



Prediction and Risk Stratification of Cardiovascular Disease in Diabetic Kidney Disease Patients

Jingjing Ren^{1,2,3,4,5}, Dongwei Liu^{1,2,3,4}, Guangpu Li^{1,2,3,4,5}, Jiayu Duan^{1,2,3,4,5*}, Jiancheng Dong^{5*} and Zhangsuo Liu^{1,2,3,4*}

¹ Department of Integrated Traditional and Western Nephrology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, ² Research Institute of Nephrology, Zhengzhou University, Zhengzhou, China, ³ Henan Province Research Center for Kidney Disease, Zhengzhou, China, ⁴ Key Laboratory of Precision Diagnosis and Treatment for Chronic Kidney Disease in Henan Province, Zhengzhou, China, ⁵ Clinical Research Center of Big-data, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:

Chayakrit Krittanawong,
New York University, United States

Reviewed by:

Luis Miguel Ruilope,
University Hospital October 12, Spain

John Adeoye,

The University of Hong Kong, Hong
Kong SAR, China
Vijay Viswanathan,

Prof. M. Viswanathan Diabetes
Research Center, India

Filippo Cademartiri,

Gabriele Monasterio Tuscany
Foundation (CNR), Italy

*Correspondence:

Zhangsuo Liu
zhangsuoliu@zzu.edu.cn
Jiancheng Dong
dongjc@ntu.edu.cn
Jiayu Duan
jyduan@hotmail.com

Specialty section:

This article was submitted to
Cardiovascular Epidemiology and
Prevention,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 20 April 2022

Accepted: 30 May 2022

Published: 24 June 2022

Citation:

Ren J, Liu D, Li G, Duan J, Dong J
and Liu Z (2022) Prediction and Risk
Stratification of Cardiovascular
Disease in Diabetic Kidney Disease
Patients.
Front. Cardiovasc. Med. 9:923549.
doi: 10.3389/fcvm.2022.923549

Background: Diabetic kidney disease (DKD) patients are facing an extremely high risk of cardiovascular disease (CVD), which is a major cause of death for DKD patients. We aimed to build a deep learning model to predict CVD risk among DKD patients and perform risk stratifying, which could help them perform early intervention and improve personal health management.

Methods: A retrospective cohort study was conducted to assess the risk of the occurrence of composite cardiovascular disease, which includes coronary heart disease, cerebrovascular diseases, congestive heart failure, and peripheral artery disease, in DKD patients. A least absolute shrinkage and selection operator (LASSO) regression was used to perform the variable selection. A deep learning-based survival model called DeepSurv, based on a feed-forward neural network was developed to predict CVD risk among DKD patients. We compared the model performance with the conventional Cox proportional hazards (CPH) model and the Random survival forest (RSF) model using the concordance index (C-index), the area under the curve (AUC), and integrated Brier scores (IBS).

Results: We recruited 890 patients diagnosed with DKD in this retrospective study. During a median follow-up of 10.4 months, there are 289 patients who sustained a subsequent CVD. Seven variables, including age, high density lipoprotein (HDL), hemoglobin (Hb), systolic blood pressure (SBP), smoking status, 24 h urinary protein excretion, and total cholesterol (TC), chosen by LASSO regression were used to develop the predictive model. The DeepSurv model showed the best performance, achieved a C-index of 0.767(95% confidence intervals [CI]: 0.717–0.817), AUC of 0.780(95%CI: 0.721–0.839), and IBS of 0.067 in the validation set. Then we used the cut-off value determined by ROC (receiver operating characteristic) curve to divide the patients into different risk groups. Moreover, the DeepSurv model was also applied to develop an online calculation tool for patients to conduct risk monitoring.

Conclusion: A deep-learning-based predictive model using seven clinical variables can effectively predict CVD risk among DKD patients and perform risk stratification. An online calculator allows its easy implementation.

Keywords: machine learning, diabetic kidney disease, cardiovascular disease, prediction model, risk stratification

INTRODUCTION

Diabetic kidney disease (DKD) has been one of the most serious diabetic microvascular complications, implicating up to 50% of patients with diabetes and becoming the major cause of the end-stage renal disease (ESRD) worldwide (1–3). DKD patients are at high risk of developing cardiovascular disease (CVD), bringing a heavy burden on the public health system (4, 5). Reduction in renal function is considered as an independent risk factor and predictor of CVD (6, 7). Therefore, DKD patients are more susceptible to CVD than the general population, resulting in much worse functional outcomes, morbidity, and mortality (8, 9). Meanwhile, they also have under-treated problems because of the lack of awareness of the CVD risks. However, the mechanisms by which CVD occurs in patients with DKD have also not been fully clarified. The individual performance and prognosis are often heterogeneous. Accordingly, improving awareness of cardiovascular risk factors and conducting early intervention in high-risk patients may improve prognosis and slow the progress.

There are many risk factors associated with the high prevalence of CVD in DKD patients. Studies of risk factors and predictive tools for CVD are common among the general population, such as the Framingham, QRISK, and China-PAR models (10–12). But these models often excluded the populations with a decline in kidney function, so they cannot be fully applied to the DKD patients with high risk of cardiovascular events (13, 14). Researches on cardiovascular risk factors in this population are still limited. Network Initiative Cardiovascular and Renal Clinical Trialists (INI-CRCT) also advise that we should attach more importance to these patients to enhance their cardiovascular outcomes (15). Therefore, identifying these cardiovascular risk factors and patients at high risk is still a challenging problem.

In recent years, artificial intelligence, particularly deep learning is developing rapidly and has been applied to a variety of medical fields, such as disease prediction (16), machine vision (17) and diagnostic study (18), etc. Deep learning is superior in handling different types of data for its strong computing power (19). So far, there have been many deep learning methods were developed for survival analysis (20, 21). Katzman et al. also developed a novel deep feed-forward neural network based on Cox assumption called DeepSurv, which combined survival analysis with deep learning and had the advantage to perform a prediction of time-to-event data. It has been successfully applied in the survival analysis of multiple diseases and showed promising performance in predicting patients' outcomes, such as oncological diseases, Covid-19, and atherosclerotic cardiovascular disease (22–26). Several online calculation tools were constructed based on the DeepSurv method (27–29).

In this study, we aimed to develop a predictive model based on the deep learning method to predict the progression of CVD in DKD patients. It also can help us to investigate the associated risk factors, to provide treatment recommendations for better cardiovascular outcomes and support personalized medicine. Patients were stratified into different risk subgroups using the output risk values from the model. We further evaluated the performance of the DeepSurv model with the classical Cox

proportional hazards (CPH) model and the Random survival forest (RSF) model using the concordance index (C-index), to prove that machine learning can effectively improve the predictive performance. Additionally, an easy-used online tool for calculating the incidence rate of CVD in DKD patients based on the predictive model was developed.

MATERIALS AND METHODS

Study Population

We retrospectively reviewed the patients who were diagnosed with DKD from the electronic medical records of the First Affiliated Hospital of Zhengzhou University from January 2013 to January 2020. DKD is defined as diabetic patients who performed persistent urinary albumin excretion or a reduction in estimated glomerular filtration rate (eGFR) for more than 3 months (2). Patients were required to have at least two hospitalizations, which allowed us to record patients' baseline data before a CV event (if any). The exclusion criteria were: (1) patients with incomplete clinical information ($n = 17$); (2) age <18 years ($n = 1$); (3) patients had history of cardiovascular diseases or coronary revascularization ($n = 1,065$); (4) patients with surgery, infection or injured at baseline ($n = 18$); and (5) patients with autoimmune disease or malignant tumor ($n = 45$). The selection process of patients is shown in **Supplementary Figure 1**.

Cardiovascular Outcomes

The CVD outcomes in this study were the first occurrence of a subsequent CVD, including coronary heart disease (coronary heart disease, myocardial infarction, angina, and coronary revascularization); cerebrovascular disease (hemorrhagic stroke and ischaemic stroke); congestive heart failure and peripheral arterial disease (amputations, aortic aneurysm, revascularization of the aorta or other peripheral arteries) and the combination of cardiovascular events. Outcomes were defined by the International Classification of Diseases, Tenth Revision (ICD-10) codes. The ICD-10 codes are summarized in **Supplemental Table 1**.

Statistical Analysis

we extracted the baseline patients' characteristics from the electronic medical records, including demographic details, comorbidities, physical examination measures, and laboratory values. If the missing-value ratio of a variable is more than 30%, the variable will be excluded. To impute the missing data, we applied multivariate imputation by chained equations (MICE) (30). After data imputation, we used the Z-score normalization method to normalize all variables to reduce the bias. The least absolute shrinkage and selection operator (LASSO) regression was performed to determine the significant clinical variables using 10-fold cross-validation (31). Furthermore, we used both univariate and multivariate cox regression analysis to assess the independent prognostic significance of the selected variables.

Modeling Process

The database was divided into two mutually exclusive datasets with balanced data distribution, 70% as a training set and 30% as an internal validation set. Continuous variables are presented as mean \pm standard deviation (SD) and compared between groups using the *t*-test, or as median (interquartile range [IQR]) and compared by Mann–Whitney *U*-tests, depending on whether the data is normally distributed or not. Categorical variables are shown as frequency (percentage) and compared by the chi-squared test. The cumulative incidence rate of the two sets was plotted by the Kaplan–Meier method and then compared using the log-rank test. Deep Learning-based Survival Model (DeepSurv) was used to perform the task of predicting patient-individual cardiovascular risk using the preselected variables by LASSO regression. We applied Bayesian hyperparameter optimization including learning rate decay, dropout, and other hyperparameters, to prevent overfitting (32). The list of hyperparameters of DeepSurv was provided in the **Supplementary Table 2**. More details about the DeepSurv method are available online (<https://github.com/jaredleekatzman/DeepSurv>). After splitting the datasets into the training and validation sets, we trained the model performing a 5-fold cross-validation on the training set to assess for overfitting and determine the model validity. Then we estimated the model performance in the validation set.

Model Validation

After obtaining the output predicted individual risks with the DeepSurv model, the DKD patients were then divided into high-risk and low-risk groups based on the cut-off value determined by ROC (receiver operating characteristic) curve. The cumulative incidence curves were plotted using the Kaplan–Meier method and compared using the log-rank test to visualize the difference in the predicted cumulative incidence of patients in two risk groups. The importance values of the selected variables were calculated by their component weights in the DeepSurv model, indicating the univariate contribution to the model. Finally, we compared the performance of the DeepSurv model with the CPH model and RSF model as conventional prediction methods to verify its performance. The discrimination performance of the proposed methods was assessed using the C-index, integrated Brier score (IBS) (33), and the area under the receiver-operator characteristic curve (AUC) in both the training and validation cohorts. C-index is a commonly used indicator of survival prediction, which can reflect the ability to predict time-to-event data. IBS compares the predicted survival rate with the actual status of patients. Higher C-index, higher AUC, and smaller IBS indicate a stronger fit of the model. Then we plotted the calibration curves of the predicted individual risk probabilities of CVD events at 1, 3, and 5 year. Furthermore, for these four different cardiovascular diseases included in our study, a subgroup analysis was conducted to determine which of their outcomes has the best-predicted performance using the DeepSurv models. A simple depiction of our study design is shown in **Figure 1**. The study was reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines

(**Supplementary Table 3**). Statistical analysis was completed with R v.4.1.1 and SPSS v.26 (IBM Corporation) software. Python v.3.7 and TensorFlow v.1.14 were used to implement DeepSurv models. A *p*-value <0.05 was considered statistically significant.

RESULTS

Patient Characteristics

A total of 890 patients were included in the final analysis. During a medium follow-up of 10.4 (IQR: 3.8–23.4) months, 284(31.91%) patients required rehospitalization due to subsequent cardiovascular events (128 cases of coronary heart diseases, 98 cerebrovascular diseases, 62 congestive heart failures, and 33 peripheral artery diseases; patients may have had >1 event). The median age for all patients was 52 years (IQR = 43–57), and that was 56 years (IQR = 48–65) for patients who sustained a subsequent CVD. Baseline demographic and clinical characteristics of the included patients, stratified by CVD outcomes are summarized in **Supplementary Table 4**. Patients were randomly divided into the training set (70%) and the validation set (30%), 606 patients in the training set, and 286 patients in the validation set. There were no statistically significant differences in the variables between the two groups (**Table 1**). **Figure 2** also showed that the cumulative incidence curves of the two sets were no statistically significant difference using the log-rank test ($p = 0.21$).

Variables Selection

There are 91 baseline clinical variables with at least 70% data completeness as candidate predictors used for LASSO regression. Seven variables, including age, high density lipoprotein (HDL), hemoglobin (Hb), systolic blood pressure (SBP), smoking status, 24 h urinary protein excretion, and total cholesterol (TC), were selected using the lambda with 1 SE of the minimum partial likelihood deviance. These variables are all easily available demographic, clinical characteristics, and laboratory results. Baseline characteristics of the selected variables were presented in **Table 2**. The results of the univariate analysis and multiple CPH regression analysis showed that each clinical parameter has independent prognostic power (**Supplementary Table 5**). To explore whether the correlation between variables would have an impact on the model, we applied the correlation-based heat map to calculate the correlation between every two factors. **Supplementary Figure 2** revealed that there were few correlations between the chosen variables (all correlation absolute values are <0.5).

Model Performance

After the feature selection process, all seven variables as independent predictors were used for the model development. We use the DeepSurv method to construct survival models to analyze individual CVD outcomes. A three-layer neural network with one input layer, one hidden layer, and one output layer is used to construct the predictive model. We used dropout, batch normalization, and L1 and L2 regularization

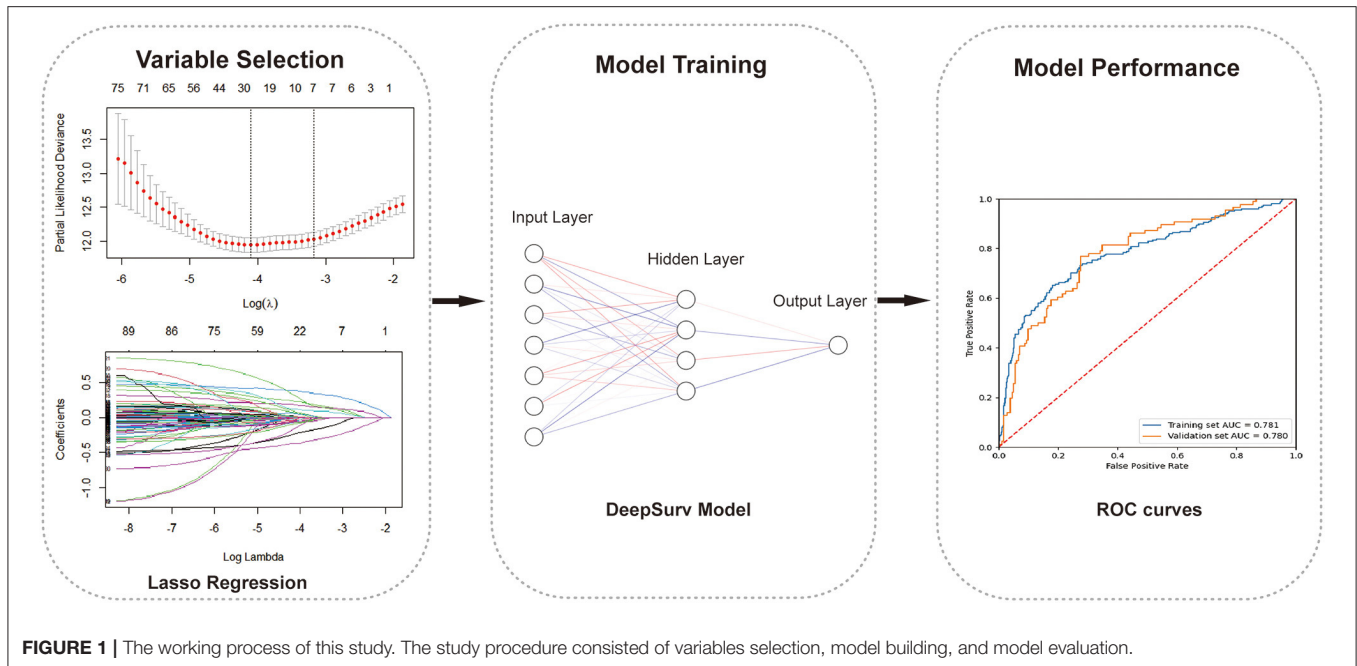


FIGURE 1 | The working process of this study. The study procedure consisted of variables selection, model building, and model evaluation.

TABLE 1 | Baseline characteristics in the training set and validation set.

	Total (n = 890)	Training set (n = 623)	Validation set (n = 267)	p-value
Smoking status				0.474
Never (%)	707 (79.4)	493 (79.1)	214 (80.2)	
Previous (%)	64 (7.2)	42 (6.7)	22 (8.2)	
Current (%)	119 (13.4)	88 (14.1)	31 (11.6)	
Age (years)	52 (45–60)	52 (45–60)	52 (45–60)	0.543
Systolic blood pressure (mmHg)	135 (126–148)	135 (126–148)	135 (125–149)	0.746
Total cholesterol (mmol/L)	4.5 (3.7–5.4)	4.6 (3.7–5.4)	4.5 (3.7–5.3)	0.615
Hemoglobin (g/L)	111 (94–131)	111 (95–130)	111 (92–132)	0.652
high density lipoprotein (mmol/L)	1.1(0.9–1.4)	1.1 (0.8–1.4)	1.1 (0.9–1.4)	0.390
24 h urinary protein (g)	2.7 (0.4–6.3)	2.7 (0.4–6.3)	2.8 (0.4–6.1)	0.296

Continuous variables are shown as mean (SD) or median (interquartile range) according to the distribution, and categorical variables are shown as frequency (percentage).

during training and selected adaptive moment (Adam) as the optimizer and Tanh as the activation function to reduce overfitting. The optimized hyperparameters for DeepSurv are shown in **Supplementary Table 6**. The 5-fold cross-validation in the training set was performed to prove the robustness of this model, with a mean AUC of 0.779(SD: 0.009), and a C-index of 0.793(SD: 0.012). To evaluate the DeepSurv model discriminative performance, we compared it with the CPH and RSF models using the C-index, AUC, and IBS. **Table 3** showed that the DeepSurv model performed best among the three survival models. The C-index of the DeepSurv model in the training and validation sets were 0.796 (95% CI: 0.761–0.831) and 0.796 (95% CI: 0.761–0.831). The AUC were 0.781 (95% CI: 0.740–0.822) and 0.780 (95% CI: 0.721–0.839) (**Figure 3**), and the IBS were 0.046 and 0.067 in

the training and validation sets respectively. We assess the importance of the variables according to their weight in the DeepSurv model, indicating the univariate contribution to the model. It revealed that older age, lower HDL, lower Hb, higher 24h urinary protein, smoking, higher SBP, and higher TC were significantly associated with the high risk of CVD (**Figure 4**). Age, HDL, and Hb were the three main relevant risk factors in the model. Subgroup analysis of four cardiovascular diseases showed that this model performed best in patients with congestive heart failure (**Table 4**), showed a C-index of 0.874 (95% CI: 0.822–0.826), and the AUC of 0.831 (95%CI: 0.770–0.892). The calibration curves of predicted event probabilities illustrated that this model has higher accuracy in predicting cardiovascular risk at 1 year (**Supplementary Figure 3**).

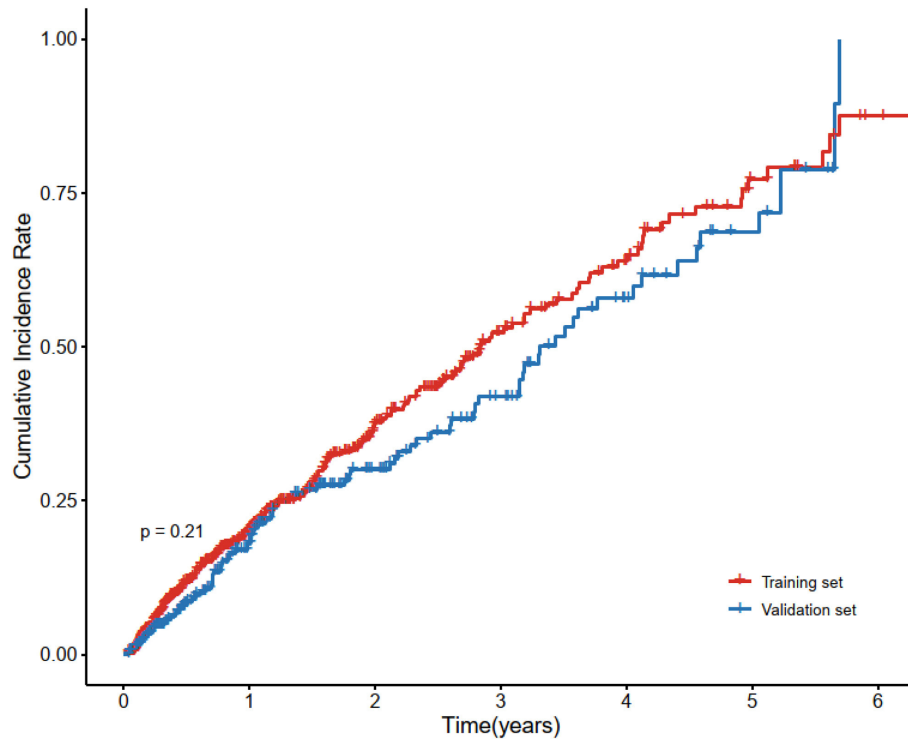


FIGURE 2 | Cumulative incidence curve of cardiovascular disease in the training set and validation set. Cardiovascular disease is the composite of coronary heart disease, cerebrovascular disease, congestive heart failure, and peripheral arterial disease. There was no statistically significant difference between the survival of the two sets using the log-rank test ($p = 0.21$).

TABLE 2 | Demographic and clinical characteristics of patients with or without CVD in the dataset.

	Total (<i>n</i> = 890)	CVD (<i>n</i> = 284)	NO CVD (<i>n</i> = 606)	<i>p</i> -value
Smoking status				<0.001
Never (%)	707 (79.4)	191 (67.3)	516 (31.5)	
Previous (%)	64 (7.2)	25 (8.8)	39 (4.1)	
Current (%)	119 (13.4)	68 (23.9)	51 (11.2)	
Age (years)	52 (45–60)	56.4 ± 11.7	51 (43–57)	<0.001
Systolic blood pressure (mmHg)	135 (126–148)	142 (131.3–160)	133 (124.8–142)	<0.001
Total cholesterol (mmol/L)	4.5 (3.7–5.4)	4.9 (4.1–5.7)	4.3 (3.6–5.2)	<0.001
Hemoglobin (g/L)	111 (94–131)	107 (91–128)	113 (95.3–133)	0.007
High density lipoprotein (mmol/L)	1.1 (0.9–1.4)	1 (0.8–1.3)	1.1 (0.9–1.4)	0.002
24 h urinary protein (g)	2.7 (0.4–6.3)	3.2 (0.7–7.1)	2.5 (0.4–6)	0.042

Continuous variables are shown as mean (SD) or median (interquartile range) according to the distribution, and categorical variables are shown as frequency (percentage).

Risk Stratification

Furthermore, after calculating the predicted individual risk score, patients were divided into high-risk and low-risk groups based on the risk cut-off values (the sensitivity of 76.7% and specificity of 72.4%). 410 patients were classified as high risk and 480 patients were classified as low risk group. We plotted the cumulative incidence curves for the two risk subgroups. **Figure 5** illustrates that risk stratification based on the DeepSurv model

can successfully stratify patients into different risk groups with significant differences ($p < 0.01$).

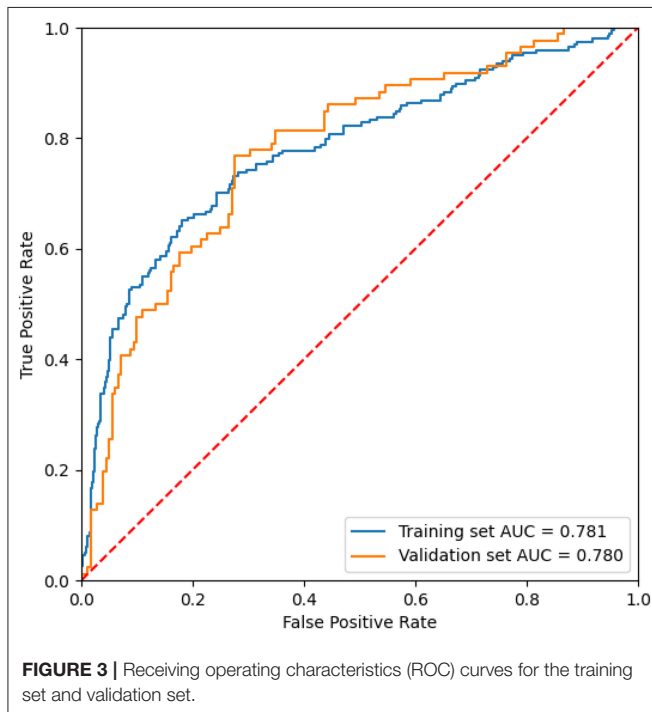
Model Visualization

The best performing model, the DeepSurv model, was used to construct an easy-used online tool to predict CVD risk in DKD patients (<http://model.51ehealth.com/>). It can calculate the individual CVD risk and monitor the trend of the risk,

TABLE 3 | Performance of different models.

	C-index (95%CI)		AUC (95%CI)		IBS	
	Training set	Validation set	Training set	Validation set	Training set	Validation set
DeepSurv	0.796 (0.761–0.831)	0.767 (0.717–0.817)	0.781 (0.740–0.822)	0.780 (0.721–0.839)	0.046	0.067
CPH	0.755 (0.719–0.790)	0.745 (0.691–0.800)	0.737 (0.698–0.776)	0.724 (0.659–0.789)	0.177	0.194
RSF	0.721 (0.681–0.761)	0.753 (0.700–0.806)	0.723 (0.680–0.766)	0.765 (0.704–0.826)	0.064	0.148

Model performance of DeepSurv, RSF, and CPH model in terms of C-index, AUC and IBS. CPH, cox proportional hazards regression; DeepSurv, deep learning-based survival model; RSF, random survival forest model; IBS, integrated brier scores; AUC, the area under the receiver-operator characteristic curve.



providing a more intuitive and understandable way to interpret the predictive model. The DeepSurv model was able to plot the predicted Kaplan-Meier survival curves for each patient. Meanwhile, for understanding convenience, we transformed the predicted time-to-event curves output from the model into curves of event incidence rates. DKD patients can input their personal information to get their risk stratification and the incidence of CVD for 1, 3, and 5 year. The interface of this risk calculator is shown in **Figure 6**. This online tool can also help physicians to choose the appropriate treatments and provide individual recommendations for the individuals to improve outcomes based on the output risk values.

DISCUSSION

The occurrence and progression of CVD is a crucial factor contributing to poor outcomes in patients with DKD (34). In this study, we developed a deep learning-based predictive model and an online tool to predict the CVD risk in DKD

patients. Our model used seven clinical variables, including age, smoking status, SBP, TC, Hb, HDL, and 24 h urinary protein, as independent predictors, and had a promising performance in the validation set. Different from the previous CVD models, our model targets a specific population: DKD patients, which allows it to be applied with greater accuracy. The C-index, ROC curves, and IBS indicated that this deep-learning model had better model discrimination in analyzing patient-individual survival outcomes than the traditional models. Furthermore, the ability to classify patients into different risk groups based on their prognosis may benefit patients by identifying high CVD risk patients and attaching more importance to them. The application of an online tool translates the predictive model into clinical practice, which may be useful for risk calculation and risk monitoring in practical clinical applications. As a result, physicians can determine the most appropriate treatment strategy to implement personalized management based on the results of risk stratification and even improve the CVD outcomes.

This study applied deep learning methods to develop models for prediction and risk stratification of CVD among DKD patients without overfitting being observed, demonstrating that this deep learning-based survival predictive model showed better performance compared to the conventional statistical method. Despite the CPH model being the most widely used approach for survival analysis in analyzing time-to-event survival data (35–37), it has its inherent drawbacks (38). The Cox model assumes the effect of each covariate is proportional and it is unable to properly model non-linearities and interaction effects. Deep-learning methods can learn to solve non-linear and intricate relationships between covariates and individual outcomes efficiently and have advantages in processing large amounts and various types of data (39). But many deep learning methods also have problems with weak interpretation in clinical practice. In this study, DeepSurv methods can combine the deep learning method with the traditional Cox assumption for survival analysis of the non-linear effect using clinical variables to predict the CVD risk of DKD patients. We reveal that this model can significantly improve the prediction performance in terms of the C-index. It not only has good model discrimination but it also can be applied for clinical use for its good interpretability. It can generate predicted Cumulative incidence curves for individuals, thus identifying them into different risk groups. The superior performance of the DeepSurv model demonstrates its ability to handle the complex association of risk factors. In addition,

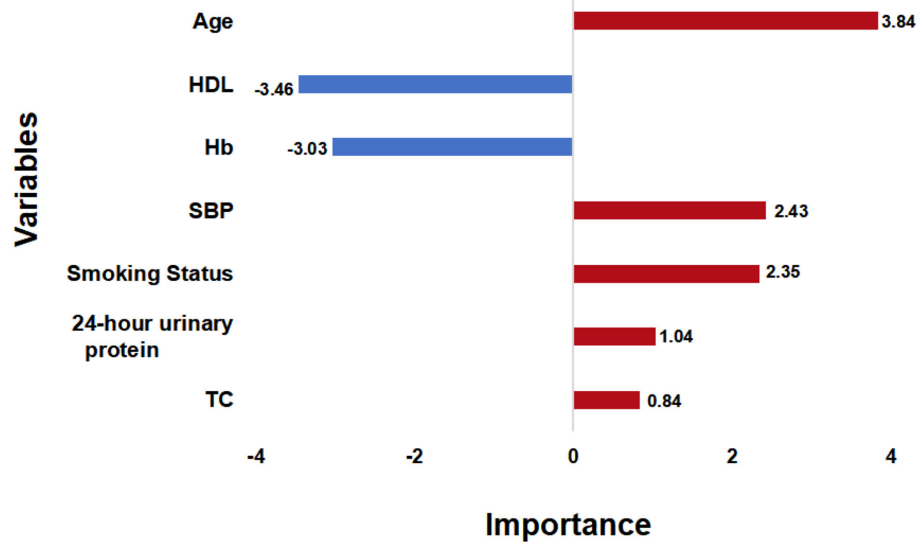


FIGURE 4 | Variable Importance. The importance score of the selected variables is calculated by their weights in the DeepSurv model. SBP, systolic blood pressure; TC, total cholesterol; Hb, hemoglobin; HDL, high density lipoprotein.

TABLE 4 | Subgroup analysis of the performance of different cardiovascular diseases.

Outcomes	C-index (95%CI)	AUC (95%CI)
Coronary heart disease	0.812 (0.768–0.856)	0.771 (0.722–0.820)
Cerebrovascular disease	0.844 (0.802–0.885)	0.806 (0.759–0.853)
Congestive heart failure	0.874 (0.822–0.826)	0.831 (0.770–0.892)
Peripheral artery disease	0.861 (0.789–0.934)	0.726 (0.646–0.806)

CI, confidence intervals.

the DeepSurv model has been widely applied to many survival analyses with a favorable prediction value (22, 23, 29). It also can provide a framework on which more datasets can be trained in the future in a broader population.

Numerous risk factors have been proved to be related to the high incidence of CVD in some DKD patients. In this study, we also demonstrated several recognized similar traditional risk predictors, consisting of age, SBP, smoking status, HDL, and TC, in consistent with the previous studies (40). A meta-analysis has shown that these traditional risk factors have been proved in previous classical predictive models based on the general population, including age, blood pressure, smoking status, and cholesterol levels (41). Age is a generally recognized risk factor for CVD (42). During aging, cardiac structural changes and functional dysfunction often occur caused by injury in fundamental cardiomyocytes (43). Hypertension is highly prevalent in chronic kidney disease (CKD) patients, which is widely recognized as a risk factor for the development of CVD (44–47). Reducing blood pressure is an important treatment strategy that not only slows the progression of renal failure

but also decreases the risk of cardiovascular disease (48). Smoking is regarded as a crucial and modifiable predictor of the progression of CVD (49). The Study of Heart and Renal Protection (SHARP) found that smoking attributed to the high risk of vascular adverse complications among patients with chronic kidney disease, and may be changed by quitting smoking (50, 51). The potential benefits of cessation are even greater than those of pharmacological treatment for cardiovascular protection (52). There was a strong and inverse correlation between HDL and CVD risk in the DKD population (53). Through reverse cholesterol transport (RCT), HDL can protect against plaque formation and development in the prevention of CVD (54). TC concentration measurement is also proved to be important in the evaluation of CVD risk factors. Besides these traditional risk factors (55), we also found that anemia and high proteinuria play a significant role in the incidence of CVD among DKD patients. Anemia was found significantly associated with the occurrence of CVD events. Anemia can cause changes in ventricular structure (56). Long-term anemia will lead to decreased oxygen capacity and utilization disorders. The compensatory hyperdynamic circulation is needed to maintain normal oxygen supply, resulting in increased cardiac output and left ventricular hypertrophy (57). Anemia often contributes to recurrent and progressive cardiac and renal deterioration, which is also called cardiorenal anemia syndrome (CRAS) (58). Proteinuria, which can reflect kidney lesions, is also a predictive factor of cardiovascular events and mortality (59). A meta-analysis also showed that participants with proteinuria had a higher risk of stroke than non-participants (33). All these included variables were independently correlated with an increased risk of CVD progression. Identifying and increasing awareness of these risk factors for CVD is essential in the early intervention and appropriate treatment of DKD

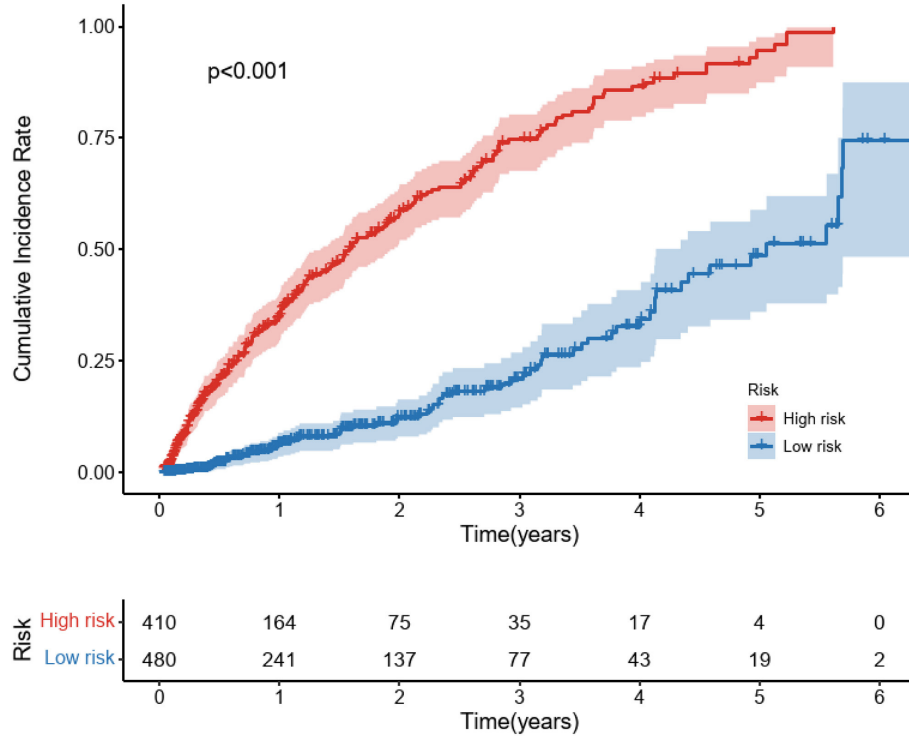
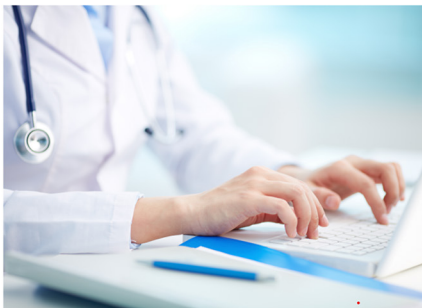


FIGURE 5 | Cumulative Incidence curves for predicted cardiovascular disease among diabetic kidney disease patients in different risk groups. Cardiovascular disease is the composite of coronary heart disease, cerebrovascular disease, congestive heart failure, and peripheral arterial disease. Patients were stratified into a high-risk group and a low-risk group based on the cut-off value of the ROC curve. The *P*-values between the high-risk and low-risk subgroups were calculated by the log-rank test.

A Calculation Tool Predict Cardiovascular Disease Risk in Diabetic Kidney Disease Patients



This calculation tool is designed to calculate the risk of developing a future cardiovascular disease. It is only valid for diabetic kidney disease patients without the history of cardiovascular disease.

Cardiovascular disease including coronary heart disease, congestive heart failure, cerebrovascular disease (subarachnoid haemorrhage is not included) and peripheral artery disease.

Enter the information below with your most recent values, then click Calculate.

Smoking status

Age

Systolic pressure mmHg

Total cholesterol mmol/L

Hemoglobin g/L

High density lipitein mmol/L

24-hour urinary protein excretion
 g

Calculate

FIGURE 6 | The interface of the online calculation tool. This online calculation tool is used to predict the cardiovascular disease risk among diabetic kidney disease patients.

patients. New integrating approaches to prognostic factors could also increase the accuracy of prediction.

There are several limitations to this study. Firstly, this study was a retrospective single-center study. Further prospective research and multicenter datasets are needed to test the generalizability and validity of the model. Secondly, a relatively small number of patients included is also a limitation in our study. Although deep learning methods have advantages in processing data with small sample sizes, Replication in a broader population is needed to confirm the superior predictive potential. Thirdly, our model used only clinical variables. Although the use of these easily accessible variables facilitates the generalization and application of the models, multi-dimensional variables such as medical imaging, omics data, and histopathological information, may also have clinical significance in the occurrence of CVD. Finally, this predictive model is based on the Chinese population, and further validation is needed to verify whether it can be applied to other ethnicities.

In conclusion, we developed and validated a new predictive model with good discrimination to estimate CVD risk among patients with DKD using seven readily available clinical variables. A user-friendly online tool based on this model was developed for clinical implementation and patient surveillance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

REFERENCES

- Microvascular complications and foot care: standards of medical care in diabetes-2019. *Diabetes care.* (2019) 42(Suppl. 1):S124–S38. doi: 10.2337/dc19-S011
- Selby NM, Taal MW. An updated overview of diabetic nephropathy: diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes Obes Metab.* (2020) 22(Suppl. 1):3–15. doi: 10.1111/dom.14007
- Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. *Diabetologia.* (2019) 62:3–16. doi: 10.1007/s00125-018-4711-2
- Jankowski J, Floege J, Fliser D, Böhm M, Marx N. Cardiovascular disease in chronic kidney disease: pathophysiological insights and therapeutic options. *Circulation.* (2021) 143:1157–72. doi: 10.1161/CIRCULATIONAHA.120.050686
- Maqbool M, Cooper ME, Jandeleit-Dahm KAM. Cardiovascular disease and diabetic kidney disease. *Semin Nephrol.* (2018) 38:217–32. doi: 10.1016/j.semnephrol.2018.02.003
- Kühn A, van der Giet M, Kuhlmann MK, Martus P, Mielke N, Ebert N, et al. Kidney function as risk factor and predictor of cardiovascular outcomes and mortality among older adults. *Am J Kidney Dis.* (2021) 77:386–96.e1. doi: 10.1053/j.ajkd.2020.09.015
- Chen J, Budoff MJ, Reilly MP, Yang W, Rosas SE, Rahman M, et al. Coronary artery calcification and risk of cardiovascular disease and death among patients with chronic kidney disease. *JAMA Cardiol.* (2017) 2:635–43. doi: 10.1001/jamacardio.2017.0363
- Kelly D, Rothwell PM. Disentangling the multiple links between renal dysfunction and cerebrovascular disease. *J Neurol Neurosurg Psychiatry.* (2020) 91:88–97. doi: 10.1136/jnnp-2019-320526
- Johansen KL, Garimella PS, Hicks CW, Kalra PA, Kelly DM, Martens S, et al. Central and peripheral arterial diseases in chronic kidney disease: conclusions from a kidney disease: improving global outcomes (Kdigo) controversies conference. *Kidney Int.* (2021) 100:35–48. doi: 10.1016/j.kint.2021.04.029
- Corlin L, Short MI, Vasani RS, Xanthakis V. Association of the duration of ideal cardiovascular health through adulthood with cardiometabolic outcomes and mortality in the framingham offspring study. *JAMA cardiol.* (2020) 5:549–56. doi: 10.1001/jamacardio.2020.0109
- Hippisley-Cox J, Coupland C, Brindle P. Development and validation of qrisk3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ.* (2017) 357:j2099. doi: 10.2337/dc19-S011
- Yang X, Li J, Hu D, Chen J, Li Y, Huang J, et al. Predicting the 10-year risks of atherosclerotic cardiovascular disease in chinese population: the china-par project (Prediction for Ascvd Risk in China). *Circulation.* (2016) 134:1430–40. doi: 10.1161/CIRCULATIONAHA.116.022367
- Zannad F, Rossignol P. Cardiovascular outcome trials in patients with advanced kidney disease: time for action. *Circulation.* (2017) 135:1769–71. doi: 10.1161/CIRCULATIONAHA.117.027338
- Lai AC, Bienstock SW, Sharma R, Skorecki K, Beerkens F, Samtani R, et al. A Personalized approach to chronic kidney disease and cardiovascular disease: jacc review topic of the week. *J Am Coll Cardiol.* (2021) 77:1470–9. doi: 10.1016/j.jacc.2021.01.028

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Scientific Ethics Committee of the First Affiliated Hospital of Zhengzhou University. Written informed consent for participation was not required for this study in accordance with the National Legislation and the Institutional Requirements.

AUTHOR CONTRIBUTIONS

JR, JDo, DL, and JDu conceived and designed the study. GL collected the data. JR performed the statistical analysis, explained the results, and wrote the manuscript. JDu discussed the draft manuscript. JDo and ZL revised the manuscript. All the authors approved the submitted version of the manuscript.

FUNDING

This study was funded by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 82103916), the General Program of the National Natural Science Foundation of China General Project (Grant No. 81970633), and the Major public welfare special projects in Henan Province (Grant No. 201300310600).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2022.923549/full#supplementary-material>

15. Rossignol P, Agarwal R, Canaud B, Charney A, Chatellier G, Craig JC, et al. Cardiovascular outcome trials in patients with chronic kidney disease: challenges associated with selection of patients and endpoints. *Eur Heart J*. (2019) 40:880–6. doi: 10.1093/eurheartj/ehx209
16. Schena FP, Anelli VV, Trotta J, Di Noia T, Manno C, Tripepi G, et al. Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin a nephropathy. *Kidney Int*. (2021) 99:1179–88. doi: 10.1016/j.kint.2020.07.046
17. Bouteldja N, Klinkhammer BM, Bülow RD, Droste P, Otten SW, Freifrau von Stillfried S, et al. Deep learning-based segmentation and quantification in experimental kidney histopathology. *J Am Soc Nephrol*. (2021) 32:52–68. doi: 10.1681/ASN.2020050597
18. Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol*. (2021) 22:132–41. doi: 10.1016/S1470-2045(20)30535-0
19. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Intern Med*. (2019) 179:293–4. doi: 10.1001/jamainternmed.2018.7117
20. Lee C, Yoon J, Schaar MV. Dynamic-deephit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Trans Biomed Eng*. (2020) 67:122–33. doi: 10.1109/TBME.2019.2909027
21. Zhao L, Feng D. Deep neural networks for survival analysis using pseudo values. *IEEE J Biomed Health Inform*. (2020) 24:3308–14. doi: 10.1109/JBHI.2020.2980204
22. Shamout FE, Shen Y, Wu N, Kaku A, Park J, Makino T, et al. An artificial intelligence system for predicting the deterioration of covid-19 patients in the emergency department. *NPJ Digital Med*. (2021) 4:80. doi: 10.1038/s41746-021-00453-0
23. Byun SS, Heo TS, Choi JM, Jeong YS, Kim YS, Lee WK, et al. Deep learning based prediction of prognosis in non-metastatic clear cell renal cell carcinoma. *Sci Rep*. (2021) 11:1242. doi: 10.1038/s41598-020-80262-9
24. Adeoye J, Koohi-Moghadam M, Lo AWL, Tsang RK, Chow VLY, Zheng LW, et al. Deep learning predicts the malignant-transformation-free survival of oral potentially malignant disorders. *Cancers*. (2021) 13:6054. doi: 10.3390/cancers13236054
25. Moradmand H, Aghamiri SMR, Ghaderi R, Emami H. The role of deep learning-based survival model in improving survival prediction of patients with glioblastoma. *Cancer Med*. (2021) 10:7048–59. doi: 10.1002/cam4.4230
26. Hathaway QA, Yanamala N, Budoff MJ, Sengupta PP, Zeb I. Deep neural survival networks for cardiovascular risk prediction: the multi-ethnic study of atherosclerosis (mesa). *Comput Biol Med*. (2021) 139:104983. doi: 10.1016/j.combiomed.2021.104983
27. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. (2019) 9:6994. doi: 10.1038/s41598-019-43372-7
28. de Vries BCS, Hegeman JH, Nijmeijer W, Geerdink J, Seifert C, Groothuis-Oudshoorn CGM. Comparing three machine learning approaches to design a risk assessment tool for future fractures: predicting a subsequent major osteoporotic fracture in fracture patients with osteopenia and osteoporosis. *Osteoporosis Int*. (2021) 32:437–49. doi: 10.1007/s00198-020-05735-z
29. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Network Open*. (2020) 3:e205842. doi: 10.1001/jamanetworkopen.2020.5842
30. Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in r. *J Stat Softw*. (2011) 45:1–67. doi: 10.18637/jss.v045.i03
31. Ramsay IS, Ma S, Fisher M, Loewy RL, Ragland JD, Niendam T, et al. Model selection and prediction of outcomes in recent onset schizophrenia patients who undergo cognitive training. *Schizophr Res Cogn*. (2018) 11:1–5. doi: 10.1016/j.scog.2017.10.001
32. Feurer M, Jost T, Hutter F. Initializing bayesian hyperparameter optimization via meta-learning. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Austin, Texas: AAAI Press (2015)
33. Ninomiya T, Perkovic V, Verdon C, Barzi F, Cass A, Gallagher M, et al. Proteinuria and stroke: a meta-analysis of cohort studies. *Am J Kidney Dis*. (2009) 53:417–25. doi: 10.1053/j.ajkd.2008.08.032
34. Santoro D, Gembillo G, Andò G. Glomerular filtration rate as a predictor of outcome in acute coronary syndrome complicated by atrial fibrillation. *J Clin Med*. (2020) 9:1466. doi: 10.3390/jcm9051466
35. Kaze AD, Santhanam P, Erqou S, Bertoni AG, Ahima RS, Echouffo-Tcheugui JB. Microvascular disease and cardiovascular outcomes among individuals with type 2 diabetes. *Diabetes Res Clin Pract*. (2021) 176:108859. doi: 10.1016/j.diabres.2021.108859
36. Lee HH, Lee H, Cho SMJ, Kim DW, Park S, Kim HC. On-treatment blood pressure and cardiovascular outcomes in adults with hypertension and left ventricular hypertrophy. *J Am Coll Cardiol*. (2021) 78:1485–95. doi: 10.1016/j.jacc.2021.08.015
37. Hou Y, Zhou Y, Hussain M, Budd GT, Tang WHW, Abraham J, et al. Cardiac risk stratification in cancer patients: a longitudinal patient-patient network analysis. *PLoS Med*. (2021) 18:e1003736. doi: 10.1371/journal.pmed.1003736
38. McGregor DE, Palarea-Albaladejo J, Dall PM, Hron K, Chastin S. Cox regression survival analysis with compositional covariates: application to modeling mortality risk from 24 h physical activity patterns. *Stat Methods Med Res*. (2020) 29:1447–65. doi: 10.1177/0962280219864125
39. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol*. (2018) 18:e24. doi: 10.1186/s12874-018-0482-1
40. Rahman M, Hsu JY, Desai N, Hsu CY, Anderson AH, Appel LJ, et al. Central blood pressure and cardiovascular outcomes in chronic kidney disease. *Clin J Am Soc Nephrol*. (2018) 13:585–95. doi: 10.2215/CJN.08620817
41. Jiang W, Wang J, Shen X, Lu W, Wang Y, Li W, et al. Establishment and validation of a risk prediction model for early diabetic kidney disease based on a systematic review and meta-analysis of 20 cohorts. *Diabetes Care*. (2020) 43:925–33. doi: 10.2337/dc19-1897
42. Costantino S, Paneni F, Cosentino F. Aging, metabolism and cardiovascular disease. *J Physiol*. (2016) 594:2061–73. doi: 10.1113/JP270538
43. Picca A, Mankowski RT, Burman JL, Donisi L, Kim JS, Marzetti E, et al. Mitochondrial quality control mechanisms as molecular targets in cardiac aging. *Nat Rev Cardiol*. (2018) 15:543–54. doi: 10.1038/s41569-018-0059-z
44. Muntner P, Anderson A, Charleston J, Chen Z, Ford V, Makos G, et al. Hypertension awareness, treatment, and control in adults with ckd: results from the chronic renal insufficiency cohort (Cric) study. *Am J Kidney Dis*. (2010) 55:441–51. doi: 10.1053/j.ajkd.2009.09.014
45. Yan Z, Wang Y, Li S, Wang J, Zhang L, Tan H, et al. Hypertension control in adults with ckd in china: baseline results from the chinese cohort study of chronic kidney disease (C-Stride). *Am J Hypertens*. (2018) 31:486–94. doi: 10.1093/ajh/hpx222
46. Sarafidis PA, Li S, Chen SC, Collins AJ, Brown WW, Klag MJ, et al. Hypertension awareness, treatment, and control in chronic kidney disease. *Am J Med*. (2008) 121:332–40. doi: 10.1016/j.amjmed.2007.11.025
47. McMullan CJ, Hickson DA, Taylor HA, Forman JP. Prospective analysis of the association of ambulatory blood pressure characteristics with incident chronic kidney disease. *J Hypertens*. (2015) 33:1939–46; discussion 46. doi: 10.1097/hjh.0000000000000638
48. Cheung AK, Rahman M, Reboussin DM, Craven TE, Greene T, Kimmel PL, et al. Effects of intensive BP control in CKD. *J Am Soc Nephrol*. (2017) 28:2812–23. doi: 10.1681/ASN.2017020148
49. Ding N, Sang Y, Chen J, Ballew SH, Kalbaugh CA, Salameh MJ, et al. Cigarette smoking, smoking cessation, and long-term risk of 3 major atherosclerotic diseases. *J Am Coll Cardiol*. (2019) 74:498–507. doi: 10.1016/j.jacc.2019.05.049
50. Staplin N, Haynes R, Herrington WG, Reith C, Cass A, Fellström B, et al. Smoking and adverse outcomes in patients with ckd: the study of heart and renal protection (Sharp). *Am J Kidney Dis*. (2016) 68:371–80. doi: 10.1053/j.ajkd.2016.02.052
51. Palmer SC, Mavridis D, Johnson DW, Tonelli M, Ruospo M, Strippoli GFM. Comparative effectiveness of calcimimetic agents for secondary hyperparathyroidism in adults: a systematic review and network meta-analysis. *Am J Kidney Dis*. (2020) 76:321–30. doi: 10.1053/j.ajkd.2020.02.439
52. Foley RN, Herzog CA, Collins AJ. Smoking and cardiovascular outcomes in dialysis patients: the united states renal data system wave 2 study. *Kidney Int*. (2003) 63:1462–7. doi: 10.1046/j.1523-1755.2003.00860.x
53. Jia C, Anderson JLC, Gruppen EG, Lei Y, Bakker SJL, Dullaart RPF, et al. High-density lipoprotein anti-inflammatory capacity

- and incident cardiovascular events. *Circulation*. (2021) 143:1935–45. doi: 10.1161/CIRCULATIONAHA.120.050808
54. Barter P, Gotto AM, LaRosa JC, Maroni J, Szarek M, Grundy SM, et al. Hdl cholesterol, very low levels of Ldl cholesterol, and cardiovascular events. *N Engl J Med*. (2007) 357:1301–10. doi: 10.1056/NEJMoa064278
 55. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the national cholesterol education program (Ncep) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel Iii) final report. *Circulation*. (2002) 106:3143–421. doi: 10.1161/circ.106.25.3143
 56. Föhrhéc Z, Gombos T, Borgulya G, Pozsonyi Z, Prohászka Z, Jánoskúti L. Red cell distribution width in heart failure: prediction of clinical events and relationship with markers of ineffective erythropoiesis, inflammation, renal function, and nutritional state. *Am Heart J*. (2009) 158:659–66. doi: 10.1016/j.ahj.2009.07.024
 57. Alonso N, Traveset A, Rubinat E, Ortega E, Alcubierre N, Sanahuja J, et al. Type 2 diabetes-associated carotid plaque burden is increased in patients with retinopathy compared to those without retinopathy. *Cardiovasc Diabetol*. (2015) 14:33. doi: 10.1186/s12933-015-0196-1
 58. McCullough PA. Anemia of cardiorenal syndrome. *Kidney Int Suppl*. (2021) 11:35–45. doi: 10.1016/j.kisu.2020.12.001
 59. Ohsawa M, Fujioka T, Ogasawara K, Tanno K, Okamura T, Turin TC, et al. High risks of all-cause and cardiovascular deaths in apparently healthy middle-aged people with preserved glomerular filtration rate and albuminuria: a prospective cohort study. *Int J Cardiol*. (2013) 170:167–72. doi: 10.1016/j.ijcard.2013.10.076

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ren, Liu, Li, Duan, Dong and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.