipient eGFR one year after transplantation [20]. They had 707 patients and applied four different regressors including ANNs, LR, Random Forest, and SVM. Although they started with 56 features, they concluded that the most valuable variables for prediction were age and creatinine levels of donors as well as gender and weight of recipients. In another study, Tiong et al. used linear or cox regression to build nomograms that predict one-year eGFR and five-year graft survival based on the data of pre-transplant phase and 6 months after transplantation [7]. Their nomograms could successfully predict one-year eGFR and five-year graft survival. Likewise, Salvadori et al. [5] used univariate and multivariate methods to predict one-year eGFR and five-year graft survival. Unlike our study, those three eGFR prediction studies used a snapshot of patients' data from pre- and/or post- transplantation phases to perform their predictions. Thus, their tools cannot be used easily in long-term routine care processes, as we explained earlier. About using time-varying covariates (or longitudinal data) in prediction, Hariharan et al. [10] showed that the efficiency of prediction improved when one-year serum creatinine and the change of serum creatinine between 6 months and one year after transplantation were considered in predicting long-term renal graft survival. Foucher et al. [9] proposed a composite kidney transplant failure score (KTFS), which used creatinine and proteinuria at 3, 6, and 12 months after transplantation to predict graft failure in 8 years. Shabir et al. [6] used laboratory results (Scr, urea, albumin, cyclosporine trough level, hemoglobin, and eGFR) at 12 months and their change between 6 and 12-months post-transplantation for predicting transplant failure, 5-year after transplantation. Kasiske et al. [2] used eGFR at 12-month post-transplantation for outcome prediction. Tangri at al. proposed a model for predicting the progression of kidney failure in CKD patients using latest-available measures of laboratory test results, which has gained much attention recently [13]. All these works showed that applying time-varying longitudinal data improved the performance of prediction models, and confirmed our approach in using longitudinal data for prediction. The joint modeling and landmark modeling are methods that used baseline covariate and historical values of time-dependent covariates for biomarkers prediction [15,21]. Their predictions strongly depend on survival model, which suffers censoring problem and requires a great deal of computation effort to do prediction, when new record becomes available [15]. Moreover, those models could not be used in routine base [22]. Likewise, the structure and the parameters of the previously proposed prediction models for predicting eGFR [7,5,18,20] and graft loss/failure/survival [2,6,8–12,15,16,17] make it impossible to be used in routine base. Those models differ from that of ours because they only allowed a single longitudinal outcome to be predicted in solitary event times. In practice, however, clinical evaluations are likely to record multiple results for a single test along the time (multiple longitudinal data) [22]. The results of our work showed that incorporation of multiple longitudinal data improved the performance of the prediction model. In our model, using the sliding window structure of the recent historical values and using only at-hand available biomarkers at each visit time makes it possible to integrate the model into daily care processes.

Furthermore, our proposed model works similarly to how clinicians work in their daily care processes. Like a clinician, the model recognizes the relationship between previous and current health indicators of a specific patient and predicts the future value of a health indicator (health condition) for that patient. Clinicians can make a better decision when necessary information related to their patients is available in their information system or memory. This, however, practically becomes difficult when patient information grows into a larger scale. In such conditions, an appropriate prediction model embedded in electronic health records (EHR) can help clinicians to manage their patients more easily and efficiently.

This study has limitations. The main limitation of this study was availability of data only for 942 out of more than 2700 patients at the time of data collection. Also, it used data from a single center. Using large amount of historical data of transplant patients from different centers could probably improve the generalizability and stability of the model and support the model usage in daily care processes. The proposed model uses only historical value of the time-dependent covariates; however, including appropriate baseline variables at pre- and post-transplantation times (e.g., donor/recipient age and gender, Human Leucocyte Antigen Typing, cold ischemia time, and other covariates), genetic data, gene expression, drugs dosage and usage mechanisms, as well as medical interventions in long-term follow-up care into a prediction model could probably provide better performance of the model and support the move towards personalized medicine. Our dataset was intrinsically ordered (time dependent) data. Performing cross-validation on this type of data is considered problematic [35]. Although we used the AIC to distinguish a better-performed prediction model, inability to perform cross-validation may still be a limitation for this study.

6. Conclusion

The proposed hybrid model had desirable results especially when larger amount of historical data (i.e., WS = 10) was used. Using at-hand, historical data of available biomarkers and sliding windows structure, which uses patients' most recent data for prediction, makes our model an appropriate tool to be used in long-term follow-up care processes. This tool has the potential to be integrated into the existing EHRs to be used routinely. We showed that including historical values of time-varying covariates promoted the prediction model to produce better results. Smoothing eGFR based on each patient's own records (i.e., MAT) assisted predicting the trend of a TK function with less noise and fluctuation. This can be considered as a step forward towards personalized medicine

Declaration of interest

None.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Uncited reference

[26].

Acknowledgments

This study was reviewed and approved by the institutional research ethics committee of Urmia University of Medical Sciences (ethics code number IR.UMSU.REC.1396.328). This paper was extracted from the PhD project of the first author. We appreciate the support provided by the Clinical Research Development Unit of Immam Hospital, Urmia University of Medical Sciences, Urmia, Iran.

References

[1] D.J. Lo, B. Kaplan, A.D. Kirk, Biomarkers for kidney transplant rejection, Nat. Rev. Nephrol. 10 (4) (2014) 215–225.
[2] B.L. Kasiske, A.K. Israni, J.J. Snyder, M.A. Skeans, Y. Peng, E.D. Weinhandl, A simple tool to predict outcomes after kidney transplant, Am. J. Kidney Dis. 56 (5) (2010) 947–960.
[3] National Kidney Foundation, https://www.kidney.org/atoz/content/gfr, 2018 (Accessed 25 August 2018).
[4] A.S. Goldfarb-Rumyantzev, Personalized medicine and prediction of outcome in kidney transplant, Am. J. Kidney Dis. 56 (5) (2010) 817–819.
[5] M. Salvadori, A. Rosati, A. Bock, et al., Estimated one-year glomerular filtration rate is the best predictor of long-term graft function following renal transplant, Transplantation 81 (2) (2006) 202–206.

[6] S. Shabir, J.M. Halimi, A. Cherukuri, et al., Predicting 5-year risk of kidney transplant failure: a prediction instrument using data available at 1 year post transplantation, Am. J. Kidney Dis. 63 (4) (2014) 643–651.
[7] H.Y. Tiong, D.A. Goldfarb, M.W. Kattan, J.M. Alster, L. Thuita, C. Yu, A. Wee, E.D. Poggio, Nomograms for predicting graft function and survival in living donor kidney transplantation based on the UNOS registry, J. Urol. 181 (3) (2009) 1248–1255.
[8] J. Moore, X. He, S. Shabir, et al., Development and evaluation of a composite risk score to predict kidney transplant failure, Am. J. Kidney Dis. 57 (5) (2011) 744–751.
[9] Y. Foucher, P. Daguin, A. Akl, et al., A clinical scoring system highly predictive of long-term kidney graft survival, Kidney Int. 78 (12) (2010) 1288–1294.
[10] S. Hariharan, M.A. Mcbride, W.S. Cherikh, et al., Post-transplant renal function in the first year predicts long-term kidney transplant survival, Kidney Int. 62 (1) (2002) 311–318.
[11] R.S. Lin, S.D. Horn, J.F. Hurdle, A.S. Goldfarb-Rumyantzev, Single and multiple time-point prediction models in kidney transplant outcomes, J. Biomed. Inform. 41 (6) (2008) 944–952.
[12] F. Shadabi, D. Sharma, Comparison of artificial neural networks with logistic regression in prediction of kidney transplant outcomes, IEEE (2009) 543–547.
[13] N. Tangri, L.A. Stevens, J. Griffith, et al., A predictive model for progression of chronic kidney disease to kidney failure, JAMA 305 (15) (2011) 1553–1559.
[14] N. Tangri, L.A. Inker, B. Hiebert, et al., A dynamic predictive model for progression of CKD, Am. J. Kidney Dis. 69 (4) (2017) 514–520.
[15] L. Li, S. Luo, B. Hu, T. Greene, Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease, Stat Biosci. (2016) 1–22.
[16] M.H. de Bruijne, Y.W. Sijpkens, L.C. Paul, R.G. Westendorp, H.C. van Houwelingen, A.H. Zwinderman, Predicting kidney graft failure using time-dependent renal function covariates, J. Clin. Epidemiol. 56 (5) (2003) 448–455.
[17] A.H. Hashemian, B. Beiranvand, M. Rezaei, A. Bardideh, E. Zand-Karimi, Comparison of artificial neural networks and cox regression models in prediction of kidney transplant survival, IJABBR 1 (10) (2013) 1204–1212.
[18] L. Pape, T. Ahlenstiel, C.D. Werner, A. Zapf, Development and validation of a new statistical model for prognosis of long-term graft function after pediatric kidney transplantation, Pediatr. Nephrol. 28 (3) (2013) 499–505.
[19] X. Liu, N. Li, L. Lv, Y. Fu, C. Cheng, C. Wang, Y. Ye, S. Li, T. Lou, Improving precision of glomerular filtration rate estimating model by ensemble learning, J. Transl. Med. 15 (1) (2017 Dec) 231.
[20] J. Lasserre, S. Arnold, M. Vingron, P. Reinke, C. Hinrichs, Predicting the outcome of renal transplantation, J. Am. Med. Inform. Assoc. 19 (2) (2011 Aug 28) 255–262.
[21] D. Rizopoulos, Joint models for longitudinal and time-to-event data: With applications in R, CRC Press, 2012 Jun 22.
[22] G.L. Hickey, P. Philipson, A. Jorgensen, R. Kolamunnage-Dona, Joint modeling of time-to-event and multivariate longitudinal outcomes: recent developments and issues, BMC Med. Res. Methodol. 16 (1) (2016 Dec) 117.
[23] T. Greene, L. Li, From static to dynamic risk prediction: time is everything, Am. J. Kidney Dis. 69 (4) (2017) 492–494.
[24] A.S. Salih, A. Abraham, Novel ensemble decision support and health care monitoring system, JNIC 2014 (2) (2014) 041–51.
[25] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 1 (50) (2003) 159–175.
[26] D. Cockcroft, M.D. Gault, Prediction of creatinine clearance from serum creatinine, Nephron 16 (1976) 31–41.
[27] C. Leys, C. Ley, O. Klein, P. Bernard, L. Licata, Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol. 49 (4) (2013 Jul 1) 764–766.
[28] S. Hansun. A new approach of moving average method in time series analysis. In New Media Studies (CoNMedia), Conference on 2013 Nov 27 pp. 1–4. IEEE; 2013.
[29] S. de Lusignan, N. Poh, Data-modeling and visualisation in chronic kidney disease (CKD): a step towards personalised medicine, J. Innov. Health Inform. 19 (2) (2011) 57–63.
[30] S. Haykin, Neural Networks: A comprehensive foundation, McMillan, New York, USA, 1999.
[31] F.E. Ahmed, Artificial neural networks for diagnosis and survival prediction in colon cancer, Mol. Cancer 4 (1) (2005) 29.
[32] D.J. Sargent, Comparison of artificial neural networks with other statistical approaches: results from medical data sets, Cancer 91 (8) (2001) 1636–1642.
[33] H. Akaike, Fitting autoregressive models for prediction, Ann. Inst. Stat. Mathem. 21 (1) (1969) 243–247.
[34] D.C. Montgomery, G.C. Runger, Applied statistics and probability for engineers, John Wiley & Sons, 2010 Mar 22.
[35] C. Bergmeir, R.J. Hyndman, B. Koo, A note on the validity of cross-validation for evaluating autoregressive time series prediction, Computat. Stat. Data Anal. 120 (4) (2018) 70–83.