



OPEN

Development of risk models for early detection and prediction of chronic kidney disease in clinical settings

Pegah Bahrami^{1,6}, Davoud Tanbakuchi^{1,6}, Monavar Afzalaghaei^{3,4}, Majid Ghayour-Mobarhan⁵✉ & Habibollah Esmaily^{2,4}✉

Chronic kidney disease (CKD) imposes a high burden with high mortality and morbidity rates. Early detection of CKD is imperative in preventing the adverse outcomes attributed to the later stages. Therefore, this study aims to utilize machine learning techniques to predict CKD at early stages. This study uses data obtained from a large longitudinal cohort study. The features include patients' sociodemographic, anthropometric, and laboratory tests that are mostly associated with CKD based on national and international studies. Missing data and outliers were deleted using listwise and interquartile range techniques, respectively. Data initially remained imbalanced to investigate the ability of models to work on imbalanced datasets. Stratified K-folds cross-validation, a robust approach that performs well on imbalanced data, was further performed to enhance the splitting. Interestingly, an interaction was found between age and gender where contrasting data was generated, therefore, to avoid this interaction gender-specific algorithms were developed. Four main algorithms and four algorithms using the stratified K-folds cross-validation technique, consisting of gender-specific Random Forest and feedforward Neural Networks were developed using the preprocessed data of 6855 participants. The RF model in women exhibited the highest AUC of 0.90 followed closely by 0.89 in their NN model. Both models constructed for men yielded an AUC of 0.88. Sensitivity scores were higher in men compared to women. Models demonstrated subpar results regarding specificity, however, the high precision and F1 scores, make the models extremely valuable in a clinical setting to accurately identify CKD cases while minimizing false positive diagnoses. Moreover, the results from stratified K-fold cross-validation indicated that the NN models were more sensitive to the imbalanced dataset and demonstrated a marked increase in performance, particularly specificity, after this approach. These data offer valuable insights for the development of future risk stratification models for CKD.

Keywords Chronic kidney disease, Risk factors, Early diagnosis

With a global and national prevalence of 9.5% and 15.1%, respectively, chronic kidney disease (CKD) stands as a serious health problem^{1,2}. Among non-communicable diseases, CKD contributes to a significant rate of morbidities and mortalities, believed to be primarily caused by cardiovascular events in these patients³. According to the Global Burden of Disease (GBD) studies, there has been a prominent increase in the number of deaths and disability-adjusted life years (DALY) attributed to CKD and kidney dysfunction (KD) in 2021^{4,5}. This heightened burden is more prominent in lower-income countries⁴. Moreover, GBD 2019 reports indicated that KD was responsible for about 3.16 million deaths and 76.5 million DALYs around the globe⁶.

Treatment of CKD once it has reached the end stages is extremely difficult and the disease is often irreversible at those stages⁷. Moreover, renal replacement therapies and kidney transplants pose massive financial burdens, in addition to their prospective side effects¹. Therefore, early detection of CKD may be effective in alleviating the

¹School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ²Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran. ³Department of Statistics and Epidemiology, Faculty of Health, Mashhad University of Medical Sciences, Mashhad, Iran. ⁴Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. ⁵International UNESCO center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, Iran. ⁶Pegah Bahrami and Davoud Tanbakuchi contributed equally. ✉email: ghayourm@mums.ac.ir; esmailyh@mums.ac.ir

subsequent adverse events following the disease, however, since CKD remains asymptomatic till the late stages early detection is a challenge⁷. Even though, there is no integrated guideline for screening CKD yet, the Kidney Disease: Improving Global Outcomes (KDIGO) has suggested a precision and individualized approach using patients' risk factors and their likelihood of developing CKD⁸. Therefore, the establishment of valid predictive and risk models is the main pillar in CKD management⁹. Age, history of hypertension, diabetes, cardiovascular diseases, and obesity are closely related to CKD^{10,11}. Additionally, smoking and low physical activity have been shown to contribute to the formation, and progression of CKD evidenced by lower GFR, higher proteinuria, and lower post-kidney transplant survival rates^{12,13}. Moreover, cohort studies and risk models have found that patients with chronic exposure to acute phase reactants like human C-reactive protein (CRP) and pro-inflammatory factors like Uric Acid are at greater risk of renal impairment^{10,14}. Implementing practical predictive models using these factors may help in screening high-risk patients in clinical settings.

Machine learning (ML) algorithms offer robust predictive models and have attracted massive attention in the field of chronic diseases. ML predictive models have exhibited the potential to enhance the quality of screening methods, diagnosis accuracy, and overall clinical decision-making in chronic non-communicable diseases¹⁵. Supervised classification ML models leverage the information gained from a given dataset to predict the outcome of interest¹⁶. Despite efforts made to employ these models in chronic kidney disease prediction, a generalizable conclusion has yet to be drawn from the results¹⁶. To fill this gap, this study aims to investigate practical predictive factors for early CKD detection by utilizing Random Forest (RF) and Neural Network (NN) algorithms in a large population in Iran.

Materials and methods

Study population

This study was conducted using the Mashhad Stroke and Heart Atherosclerotic Disorders (MASHAD) Study database. The MASHAD cohort study was initiated in 2010 using a stratified cluster random sampling technique in three regions of Mashhad. The details of the study's methodology are discussed elsewhere¹⁷. The original database consisted of information on 9704 participants. Figure 1 shows the workflow and a schematic description of the study subjects.

Anthropometric assessments

Anthropometric measurements including height, weight, and body mass index (BMI) were evaluated according to the standard measurement techniques¹⁸. Height (cm) was measured with a tape measure to the nearest 0.01 mm. Weight (kg) was measured to the nearest 0.1 kg using electronic scales. BMI was calculated using the participants' weight in kilograms divided by the square of the height in meters (kg/m²).

Laboratory evaluation

All participants were advised to undergo 14 h overnight fasting and the blood samples were obtained in the following morning. To keep the study coherent all samples were evaluated in the same laboratory¹⁷. BT3000 biochemical analyzer was used for evaluating the following parameters: fasting blood glucose (FBG), Uric Acid (UA), high sensitivity C-reactive protein (hs CRP), Lowdensity lipoprotein (LDL), total cholesterol (TC), Triglyceride (TG), high-density lipoprotein (HDL), calcium (Ca), phosphorus (P), aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase (ALP), and creatinine (Cr). Type 2 Diabetes Mellitus (DM)

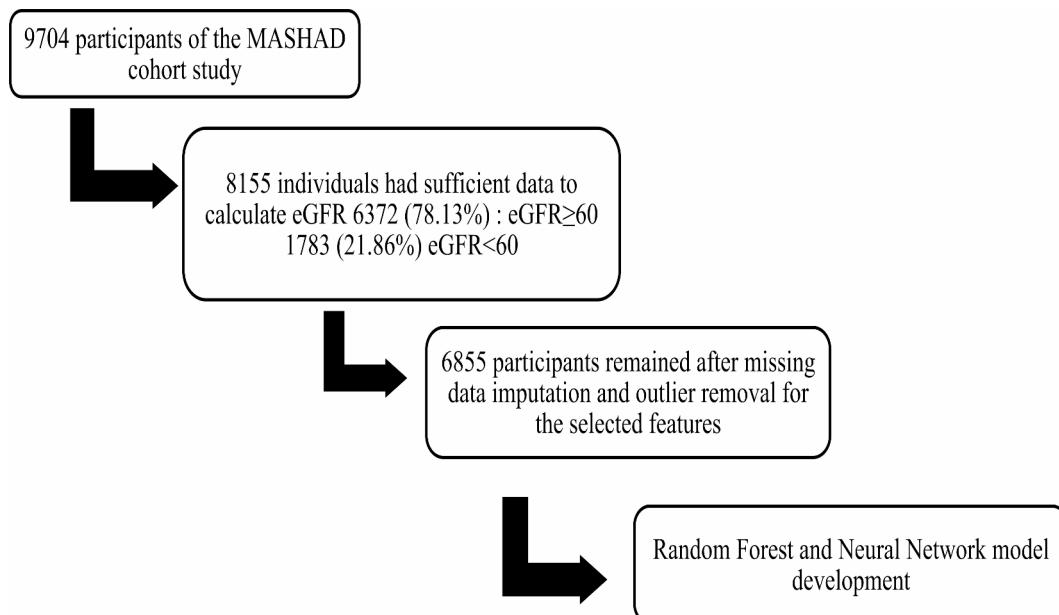


Fig. 1. Data Workflow.

was assigned based on WHO definitions as $\text{FBG} \geq 126 \text{ mg/dL}$ or in participants receiving oral hypoglycemic or insulin treatment^{17,19}.

Blood pressure assessment

Systolic (SBP) and diastolic (DBP) blood pressures were measured by standard mercury sphygmomanometers. Hypertension (HTN) was defined and reported following WHO guidelines as an $\text{SBP} \geq 140 \text{ mmHg}$ or a $\text{DBP} \geq 90 \text{ mmHg}$ or in participants being on anti-hypertensive medication^{17,20}.

Assessment of physical activity

Physical activity level (PAL) was assessed based on questionnaires evaluating activities during work hours, off time, and in bed²¹. The questionnaire filled by participants was in accordance with the James and Schofield human energy requirements equations²¹.

Assessment of other variables

Demographic, social, and medical history features including age, gender, smoking status, and cardiovascular disease (CVD) history were recorded by healthcare professionals and a nurse interview.

Feature selection and target assignment

For the purposes of this study, participants' demographic and anthropometric information including age, gender, and BMI were extracted from the database. Data obtained regarding medical and social history included a history of HTN, DM, CVD, smoking, and PAL, respectively. Laboratory data used for this study included Cr (mg/dL), hs.CRP (mg/L), FBG (mmol/L), and uric acid (mg/dL). The features used for model development in this study were decided after a thorough investigation of pivotal factors demonstrated in the KDIGO guidelines¹¹, national CKD burden study²², and availability of data in our dataset.

The Kidney Disease Quality Outcome Initiative's (KDOQI) definition, defines CKD as GFR below 60 for more than 3 months without regard for the underlying cause²³. Following this initiative, this study used a GFR cut-off of 60 (mL/min/1.73 m²) to evaluate whether the participants have chronic kidney diseases or not. The Modification of Diet in Renal Disease (MDRD) equation has shown superior results in terms of accuracy compared with other methods in assessing the Glomerular filtration rate (GFR)²⁴. Therefore, the MDRD equation was used for calculating GFR using participants' age, gender, and creatinine levels.

Data preprocessing

We used a one-hot coding approach for categorical variables including gender, HTN, DM, CVD, smoking status, and CKD status. Missing data imputation was implemented using the list-wise deletion technique. Afterward, outliers were detected with the interquartile (IQR) technique and removed.

Initially, despite our dataset being highly imbalanced, we opted for not performing over- or under-sampling techniques on the dataset to investigate whether the models perform well in an unbalanced dataset. Subsequently, for the neural network algorithms, data were normalized and standardized with scaling and centering approaches. Lastly, the dataset was split into training and testing sets with 75% and 25% allocations, respectively.

To reduce the possible bias induced by our imbalanced dataset, we opted for a splitting method that offers a robust performance on imbalanced data. Therefore, stratified K-fold cross-validation was employed. Stratified K-fold cross-validation is a robust method that maintains the proportion of each class (folds) similar to the whole dataset, thereby preserving class distribution and offering a more enhanced approach for unequal data distributions²⁵. Thus, stratified K-fold cross-validation with 10 folds ($K=10$) was performed in this study.

Random forest model

Random Forest is a complex ensemble learning model based on a multitude of regression trees. This supervised machine learning algorithm uses a similar node splitting criteria (Information Gain, Gini Index, Entropy) with the Decision Tree model²⁶. However, unlike Decision Tree, RF benefits from Bootstrap Aggregation (Bagging) and Feature Randomness, which means that RF uses a random subset of the training data for each tree and a random subset of the features at each split²⁶. The final decision in RF is reached by an accumulative prediction of the average values of all individual trees. These differences allow RF to provide more robust and generalizable data by avoiding overfitting. We used the Gini index criteria for node splitting and 500 trees to construct the forest. After the model construction, features were demonstrated as the most important to least, and partial dependency and ROC curve plots were obtained.

The Gini Index equation is described below:

$$G = \sum_{i=1}^C p(i) * (1 - p(i))$$

C: number of classes,

p(i): The probability of picking the data point with the class i.

Neural network model

Artificial neural networks are computational models resembling neurotransmission in the human brain with a complex structure consisting of predictors as inputs, hidden layers, and their processors (neurons), and the outcome variables as the final output²⁷. Inputs are weighted randomly prior to "learning" and then the weighted sum of the inputs, presented to the hidden layer, is used to measure the strengths of the output. Owing to their

unique ability to learn, NNs are capable of processing more complex patterns. Learning is defined as the repeated adjustment of weights based on measured output and the training set output²⁷. This study used a feed-forward back propagation neural network model with a sigmoid activation function. This approach allows the network to form nodal connections without making any loops²⁸.

Hyperparameters of the RF and NN models are demonstrated in Table 1.

Model evaluation

Assessment of the model's performance was carried out using the accuracy, sensitivity (recall), specificity, precision, F1 score, and Receiver Operating Characteristic - Area Under the Curve (ROC curve AUC) for all models.

Precision: True Positives / (True Positives + False Positives).

Sensitivity: True Positives / (True Positives + False Negatives).

Specificity: True Negatives / (True Negatives + False Positives).

F1 score: $2 * (\text{Precision} * \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$.

Model interpretation

The Local Interpretable Model-Agnostic Explanations (LIME) algorithms are widely used for providing understandable explanations for ML predictions²⁵. Therefore, LIME algorithms were implemented to enhance the understandability and interpretability of the outcomes.

Statistical analysis

In this study the quantitative results were stated as mean and standard deviation and the qualitative results as frequency and percentage. Shapiro-Wilk test was used to evaluate the normality of the data. Based on normality Independent T-test or Mann-Whitney test was employed to compare quantitative data. The chi-square test was used to compare the categorical variables. Descriptive and inferential analysis and data preprocessing were performed in SPSS version 27. Model development and hyperparameter tuning were performed using R programming software 4.3.2 “randomForest” and “neuralnet” packages.

Ethical approval

Written informed consent was obtained from all MASHAD study participants. This study was performed in accordance with the World Medical Association Declaration of Helsinki and approved by the ethics committee of Mashhad University of Medical Sciences.

Results

Cohort characteristics

Estimated GFR values were computed for 8155 adult participants with an average age of 58.69 ± 9.03 in men and 57.04 ± 8.70 in women, out of whom 6372 (78.13%) individuals had GFR rates equal to or over 60 mL/min/1.73 m², while the remaining 1783 (21.86%) GFR values fell below 60 mL/min/1.73 m². Table 2 summarizes the baseline characteristics of the dataset by CKD status. Average age, PAL, UA, and FBG levels were significantly higher among the CKD-positive group, conversely, the BMI and hs.CRP values were lower. After the dataset was processed for the selected features, 258 missing variables and 1042 outlier values were excluded, and the models were developed on a dataset comprising 6855 participants.

Adjustment of models based on feature interactions

Interestingly, an interaction between age and gender was observed regarding their impact on the incidence of CKD. Consequently, a detailed analysis of the results from both RF and NN models was conducted, in which BMI was identified as a protective factor against kidney failure. Furthermore, it was discovered that this protective effect occurred due to the mentioned interaction. Therefore, the results of the two models were analyzed separately by gender and age to adequately address this interaction effect.

In that regard, since menopause is a known contributor to CKD in women²⁹, we used the population of women above the age of 50 for model development.

Since NN models were aimed to be performed for men and women separately, the hidden layers and neurons were developed for each gender separately, and 2 hidden layers with (7,3) and (5,3) neurons, were implemented for men and women, respectively.

Model performance

The accuracy of all models was high with 86.34% and 84.58% for RF and NN models in women and 89.07% and 84.84% for men respectively. Sensitivity scores were higher in men with 95.48% and 90.52% for RF and NN compared to 92.26% and 88.44% in women. RF models in both men and women yielded moderate specificity

Model	Risk model in women	Risk model in men
Random Forest	variable, n_estimator = 500, criterion = gini, max_depth = 49, max_features = 8, min_samples_leaf = 15, min_samples_split = 15	variable, n_estimator = 500, criterion = gini, max_depth = 79, max_features = 8, min_samples_leaf = 10, min_samples_split = 10
Neural Network	Activation function = sigmoid, layers = 2, neurons at each layer = (5,3), linear, output = FALSE, threshold = 0.1, learning rate = 30,000	Activation function = sigmoid, hidden layers = 2, neurons at each layer = (7,3), linear, output = FALSE, threshold = 0.1, learning rate = 35,000

Table 1. Hyperparameters of the models.

		CKD -	CKD +	P-value
Age		55.40 (± 8.01)	65.69 (± 7.29)	$P < 0.001$
Gender	Male	1980 (36.78%)	782 (54.34%)	$P < 0.001$
	Female	3402 (63.21%)	657 (45.65%)	
HTN	-	3792 (79.60%)	871 (60.61%)	$P < 0.001$
	+	1579 (29.39%)	566 (10.53%)	
DM	-	4286 (79.78%)	1076 (74.87%)	$P < 0.001$
	+	1086 (20.21%)	361 (25.12%)	
CVD	-	4364 (85.96%)	1108 (77.15%)	$P < 0.001$
	+	754 (14.04%)	328 (22.84%)	
BMI		29.00 \pm 4.39	24.92 \pm 3.96	$P < 0.001$
Cr		1.06 (± 0.17)	1.56 (± 4.64)	$P = 0.00$
BUN		31.70 (± 8.38)	42.15 (± 116.62)	$P = 0.00$
Chol		206.23 (± 43.26)	201.77 (± 46.84)	$P = 0.00$
TG		148.02 (± 80.98)	135.88 (± 72.82)	$P = 0.00$
Ca		9.68 (± 2.39)	9.71 (± 0.63)	$P = 0.00$
P		4.06 (± 7.75)	3.90 (± 0.48)	$P = 0.87$
FBG		112.25 (± 41.30)	116.16 (± 47.14)	$P = 0.00$
Uric Acid		4.98 (± 1.51)	5.51 (± 1.77)	$P < 0.001$
HDL		49.21 (± 12.68)	49.50 (± 16.47)	$P = 0.83$
LDL		117.01 (± 34.52)	114.40 (± 36.80)	$P = 0.00$
AST		22.26 (± 15.17)	22.32 (± 14.76)	$P = 0.006$
ALT		21.38 (± 130.02)	17.29 (± 14.90)	$P = 0.00$
ALP		223.76 (± 79.96)	232.35 (± 72.60)	$P = 0.00$
T Bili		0.86 (± 1.79)	0.87 (± 0.35)	$P = 0.00$
D Bili		0.3 (± 3.34)	0.26 (± 0.13)	$P = 0.00$
hs.CRP		2.61 (± 3.80)	2.36 (± 3.83)	$P < 0.001$
PAL		1.70 (± 0.37)	1.86 (± 0.40)	$P < 0.001$
GFR		82.62 (± 14.92)	49.91 (± 8.41)	$P = 0.00$

Table 2. Baseline characteristics.

Female Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
Random Forest	86.34%	92.26%	53.57%	91.66%	91.96%	0.90
Stratified K-fold Random Forest	86.02%	97.40%	31.25%	87.20%	92.02%	0.94
Neural Network	84.58%	88.44%	68.29%	92.16%	87.92%	0.89
Stratified K-fold Neural Network	94.32%	96.82%	83.01%	96.26%	96.54%	0.95
Male Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
Random Forest	89.07%	95.48%	53.57%	91.92%	93.63%	0.88
Stratified K-fold Random Forest	85.71%	97.53%	54.84%	84.94%	90.80%	0.91
Neural Network	84.84%	90.52%	72.73%	87.61%	89.04%	0.88
Stratified K-fold Neural Network	92.82%	94.15%	89.49%	95.66%	94.90%	0.94

Table 3. Model Performance.

values (53.57%), while the NN models demonstrated higher specificity in both genders (68.29% and 72.73% in women and men respectively).

To further assess the feasibility of models in clinical practice precision and F1-score were also evaluated. The precision (women: RF 91.66%, NN 92.16%, men: RF 91.92%, NN 87.61%) and F1-score (women: RF 91.96%, NN 87.92%, men: RF 93.63%, NN 89.04%) were high in all models. When comparing the two genders, precision was slightly higher in the NN model for women, and the F1 scores were higher in both models attributed to men. Table 3 summarizes the performance metrics of all models.

The receiver operation characteristic (ROC) curve and the associated area under the curve (AUC) of the models are demonstrated in Fig. 2. The RF model in women exhibited the highest AUC of 0.90 followed closely by 0.89 in their NN model. Both models constructed for men yielded an AUC of 0.88.

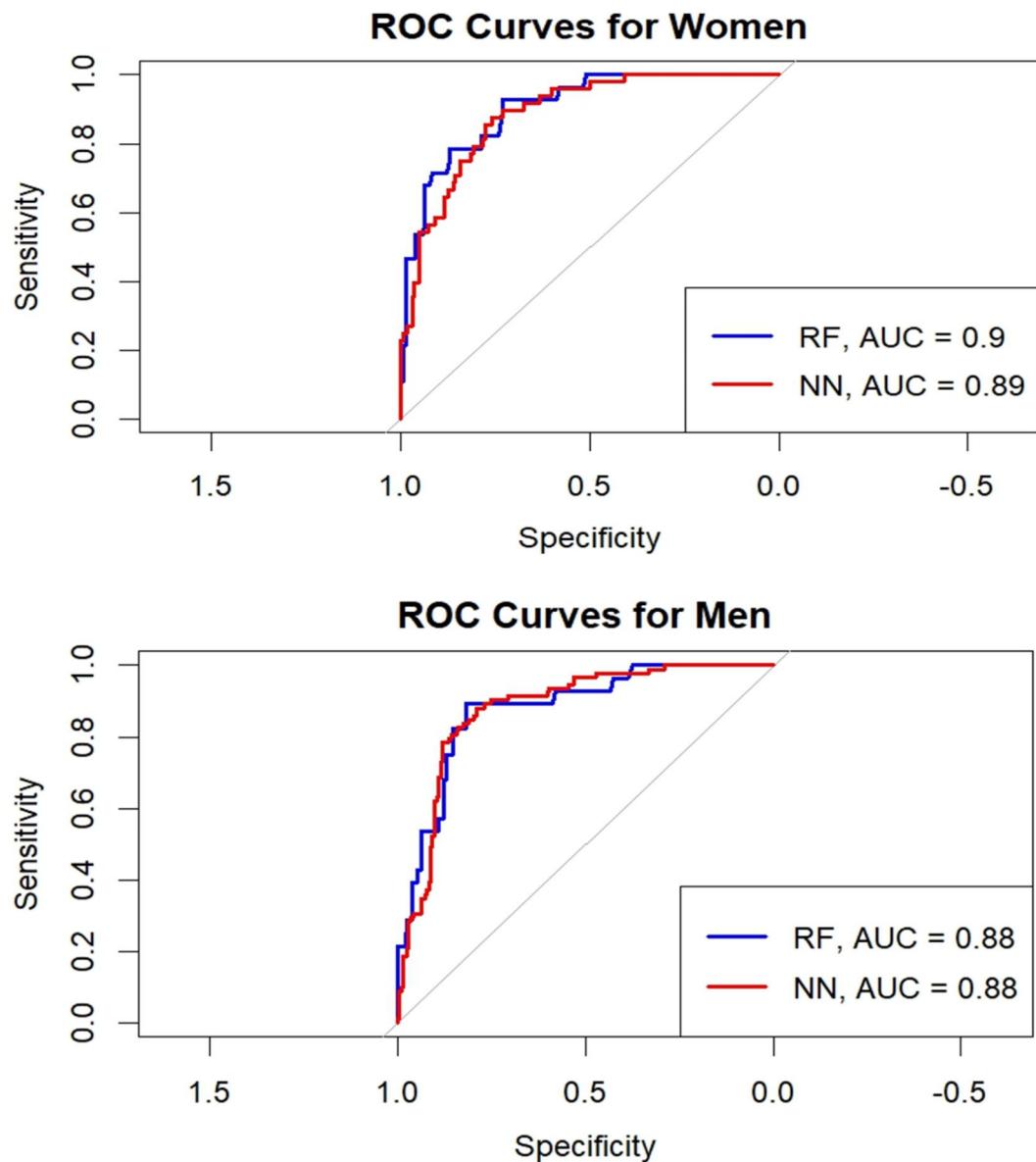


Fig. 2. Roc curves for men and women.

Implementation of stratified K-fold cross-validation demonstrated enhanced performance in all aspects in the NN models of both men and women, whereas it did not offer higher performance ability for RF models, except for a consistent increase in sensitivity for both men and women (Table 3).

Variable importance and interpretation of models

The importance of each feature and data on how various features influenced the outcomes were evaluated and the data are presented as Supplementary Figures S1-S7. The importance of the factors influencing the risk models differed between the RF (Supplementary Figures S1, S2) and NN (Supplementary Figures S3, S4) models. The RF model in women recognized BMI, age, PAL, UA, hs.CRP and FBG were the most important features, whereas although the features were the same in men, they differed in order: BMI, UA, age, PAL, hs.CRP, and FBG. The remaining variables including a history of DM, HTN, CVD, and smoking were reportedly the least important. Supplementary Figures S5 and S6 display the partial dependency plots of the important features in RF.

While BMI was still the most prominent feature in the NN models, categorical variables like HTN, DM, and gender played more important roles in NN models with BMI, FBG, gender, DM, HTN, and BMI, HTN, FBG, gender, DM being the important feature in women and men respectively.

Discussion

Unlike the development of CKD into ESRD, the detection of CKD at the initial stages has yet to yield concise prediction criteria. This study utilized random forest and artificial neural network machine learning algorithms

to create risk prediction models for the presence of CKD evidenced by eGFR levels below 60 mL/min/1.73 m². Furthermore, taking into consideration that CKD is found to be more prevalent in women at peri- and post-menopause ages compared to younger women²⁹, we trained and tested our models separately for women above 50 and men to ensure more clinically accurate results. All 4 models developed for men and women demonstrated high accuracy (over 80% in all models). Nevertheless, accuracy is not a suitable metric for evaluation in an imbalanced dataset as the class dominating the data will automatically influence the prediction. Therefore, more attention should be directed toward other performance metrics when interpreting our results. Sensitivity, precision, F1-score, and AUC were high in all models. However, models demonstrated subpar results regarding specificity with 53.57% in both RF models, 68.29%, and 72.73% in NN models for women and men respectively. The high precision and F1 scores, make the models extremely valuable in a clinical setting due to their innate ability to accurately identify CKD cases while minimizing false positive diagnoses. Our study also employed stratified K-fold cross-validation to investigate CKD prediction. The results further indicated that RF was a more robust approach for our imbalanced dataset, however, the stratified K-fold cross-validation in NN models demonstrated better outcomes and particularly higher specificity (83.01% for women and 89.49% for men). These data indicate that the NN models were more impacted by the imbalanced data compared to RF models.

In two studies using the University of California Irvine (UCI) dataset, RF and NN models have proved superior in CKD prediction compared to other supervised learning systems^{30,31}. The studies implemented the recursive feature elimination (RFE) technique and successfully gained 100% accuracy and sensitivity in RF³⁰ and an AUC of 1.0 for NN³¹ with Hb, Rbc, Packed Cell Volume, Alb, SG, HTN, and blood Urea as the most influencing features. Despite the very promising results, the relatively small sample size inhibits the findings from being generalizable in real-world clinical settings. When compared in a large dataset of 134,895 individuals in Korea, the auto-encoder neural network model demonstrated better CKD classification ability compared to RF with the highest sensitivity and specificity at stage 1 (99.76%, 99.79%)³². The auto-encoder NN is a derivative of the neural network, integrating unsupervised and supervised learning methods, and has the same nodes at the output as the input³². Despite random forest's relatively lower sensitivity at classifying higher stages of CKD, it significantly excelled in the first two stages with 0.99 sensitivity³². The features utilized in Jeong et al.'s study differed from ours with a higher number of laboratory data focusing on liver function, and lipid profile³². When comparing the results of various ML models in CKD, it is crucial to bear the racial and ethnic variances in mind as they play a significant role in CKD development, prevalence, and risk factors^{33,34}. Hence, developing nation-based CKD predictive models is of immense importance. To our knowledge, no study has yet proposed an ML model to predict the incidence of CKD at early stages in the Iranian population.

High BMI is associated with deterioration of kidney function³⁵, and based on the Framingham Heart Study, obesity, defined as BMI over 30, is a predictor of CKD in patients with no prior history of chronic kidney disease³⁶. Additionally, new onset al.buminuria is more prominently observed in obesity³⁷. The exact mechanism underlying the influence of obesity and weight gain in promoting renal dysfunction remains unknown. However, it is believed that when faced with obesity, kidneys instruct a compensatory hyperfiltration to meet the elevated metabolic demand, damaging the glomerular filtration barrier, and leading to glomerular hyper-permeability (36). Agreeing with this notion, BMI is the most influential factor in predicting CKD in our proposed models.

An interesting finding in this study is that the RF partial dependency plots in men demonstrated higher CKD prediction in men with BMI below 20 kg/m² rather than those with obesity, whereas in women the model was able to acceptably predict CKD in both lean, overweight, and obese values of BMI, albeit with more steep slopes in the overweight and an overall better predictive performance in BMI below 20 kg/m². Therefore, it can be concluded that weight has a paradoxical effect on CKD. This hypothesis is supported by several studies where lower BMI values were significantly associated with higher mortality rates and adverse renal events^{38,39}. Further, BMI-stratified studies are needed to discover the exact relationship between BMI values and CKD incidence.

Despite the imperative role of obesity in predicting CKD, few studies have investigated BMI in their risk models. In a large prediction study conducted on the Taiwanese population, BMI was not observed as one of the foremost factors contributing to CKD⁴⁰. In addition to their different ML models from our study, ethnic and behavioral differences may be of cause. This hypothesis is further confirmed by the higher burden of CKD in Iran compared to other countries globally²², which is most likely attributed to the higher incidence of CKD risk factors like obesity⁴¹.

Aging has widely been associated with diminished kidney function⁴². There is a decline in GFR values as age increases, ascribed to the inherent age-related physiologic changes in renal mass, cardiac output, and the vasoconstriction-dilation balance⁴². Age was among the most influential factors in our models. Consistent with our results, the large-scale study by Aoki et al., which investigated the predictive ability of a random forest model in assessing the GFR decline of over 30% over 5 years, demonstrated aging as an important factor influencing the model's high performance (AUC 0.85).

Hyperuricemia, described as serum UA levels exceeding 7 mg/dL in men and 6 mg/dL in women, is a prominent contributing factor in the development of new-onset CKD and its progression owing to the deposition of UA in the renal collecting tubes, and changes in the renal endothelium^{43,44}. In addition to gout, asymptomatic hyperuricemia also promotes renal dysfunction and CKD⁴³. The findings of this study, in consistency with a previous study of CKD ML models⁴⁰, identified serum UA levels as one of the most important indicators of the predictive models. Interestingly, our findings suggested that serum UA levels can predict CKD beyond the hyperuricemia definition. Estimated UA levels below 4 mg/dL and over 7 mg/dL in men, and UA levels over 5 mg/dL in women were indicative of CKD incidence. Similar to our results, a retrospective cohort study by Nakayama et al. with 138,511 participants demonstrated that serum UA levels below 4 mg/dL and over 10 mg/dL in men and UA levels below 4 mg/dL and over 7 mg/dL in women showed a more significant risk of CKD incidence or proteinuria in the male population⁴⁵. The precise cut-off levels of low and high UA in men and

women differ between studies^{45,46}, leaving room for further investigations. Variances in hormonal levels between the two genders further, necessitate the evaluation of the two genders and menopausal women separately⁴⁷.

A discrepancy was observed in our results regarding the important features in NN models compared to RF models. Although BMI and FBG were common between the models, conventional predictors of CKD like HTN and DM were of more importance in predicting CKD in the neural network. Age, UA, and PAL did not hold high importance in the NN models. When interpreting the important features of the models, it is noteworthy to bear in mind that while NN models are great in providing predictive insights, these models are less reliable in providing important features. This discrepancy is set in the key differences in the process of these models. Random Forest, owing to its ensemble method, is a robust method to determine the key variables contributing to the outcome and does so by evaluating the impact of each variable in decreasing the impurity of each tree node. Moreover, NN models were more sensitive to the imbalanced dataset and therefore offered more robust outcomes after the application of stratified K-fold cross-validation. These data offer insight for the development of future risk stratification models derived from large clinical datasets.

The findings of this study are supported by the large number of participants, in-depth and gender-specific analysis, and high performance of the models. The robust sensitivity, precision, and F1 scores indicate the practicality of our models for the means of screening. Even though the features utilized in this study are cost-benefit, this study did not incorporate important laboratory variables including albumin and urine protein levels. Although these variables could potentially add complexity to the models, the inclusion of these elements can increase the accuracy of the results.

Conclusion

This study presents robust risk models for predicting the incidence of CKD in the Iranian population. With high sensitivity, precision, and F1-score, the models provide an in-depth prediction of CKD in the general population. The robust performance coupled with their cost-effective features indicates that the models may serve as valuable and practical screening tools in the general Iranian population.

Data availability

The authors confirm that the data supporting the findings of this study are available from the corresponding authors upon request.

Received: 1 July 2024; Accepted: 18 December 2024

Published online: 30 December 2024

References

1. Jadoul, M., Aoun, M., Masimango, M. & Imani The major global burden of chronic kidney disease. *Lancet Glob Health*. **12** (3), e342–e343 (2024).
2. Bouya, S. et al. Prevalence of chronic kidney Disease in Iranian General Population: a Meta-analysis and systematic review. *Ther. Apher Dial.* **22** (6), 594–599 (2018).
3. Vondenhoff, S., Schunk, S. J. & Noels, H. Increased cardiovascular risk in patients with chronic kidney disease. *Herz* **49** (2), 95–104 (2024).
4. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of Disease Study 2021. *Lancet*, **403**(10440): pp. 2162–2203. (2024).
5. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*, **403**(10440): pp. 2100–2132. (2024).
6. Zhang, S. et al. Global, regional, and national burden of kidney dysfunction from 1990 to 2019: a systematic analysis from the global burden of disease study 2019. *BMC Public. Health.* **23** (1), 1218 (2023).
7. Levin, A. et al. Perspectives on early detection of chronic kidney disease: the facts, the questions, and a proposed framework for 2023 and beyond. *Kidney Int.* **103** (6), 1004–1008 (2023).
8. Shlipak, M. G. et al. The case for early identification and intervention of chronic kidney disease: conclusions from a kidney disease: improving global outcomes (KDIGO) Controversies Conference. *Kidney Int.* **99** (1), 34–47 (2021).
9. Echouffo-Tcheugui, J. B. & Kengne, A. P. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med.* **9** (11), e1001344 (2012).
10. Kawasoe, S. et al. Development of a risk prediction score and equation for chronic kidney disease: a retrospective cohort study. *Sci. Rep.* **13** (1), 5001 (2023).
11. Stevens, P. E. et al. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int.* **105** (4), S117–S314 (2024).
12. Lang, S. M. & Schiff, H. Smoking status, cadmium, and chronic kidney disease. *Ren. Replace. Therapy*. **10** (1), 17 (2024).
13. Robinson-Cohen, C. et al. Physical activity and change in estimated GFR among persons with CKD. *J. Am. Soc. Nephrol.* **25** (2), 399–406 (2014).
14. Gao, J. et al. The cumulative exposure to high-sensitivity C-Reactive protein predicts the risk of chronic kidney diseases. *Kidney Blood Press. Res.* **45** (1), 84–94 (2019).
15. Delpino, F. M. et al. Machine learning for predicting chronic diseases: a systematic review. *Public. Health.* **205**, 14–25 (2022).
16. Rashid, J. et al. An augmented Artificial Intelligence Approach for Chronic diseases Prediction. *Front. Public. Health.* **10**. (2022).
17. Ghayour-Mobarhan, M. et al. Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation. *Int. J. Public. Health.* **60** (5), 561–572 (2015).
18. Norton, K. I. *Standards for Anthropometry Assessment, in Kinanthropometry and Exercise Physiologyp*. 68–137 (Routledge, 2018).
19. World Health, O. *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications: Report of a WHO Consultation. Part 1, Diagnosis and Classification of Diabetes Mellitus* (World health organization, 1999).
20. World Health, O. & International Society of Hypertension Writing. and G., *World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension*. *Journal of hypertension*, **2003**. **21**(11): pp. 1983–1992. (2003).
21. Gerrior, S., Juan, W. & Basiotis, P. An easy approach to calculating estimated energy requirements. *Prev. Chronic Dis.* **3** (4), A129 (2006).
22. Nejadghaderi, S. A. et al. Trends of national and sub-national burden attributed to kidney dysfunction risk factor in Iran: 1990–2019. *Frontiers in Endocrinology*, **14**. (2023).

23. Levey, A. S. et al. Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). *Kidney Int.* **67** (6), 2089–2100 (2005).
24. Tarwater, K. Estimated glomerular filtration rate explained. *Mo Med.* **108** (1), 29–32 (2011).
25. Widodo, S., Brawijaya, H. & Samudi, S. *Stratified K-fold cross validation optimization on machine learning for prediction*. Sinkron: jurnal dan penelitian teknik informatika, **6**(4): pp. 2407–2414. (2022).
26. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
27. Myers, J. et al. A neural network approach to predicting outcomes in heart failure using cardiopulmonary exercise testing. *Int. J. Cardiol.* **171** (2), 265–269 (2014).
28. Sazli, M. H. *A brief review of feed-forward neural networks*. Communications Faculty of Sciences University of Ankara Series A2- A3. *Phys. Sci. Eng.*, **50**(01). (2006).
29. Dines, V. A. & Garovic, V. D. Menopause and chronic kidney disease. *Nat. Rev. Nephrol.* **20** (1), 4–5 (2024).
30. Senan, E. M. et al. *Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques*. Journal of Healthcare Engineering, 2021. (2021).
31. Singh, V., Asari, V. K. & Rajasekaran, R. *A Deep Neural Network for Early Detection and prediction of chronic kidney disease*. *Diagnostics*, **12**. (2022).
32. Jeong, B. et al. Comparison between statistical models and machine learning methods on classification for highly imbalanced multiclass kidney data. *Diagnostics* **10** <https://doi.org/10.3390/diagnostics10060415> (2020).
33. Choi, Y. et al. Racial differences and Contributory Cardiovascular and non-cardiovascular risk factors towards chronic kidney Disease Progression. *Vasc Health Risk Manag.* **19**, 433–445 (2023).
34. Wang, F. et al. Prevalence and risk factors for CKD: a comparison between the adult populations in China and the United States. *Kidney Int. Rep.* **3** (5), 1135–1143 (2018).
35. Herrington, W. G. et al. Body-mass index and risk of advanced chronic kidney disease: prospective analyses from a primary care cohort of 1.4 million adults in England. *PLoS One.* **12** (3), e0173515 (2017).
36. Ting, S. M. S. et al. Overweight, obesity and chronic kidney disease. *Nephron Clin. Pract.* **112** (3), c121–c127 (2009).
37. Garofalo, C. et al. A systematic review and meta-analysis suggests obesity predicts onset of chronic kidney disease in the general population. *Kidney Int.* **91** (5), 1224–1235 (2017).
38. Ouyang, Y. et al. Underweight is an independent risk factor for renal function deterioration in patients with IgA nephropathy. *PLoS One.* **11** (9), e0162044 (2016).
39. Ricardo, A. C. et al. Healthy lifestyle and risk of kidney disease progression, atherosclerotic events, and death in CKD: findings from the chronic renal insufficiency cohort (CRIC) study. *Am. J. Kidney Dis.* **65** (3), 412–424 (2015).
40. Chiu, Y. L. et al. *Health data-driven machine learning algorithms applied to risk indicators assessment for chronic kidney disease. Risk Manage. Healthc. Policy*, : pp. 4401–4412. (2021).
41. Djalalinia, S. et al. Patterns of obesity and overweight in the Iranian Population: findings of STEPs 2016. *Front. Endocrinol. (Lausanne)*, **11**, 42 (2020).
42. Weinstein, J. R. & Anderson, S. The aging kidney: physiological changes. *Adv. Chronic Kidney Dis.* **17** (4), 302–307 (2010).
43. Li, L. et al. Is hyperuricemia an independent risk factor for new-onset chronic kidney disease? A systematic review and meta-analysis based on observational cohort studies. *BMC Nephrol.* **15** (1), 122 (2014).
44. Gherghina, M. E. et al. Uric acid and oxidative stress—relationship with Cardiovascular, metabolic, and renal impairment. *Int. J. Mol. Sci.* **23** <https://doi.org/10.3390/ijms23063188> (2022).
45. Nakayama, S. et al. Detailed association between serum uric acid levels and the incidence of chronic kidney disease stratified by sex in middle-aged adults. *Atherosclerosis* **330**, 107–113 (2021).
46. Mori, K. et al. U-shaped relationship between serum uric acid level and decline in renal function during a 10-year period in female subjects: BOREAS-CKD2. *Hypertens. Res.* **44** (1), 107–116 (2021).
47. Li, G. et al. *The dose-response relationship between sex hormones and hyperuricemia in different gender: NHANES 2013–2016*. (2022). Frontiers in Endocrinology, **13**.

Author contributions

All authors have read and approved the manuscript. MA, MGM, and HE contributed to the Study concept and design. MA, MGM, and HE contributed to methodology of the projects and supervised the project. PBT and DT contributed to analyzing and interpreting the data and drafted the manuscript. MGM, and HE critically revised the manuscript for important intellectual content.

Funding

The collecting of clinical data was financially supported by Mashhad University of Medical Sciences.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83973-5>.

Correspondence and requests for materials should be addressed to M.G.-M. or H.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024