Original Research Article

Machine-learning prediction of kidney failure occurrence based on regular health check-up data: a nationwide cohort dataset in South Korea

Gahee Lee[1#], Seokjun Kim[2#], Seohyun Hong[2#], Soo-Young Yoon[3], Hyeon Seok Hwang[3], Ai Koyanagi[4], Lee Smith[5], Hayeon Lee[1], Jinseok Lee[1*]

[1]Department of Biomedical Engineering, Kyung Hee University, Yongin, South Korea
[2]Department of Medicine, Kyung Hee University College of Medicine, Seoul, South Korea
[3]Division of Nephrology, Department of Internal Medicine, Kyung Hee University Medical Center, Kyung Hee University College of Medicine, Seoul, Korea
[4]Research and Development Unit, Parc Sanitari Sant Joan de Deu, Barcelona, Spain
[5]Centre for Health, Performance and Wellbeing, Anglia Ruskin University, Cambridge, UK

Abstract

Objective: While prior research developed numerous machine learning (ML) models for disease prediction and demonstrated their potential, there has yet to be research on models concerning the prediction of kidney failure and its associated risk factors using simple, regular health check-up data. Thus, this study aimed to develop a ML model capable of predicting the risk of kidney failure within five years after receiving regular health check-ups and to identify key contributory factors in the prediction of kidney failure.

Methods: We utilized the National Health Insurance Service-National Sample Cohort (NHIS-NSC; n=973,303) to predict kidney failure. The outcome of interest was the risk of kidney failure onset within five years after receiving regular health check-ups. We evaluated the performance of various ML models, including ensemble models, and identified the model with the best predictive capability. Additionally, we conducted a feature importance analysis and ranked the relative significance of each feature.

Results: The final dataset used in this study consisted of 1,048,422 cases using multiple records of participants' annual health check-ups, including patients with kidney failure (n=13,156 [1.27%]). The best-performing model was the double-ensemble, which includes all features and excludes age, consisting of ridge regression and LightGBM, with an AUROC of 0.754, accuracy of 0.693, specificity of 0.693, sensitivity of 0.691, and balanced accuracy of 0.692 on the test dataset. Finally, the five most important features in predicting kidney failure were age, body mass index, fasting blood glucose, diastolic blood pressure, and systolic blood pressure.

Conclusions: The study emphasizes the potential of ML models in predicting kidney failure occurrences within five years after annual health check-up data, encouraging broader implementation of such models to enhance public health and decrease the prevalence of kidney failure by preventive intervention. Additionally, the feature importance list may provide valuable insight to clinicians for proactive action in relation to kidney failure.

Keywords: machine learning, kidney failure, prediction, LightGBM, logistic regression

1. Introduction

Kidneys play a pivotal role in the human body as essential regulators of waste filtration, fluid balance, and electrolyte equilibrium. Kidney failure, characterized by the loss of renal function, stands as a multifaceted and pressing global health concern, impacting millions worldwide.[1] Kidney failure adversely impacts health-related life quality, encompassing physical, emotional, and social well-being.[2] Furthermore, prevailing trends indicate an

*Correspondence:
Jinseok Lee
E-mail: gonasago@khu.ac.kr

ORCID
Jinseok Lee
https://orcid.org/0000-0002-8580-490X

increasing incidence of chronic and end-stage kidney disease.[3] Thus, there have been a demand for studies focusing on this issue from various perspectives, particularly those involving the prediction of kidney failure occurrence.

The prediction of kidney failure necessitates the incorporation of numerous variables. Consequently, developing a comprehensive predictive program is challenging due to its inherent complexity. Machine learning (ML) has been increasingly recognized as a viable approach to enhance the precision of kidney failure prediction and has already demonstrated its potential in predicting various diseases.[4-6] In this context, our specific objective was to predict kidney failure utilizing nationwide, readily accessible regular health check-up data from individuals in South Korea by the development and assessment of multiple ML models to identify the optimal predictive model. Additionally, while age is commonly assumed as a prominent factor in predicting kidney failure, this study endeavored to provide a comprehensive perspective on the relevance of other factors and to compare their significance with that of age.[7, 8]

2. Methods

2.1 Study population and data sources

We used national, large-scale, and general-population-based cohort data from South Korea; The National Health Insurance Service-National Sample Cohort (NHIS-NSC; total n=973,303).[9-11] The cohort was built and provided by the NHIS of South Korea based on participants aged 20 years and above between 1 January 2002 and 31 December 2013. The results of nationwide regular health check-ups were used. This cohort study has the following advantageous characteristics: (1) The Korean government has been conducting mandatory regular health check-ups on all Koreans aged 20 years and above every year; (2) It records whether the individual has been diagnosed with kidney failure every year; (3) To ensure confidentiality, the Korean government anonymized all individual-related information.[9] Participants with insufficient information, who died, or with unknown kidney failure status were excluded. Figure S1 illustrates the data selection process from NHIS-NSC. The research protocol was authorized by the Institutional Review Board of Kyung Hee University and NHIS of Korea, and written informed consent was waived by the ethics committee, owing to the routinely collected data.

2.2 Covariate definitions

To develop a ML model capable of predicting kidney failure within five years, we acquired several variables as follows: age (continuous); sex (male and female); household income (basic livelihood recipient, 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, and 90-100 percentile); region of residence (rural and urban); body mass index (BMI; continuous)[12]; systolic blood pressure (continuous); diastolic blood pressure (continuous); fasting blood glucose (FBG; continuous); serum total cholesterol (continuous); hemoglobin (continuous); aspartate transaminase (continuous); alanine transaminase (continuous); γ-glutamyl transpeptidase (continuous); history of kidney related diseases (Table S1); history of diabetes mellitus, stroke, and hypertension; smoking status (never, ex-, and current smoker); alcoholic drink consumption (<1, 1-2, 3-4, and ≥5 days per week); and moderate-to-vigorous physical activity (0, 1-2, 3-4, 5-6, and 7 days per week). All included features for the artificial

intelligence (AI) model are summarized in Table S2.

2.3 Patient and public involvement

None of the patients were directly involved in designing the research questions or conducting the research. Patients were not asked for advice on the interpretation or writing of the results. There were no plans to involve patients or the relevant patient community in the dissemination of study findings.

2.4 Model development

In this study, we split the multiplied NHIS-NSC dataset (n=1,048,422) into training (n=838,562) and testing (n=209,860) datasets with a ratio of 8:2 in a stratified fashion. Table S3 summarizes the distribution of training and testing data. The testing set was used exclusively for conducting an independent test of our developed AI models and was not utilized for training or internal validation.

We first performed five-fold cross-validation using the training data to confirm its generalizability. The training dataset (n=838,562) was randomly shuffled and stratified into five equal groups, four groups were selected for training the model, and the remaining group was used for internal validation. This process was repeated five times by shifting the internal validation group for the cross-validation. Since the number of kidney failure data (n=4,844; 1.15%) was even lower than that of non-kidney failure data (n=416,077; 98.75%), we up-sampled the kidney failure data using a synthetic minority oversampling technique during the model update.[13] By balancing the two groups, we aimed to prevent bias toward the kidney failure group.

We initially applied six distinct ML algorithms: extreme gradient boosting, gradient boosting machine (GBM), light gradient-boosting machine (LightGBM), random forest, adaptive boosting, and logistic regression (LR). Subsequently, we identified the top-performing three models (GBM, LightGBM, and LR) among the six models and applied an ensemble approach by considering every possible combination. For the robustness of our model, an ensemble model was adopted.

We selected a best-achievement model among the ensemble models, which was the combination of LightGBM and LR. To attenuate age-centric bias in the predictive model,[7] we undertook an ensemble approach combining two ensemble models: the model with all features and the model without age. Fig. 1 illustrates the double-ensemble model designed to predict kidney failure within five years after receiving regular health check-ups. Each single-ensemble model provided the probabilities: p_{all} from the general model and p_{ex_age} from the model excluding age. Subsequently, p_{all} and p_{ex_age} were multiplied by model weight values, w_{all} and w_{ex_age}, respectively. To find the weight values, w_{all} and w_{ex_age}, we investigated the prediction performance by changing w_{all} and w_{ex_age}. Finally, the sum of the two multiplications was considered as the final probability for kidney failure.

The evaluation of model performance was based on five metrics, including sensitivity, specificity, accuracy, balanced accuracy, and area under receiver of characteristics (AUROC) from five-fold cross-validation. Especially due to severe data imbalance, we focused on balanced accuracy and AUROC for evaluation metrics of the main model (single-ensemble model with all features). Based on the final model, we further presented the relative importance
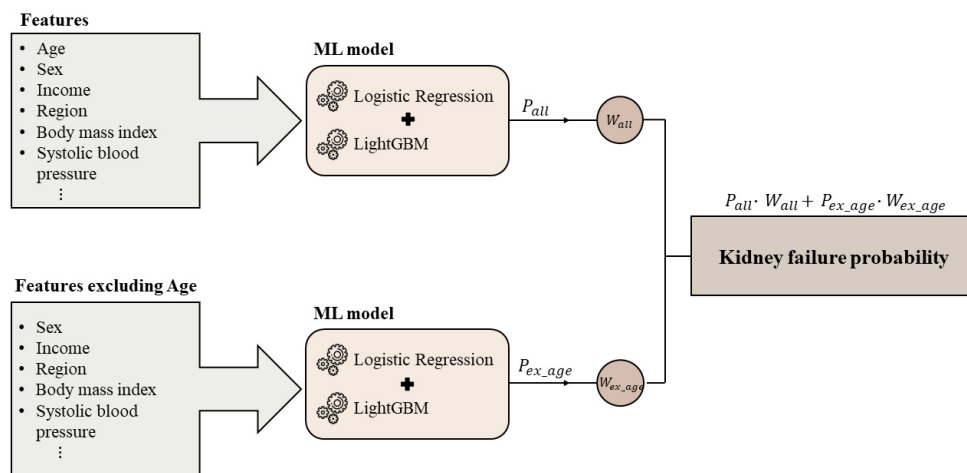
Fig. 1. Our ensemble proposed model to predict kidney failure within five years after receiving regular health check-ups.

of individual features, listing features in the order they contributed to the prediction of kidney failure occurrence within five years after receiving regular health check-ups.

We implemented the models using Python (version: 3.9.16; Python Software Foundation, Wilmington, DE, USA) with TensorFlow (version: 2.9.1), Keras (version: 2.9.0), NumPy (version: 1.21.5), Pandas (version: 1.4.4), Matplotlib (version: 3.5.2), and Scikit-learn (version: 1.0.2). All statistical analyses were performed using SAS (version 9.4; SAS Institute Inc., Cary, NC, USA).[14, 15]

3. Results

3.1 Baseline characteristics of NHIS-NSC cohort

There are 4,844 participants with kidney failure (mean age 60.1 [standard deviation, 14.0] years; 2,274 [46.9%] male) and 416,077 those without kidney failure (mean age 50.2 [standard deviation, 14.6] years; 217,345 [52.2%] male). Table 1 depicts a comparison of variables included in the NHIS-NSC.

3.2 Model performance of single-ensemble model

Table 2 summarizes the five-fold cross-validation result in comparison to other ML models. Regarding the balanced accuracy and AUROC, when considering only single models, the most noteworthy performance was demonstrated by LR (balanced accuracy, 0.691 and AUROC, 0.754), LightGBM (balanced accuracy, 0.680 and AUROC, 0.740), and GBM (balance accuracy, 0.680 and AUROC, 0.739). To further elevate prediction performance, we investigated the ensemble approach using the possible combinations of the top three single models: LR, LightGBM, and GBM. The results showed that the ensemble model with the combination of LightGBM and LR provided the highest performance than any other single or combination models. For LightGBM, we found the following optimized hyper-parameters: number of tree estimators with 300, learning rate with 0.01, the tree depth explicitly with 2, and a minimal amount of data in one leaf with 2. For LR, penalty norm with L2 (ridge), the inverse of regularization strength with 10, and maximum number of iterations with 100.

Table 1. Comparison of included variables in the discovery cohort (NHIS-NSC; n=420,921) to predict kidney failure occurrence within five years after receiving regular health check-ups

Variables	Kidney failure (n=4,844)	Non-kidney failure (n=416,077)
Age, years, mean (SD)	60.1 (14.0)	50.2 (14.6)
Sex, n (%)		
Male	2,274 (46.9)	217,345 (52.2)
Female	2,570 (53.1)	198,732 (47.8)
Region of residence, n (%)		
Urban	2,199 (45.4)	193,507 (46.5)
Rural	2,645 (54.6)	222,570 (53.5)
Household income, n (%)		
0	36 (0.7)	1,056 (0.3)
1	441 (9.1)	35,170 (8.5)
2	785 (16.2)	47,261 (11.4)
3	363 (7.5)	33,508 (8.1)
4	329 (6.8)	37,808 (9.1)
5	392 (8.1)	41,702 (10.0)
6	425 (8.8)	41,915 (10.1)
7	414 (8.6)	43,344 (10.4)
8	435 (9.0)	42,895 (10.3)
9	550 (11.4)	44,188 (10.6)
10	674 (13.9)	47,230 (11.4)
Systolic blood pressure, mmHg, mean (SD)	131.1 (19.8)	123.4 (16.9)
Diastolic blood pressure, mmHg, mean (SD)	80.9 (12.5)	77.2 (11.1)
Fasting blood glucose, mg/dL, mean (SD)	107.2 (44.7)	95.1 (28.0)
Serum total cholesterol, mg/dL, mean (SD)	200.6 (42.8)	192.8 (37.7)
Hemoglobin, g/dL, mean (SD)	13.5 (1.7)	13.9 (1.6)
Aspartate transaminase, U/L, mean (SD)	27 (18.4)	25.5 (16.9)
Alanine transaminase, U/L, mean (SD)	26.2 (23.2)	25.1 (22.3)
γ-Glutamyl transpeptidase, U/L, mean (SD)	38.4 (53.6)	34.5 (48.3)
Body mass index, kg/m^2, mean (SD)	24.4 (3.4)	23.5 (3.2)
History of chronic kidney disease related diagnosis, n (%)	333 (6.9)	4,795 (1.2)
History of diabetes mellitus, n (%)	656 (13.5)	13,524 (3.3)
History of stroke, n (%)	50 (1.0)	2,243 (0.5)
History of hypertension, n (%)	1,146 (23.7)	28,086 (6.8)
Smoking, n (%)		
Never smoker	3,700 (76.4)	293,398 (70.5)
Ex-smoker	250 (5.2)	15,800 (3.8)
Current smoker	894 (18.5)	106,879 (25.7)
Alcoholic drinks per week, n (%)		
<1	3,883 (80.2)	297,454 (71.5)

Table 1. Continued

	Total (n=420,921)	
1-2	567 (11.7)	79,161 (19.0)
3-4	226 (4.7)	26,605 (6.4)
≥5	168 (3.5)	12,857 (3.1)
Physical activity sessions per week, n (%)		
0	3,055 (63.1)	242,712 (58.3)
1-2	926 (19.1)	103,566 (24.9)
3-4	382 (7.9)	38,016 (9.1)
5-6	115 (2.4)	9,941 (2.4)
7	366 (7.6)	21,842 (5.3)

Abbreviation: SD, standard deviation.

Table 2. Five-fold cross-validation result comparison to other ML models with all features

Model	Matrix, mean (SD)				
	Accuracy	Specificity	Sensitivity	Balanced accuracy	AUROC
XGB	0.6429 (0.0050)	0.6421 (0.0052)	0.7048 (0.0151)	0.6734 (0.0051)	0.7298 (0.0051)
GBM	0.6590 (0.0065)	0.6585 (0.0067)	0.7007 (0.0093)	0.6796 (0.0051)	0.7388 (0.0050)
LGBM	0.6542 (0.0066)	0.6535 (0.0066)	0.7069 (0.0086)	0.6802 (0.0055)	0.7401 (0.0055)
Random forest	0.6232 (0.0208)	0.6220 (0.0212)	0.7191 (0.0099)	0.6706 (0.0064)	0.7284 (0.0043)
AdaBoost	0.6434 (0.0032)	0.6427 (0.0033)	0.6982 (0.0152)	0.6705 (0.0077)	0.7259 (0.0065)
LR	0.6460 (0.0017)	0.6448 (0.0019)	0.7363 (0.0175)	0.6906 (0.0082)	0.7538 (0.0091)
GBM + LGBM	0.6565 (0.0052)	0.6559 (0.0053)	0.7043 (0.0080)	0.6801 (0.0050)	0.7395 (0.0052)
GBM + LR	0.6952 (0.0038)	0.6953 (0.0039)	0.6857 (0.0178)	0.6905 (0.009)	0.7529 (0.0079)
LGBM + LR	**0.6823 (0.0038)**	**0.682 (0.0039)**	**0.6996 (0.0164)**	**0.6908 (0.0081)**	**0.7530 (0.0080)**
GBM + LGBM + LR	0.6892 (0.0046)	0.6893 (0.0047)	0.6879 (0.0152)	0.6886 (0.0076)	0.7503 (0.0072)

Abbreviations: SD, standard deviation; AUROC, area under receiver of characteristics; XGB, XGBoost; GBM, gradient boosting machine; LGBM, light gradient-boosting machine; AdaBoost, adaptive boosting; LR, logistic regression.
The bold characters were the best-performance model.

3.3 Double-ensemble model

Table 3 summarizes the prediction accuracy according to the values of w_{all} and w_{ex_age}. The results showed that when the weights of the model with all features and the model without the age feature were assigned in a ratio of 7:3 ($w_{all} = 0.7$ and $w_{ex\_age} = 0.3$), finally, it achieved the highest balanced accuracy of 0.691 and AUROC of 0.754. Table S4 presents the

Table 3. Model performance analysis of LGBM + LR double-ensemble model with weight values

w_{all}	w_{ex_age}	Accuracy, mean (SD)	Specificity, mean (SD)	Sensitivity, mean (SD)	Balanced accuracy, mean (SD)	AUROC, mean (SD)
1.0	0.0	0.6823 (0.0038)	0.682 (0.0039)	0.6996 (0.0164)	0.6908 (0.0081)	0.753 (0.0080)
0.9	0.1	0.6866 (0.0037)	0.6865 (0.0038)	0.6955 (0.0166)	0.691 (0.0082)	0.7538 (0.0079)
0.7	**0.3**	**0.6962 (0.0031)**	**0.6964 (0.0033)**	**0.6861 (0.0182)**	**0.6912 (0.0086)**	**0.7541 (0.0076)**
0.5	0.5	0.706 (0.0027)	0.7064 (0.0028)	0.6743 (0.0137)	0.6903 (0.0064)	0.7519 (0.0072)
0.3	0.7	0.7111 (0.0029)	0.7118 (0.0030)	0.6585 (0.0131)	0.6851 (0.0056)	0.7461 (0.0066)
0.1	0.9	0.7082 (0.0036)	0.7091 (0.0038)	0.6424 (0.0140)	0.6757 (0.0056)	0.7355 (0.0059)
0.0	1.0	0.7041 (0.0041)	0.705 (0.0043)	0.6356 (0.0120)	0.6703 (0.0043)	0.7286 (0.0056)

Abbreviations: AUROC, area under receiver of characteristics; LGBM, light gradient-boosting machine; LR, logistic regression.
The bold characters were the best-performance model.

results of the five-fold cross-validation comparison for the model considering all features, the model excluding age, and the final model combining the two.

3.4 Model performance in the testing dataset

Table 4 summarizes the model performance in the test dataset from the isolated testing dataset (n=209,860) from multiplied NHIS-NSC. The testing data results also show that the ensemble of models, which includes all features and excludes age, consisting of LightGBM and LR, provided the highest values for balanced accuracy of 0.692 and AUROC of 0.754. Similar to the cross-validation results, the ensemble model provided the best performance in the testing datasets. The similarity between the results from the cross-validation and the testing data indicates that overfitting or underfitting was minimal in the model.

Table 4. Comparison of the prediction performances of the prediction models on the testing dataset (n=209,860)

Model	Accuracy	Specificity	Sensitivity	Balanced accuracy	AUROC
The final double-ensemble model	**0.6933**	**0.6933**	**0.691**	**0.6922**	**0.7538**
GBM + LGBM	0.6529	0.6523	0.7004	0.6764	0.7421
GBM + LR	0.6925	0.6926	0.6845	0.6886	0.7530
LGBM + LR	0.6791	0.6788	0.7004	0.6896	0.7533
GBM + LGBM + LR	0.6851	0.6851	0.6853	0.6852	0.7513

Abbreviations: AUROC, area under receiver of characteristics; XGB, XGBoost; GBM, gradient boosting machine; LGBM, light gradient-boosting machine; AdaBoost, adaptive boosting; LR, logistic regression.

3.5 Feature importance analysis

To analyze the effect of each feature to predict kidney failure occurrence, we performed the feature importance analysis to confirm the contribution of each feature. Fig. 2 shows the ranked normalized feature importance from the final ensemble model with the combination of LightGBM and LR. The results show that age was the top contributor to predict kidney failure occurrence within five years. The next important features were BMI, FBG, diastolic blood pressure, and systolic blood pressure. Table S5 summarizes the complete ranked normalized feature importance values.

3.6 AI-driven web application

Our proposed ensemble model was deployed on our own public website (http://ai-wm.khu.ac.kr/KidneyFailure/), so that kidney failure onset within five years can be predicted based on regular health check-up data. The deployed web application, which provides results for prediction of kidney failure onset, is shown in Fig. 3. The web interface for entering information on 20 features from regular health check-up data is shown in Fig. 3. After entering the information in the web application, a user can immediately obtain the results for prediction of kidney failure onset with its probability, as shown in Fig. 3. In the web application, the features input by a user are encoded to the website server and immediately deleted upon generation of the prediction result, so that there is no risk of exposing information. In addition, there is no need to enter any information that would be regarded as private.

4. Discussion

4.1 Key findings

To our knowledge, this is the first ML model capable of predicting kidney failure with only regular health check-up data. In this investigation, we employed ML models to predict the onset
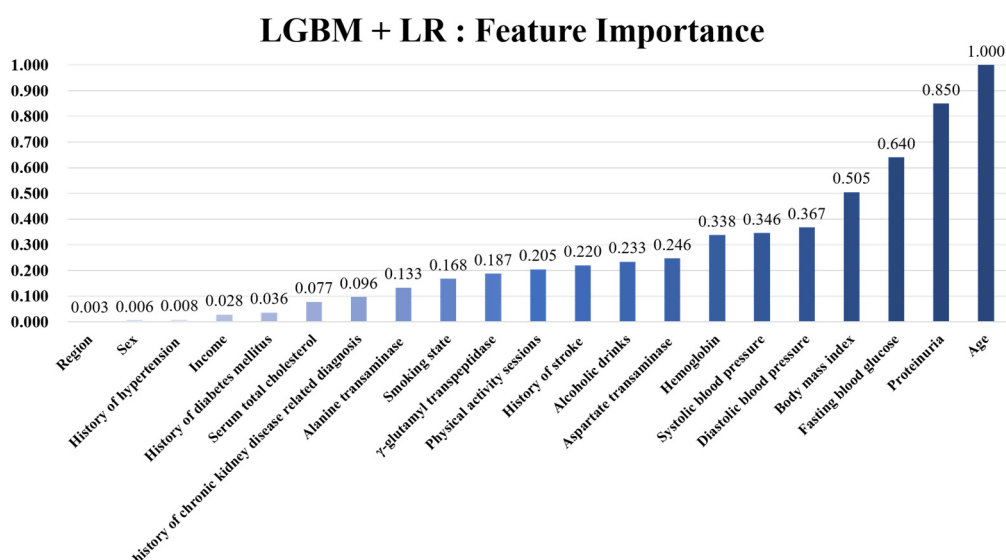

Fig. 2. Results of the ranked normalized feature importance from the ensemble model with the combination of LightGBM and LR. LGBM, light gradient-boosting machine; LR, logistic regression.

Fig. 3. Deployed web application to provide kidney failure onset within five years after receiving regular health check-ups. BMI, body mass index; SGOT_AST, aspartate transaminase; SGPT_ALT, alanine transaminase; Gamma_GTP, γ-glutamyl transpeptidase.

of kidney failure within a five-year timeframe. We evaluated each performance of 10 distinct models, including six single models and four ensemble models. Among the models, the ensemble model consisting of Light GBM and LR, which includes all features and excludes age, found the highest performance, with a balanced accuracy of 0.691 and an AUROC of 0.754 on the validation dataset and a balanced accuracy of 0.692 and an AUROC of 0.754 on the testing dataset. Feature importance analysis revealed that the five most critical predictors of kidney failure were age, FBG, systolic blood pressure, hemoglobin, and diastolic blood pressure. This feature importance analysis conducted through our model approaches highlighted features that

clinicians should pay attention to and consider as risk factors when assessing the likelihood of kidney failure occurrence.

4.2 Possible mechanisms

Previous research has explored the utilization of ML models for predicting kidney-related disease, aligning with our study's focus. Along with the positive outlook for the introduction of ML in the prediction of kidney failure,[16] several studies have demonstrated the potential of ML in predicting kidney failure. For instance, one study suggested that the ML model could predict pediatric acute kidney injury (AKI) up to 48 hours earlier than the previous diagnosis guidelines with an AUROC of 0.89 for predicting severe AKI.[17] Another study used unsupervised ML algorithms to subtype patients with CKD to enable more effective prediction by taking into account different key risk factors for each subgroup.[18] These studies as well as ours support the potential of the ML models for the prediction of kidney failures and the continued need to improve the models.

As discussed above, this study shows the continuous importance of understanding what factors are associated with the risk of kidney failure. Conventionally, age has been consistently acknowledged as a key determinant of various health issues, including kidney failure.[19] This study, while proving it, also identifies BMI and FBG as crucial factors for kidney health. A study suggested that excess weight is a common, strong, and modifiable risk factor for CKD and end-stage renal disease, showing similar results with our study.[20] Another study showed higher BMI was significantly associated with an increased risk of CKD development in hypertensive patients with normal kidney function.[21] Furthermore, a study, concluding with similar features as important, showed the association of age and BMI with kidney function and showed that a BMI of 30kg/m2 or more is associated with rapid loss of kidney function in patients with estimated glomerular filtration rate (eGFR) of at least 60mL/min per 173 m2.[22] In summary, both our study and prior research show that there is clear evidence that obesity (BMI over 30) is associated with an increased risk of kidney failure. There were studies suggesting other significant features for the diagnosis and prognosis of CKD, such as packed cell volume, specific gravity, red blood cell counts, and albumin, while there were similar features such as hemoglobin.[23] These studies show that there is a need to explain the mechanisms of suggested important features, especially for the commonly reported features in several studies.

Age, BMI, and FBG emerged as pivotal factors in predicting kidney failure, likely due to their association with diabetes and hypertension, both of which profoundly influence cardiovascular health and kidney function.[24] In contrast, factors such as sex, physical activity levels, and a history of stroke may exert a less direct influence on kidney failure or potentially affect kidney health indirectly through cardiovascular mechanisms.[25]

The primary mechanism linking BMI and FBG levels to kidney failure is their connection to type 2 diabetes.[26] High BMI often correlates with higher blood pressure and an increased likelihood of diabetes which can contribute to kidney damage. Notably, hypertension and diabetes are the leading causes of CKD according to the US Department of Health and Human Services.

FBG level is also closely associated with hypertension which can lead to the constriction of small arteries and arterioles supplying blood to the kidneys. This can compromise the eGFR.

Prolonged hypertension also exerts mechanical stress on blood vessel walls, causing vascular damage. This impairment in vascular integrity triggers inflammation and oxidative stress, exacerbating kidney deterioration. Additionally, elevated blood glucose levels can damage blood vessels within glomeruli, resulting in albuminuria, an early indicator of kidney damage.[27, 28]

Our study provides valuable insights for both clinicians and policymakers. Primarily, it underscores the significance of blood glucose and blood pressure levels in predicting kidney failure, enabling doctors to emphasize and monitor their association with kidney failure. Moreover, our model allows for the quick and efficient prediction of kidney failure using simple, regular check-up results, which can significantly aid in risk reduction through early intervention. Furthermore, given the high performance of the model developed in this study, it holds potential for commercialization as an effective predictive tool with further research and refinement. For policymakers, it is important for them to raise awareness of the contribution of features to the general population. By understanding the significance of these key features in contributing to kidney failure, individuals can proactively monitor these factors, take preventive measures, and improve their overall health through early response.

4.3 Limitations and strengths

This study has several limitations. First, although our dataset encompasses over a million cases, it primarily represents a limited Asian population. There may be unique features specific to Koreans that influence the predictive outcomes. Consequently, there is a need to assess the performance of our model using a large-scale, international external validation dataset. Second, our study did not establish a causal link between the features utilized for model training and the onset of kidney failure. Further research is required to elucidate the biological mechanisms underlying the relationships between the selected features and their influence on kidney failure.

Despite these limitations, our study provides several significant contributions. First, we employed a substantial cohort of Korean population data in our study. We utilized results from mandatory annual health check-ups imposed by the Korean government on all Koreans aged 20 and above, which ensured a robust dataset size. Utilizing a large sample size allowed us to mitigate the risk of overfitting and maintain generalizability.[29] Second, we utilized data from routine health check-ups that are conducted regularly, independent of clinical or environmental settings. This ensures the feasibility of conducting tests with high accessibility. Third, by incorporating a diverse array of features from the dataset, our study was able to offer insights into the relative importance of numerous features. This enables clinicians to identify and compare the significance of features relevant to their needs. Lastly, this model holds the potential to aid in the early detection of kidney failure. Both individuals and clinicians can assess the risk of kidney failure using only annual health check-up data, enabling prompt medical intervention. Considering the importance of early diagnosis in kidney failure, it is clear that our study offers significant value.[30]

5. Conclusion

ML is an innovative field that offers the potential to predict kidney failure with ease. In this study, we utilized a variety of features to develop several ML-based models for predicting kidney failure within five years. The ensemble models, which includes all features and excludes age, consisting of LightGBM and LR demonstrated the highest performance (validation dataset:

AUROC, 0.7541; test dataset: AUROC, 0.7538). Additionally, age, BMI FBG, diastolic blood pressure, and systolic blood pressure emerged as significant features based on their relative feature importance rankings in this model. This study holds significance as it demonstrates the feasibility of creating a highly accurate predictive model using simple and routine tests. While current national clinical guidelines for kidney failure may be cautious about adopting predictive models, such models could serve as valuable tools for preventing kidney failure in the Korean population in the future.

Capsule Summary

The study emphasizes the potential of machine learning models in predicting kidney failure occurrences within five years after annual health check-up data, encouraging broader implementation of such models to enhance public health and decrease the prevalence of kidney failure by preventive intervention.

Ethical statement

The research protocol was authorized by the Institutional Review Board of Kyung Hee University and NHIS of Korea, and written informed consent was waived by the ethics committee, owing to the routinely collected data.

Patient and public involvement

None of the patients were directly involved in designing the research questions or conducting the research. Patients were not asked for advice on the interpretation or writing of the results. There were no plans to involve patients or the relevant patient community in the dissemination of study findings.

Data Availability Statement

Data are available on reasonable request.

Transparency statement

The leading author (Dr. JL) is an honest, accurate, and transparent account of the study being reported.

Author Contribution

Dr JL had full access to all the data used in the study and took responsibility for the integrity of the data and accuracy of the data analysis. All authors approved the final version before submission. Study concept and design: GL, SK, SH, and JL; Acquisition, analysis, or interpretation of data: GL, SK, SH, HL, and JL; Drafting of the manuscript: GL, SK, SH, HL, and JL; Critical revision of the manuscript for important intellectual content: all authors; Statistical analysis: GL, SK, SH, HL, and JL; Study supervision: JL. JL supervised the study and is the guarantor for this study. GL, SK, and SH contributed equally as first authors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

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.

Provenance and peer review

Not commissioned; externally peer reviewed.

Supplementary Materials

Supplementary materials are only available online from: https://doi.org/10.54724/lc.2024.e6.

References

1. Bello AK, Levin A, Lunney M, Osman MA, Ye F, Ashuntantang GE, et al. Status of care for end stage kidney disease in countries and regions worldwide: international cross sectional survey. BMJ. 2019;367:l5873.

2. Park JI, Baek H, Jung HH. CKD and health-related quality of life: The Korea national health and nutrition examination survey. Am J Kidney Dis. 2016;67(6):851-60.

3. Oh KH, Park SK, Kim J, Ahn C. The KoreaN cohort study for outcomes in patients with chronic kidney disease (KNOW-CKD): A Korean chronic kidney disease cohort. J Prev Med Public Health. 2022;55(4):313-20.

4. Stachel A, Daniel K, Ding D, Francois F, Phillips M, Lighter J. Development and validation of a machine learning model to predict mortality risk in patients with COVID-19. BMJ Health & Care Informatics. 2021;28(1):e100235.

5. Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiology. 2021;6(6):633-41.

6. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): A modelling study of pooled datasets. Lancet. 2021;397(10270):199-207.

7. Liu P, Quinn RR, Lam NN, Elliott MJ, Xu Y, James MT, et al. Accounting for age in the definition of chronic kidney disease. JAMA Internal Medicine. 2021;181(10):1359-66.

8. Ilginel MT, Yesilyurt AO, Ozyilmaz E, Ilginel OO, Sener M, Tunay DL, et al. Evaluation of factors affecting treatment and mortality in patients over 65 years of age and without chronic disease, followed in the Intensive Care Unit due to COVID-19. Eur Rev Med Pharmacol Sci. 2023;27(17):8301-13.

9. Shin YH, Shin JI, Moon SY, Jin HY, Kim SY, Yang JM, et al. Autoimmune inflammatory rheumatic diseases and COVID-19 outcomes in South Korea: a nationwide cohort study. Lancet Rheumatol. 2021;3(10):e698-e706.

10. Yang JM, Koh HY, Moon SY, Yoo IK, Ha EK, You S, et al. Allergic disorders and susceptibility to and severity of COVID-19: A nationwide cohort study. J Allergy Clin

Immunol. 2020;146(4):790-8.

11. Lee JS, Shin JI, Kim S, Choi YS, Shin YH, Hwang J, et al. Breastfeeding and impact on childhood hospital admissions: A nationwide birth cohort in South Korea. Nat Commun. 2023;14(1):5819.

12. Eum S, Rhee SY. Age, ethnic, and sex disparity in body mass index and waist circumference: A bi-national large-scale study in South Korea and the United States. Life Cycle. 2023;3:e4.

13. Zale AD, Abusamaan MS, McGready J, Mathioudakis N. Development and validation of a machine learning model for classification of next glucose measurement in hospitalized patients. EClinicalMedicine. 2022;44:101290.

14. Cho JK, Yang H, Park J, Lee H, Nguyen A, Kattih M, et al. Association between allergic rhinitis and despair, suicidal ideation, and suicide attempts in Korean adolescents: A nationally representative study of one million adolescents. Eur Rev Med Pharmacol Sci. 2023;27(19):9248-56.

15. Kim J, Kim SC, Kang D, Kim SY, Kwon R, Yon DK, et al. Feature extraction of time series data on functional near-infrared spectroscopy and comparison of deep learning performance for classifying patients with Alzheimer's-related mild cognitive impairment: a post-hoc analysis of a diagnostic interventional trial. Eur Rev Med Pharmacol Sci. 2023;27(14):6824-30.

16. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, et al. Internal and external validation of a machine learning risk score for acute kidney injury. JAMA Network Open. 2020;3(8):e2012892-e.

17. Dong J, Feng T, Thapa-Chhetry B, Cho BG, Shum T, Inwald DP, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care. 2021;25(1):288.

18. Zheng Z, Waikar SS, Schmidt IM, Landis JR, Hsu CY, Shafi T, et al. Subtyping CKD patients by consensus clustering: the chronic renal insufficiency cohort (CRIC) Study. J Am Soc Nephrol. 2021;32(3):639-53.

19. Elliott MJ, Tam-Tham H, Hemmelgarn BR. Age and treatment of kidney failure. Curr Opin Nephrol Hypertens. 2013;22(3):344-50.

20. Nguyen S, Hsu CY. Excess weight as a risk factor for kidney failure. Curr Opin Nephrol Hypertens. 2007;16(2):71-6.

21. Xie L, Wang B, Jiang C, Zhang X, Song Y, Li Y, et al. BMI is associated with the development of chronic kidney diseases in hypertensive patients with normal renal function. J Hypertens. 2018;36(10):2085-91.

22. Lu JL, Molnar MZ, Naseer A, Mikkelsen MK, Kalantar-Zadeh K, Kovesdy CP. Association of age and BMI with kidney function and mortality: a cohort study. Lancet Diabetes Endocrinol. 2015;3(9):704-14.

23. Mahbub NI, Hasan MI, Ahamad MM, Aktar S, Moni MA, editors. Machine learning approaches to identify significant features for the diagnosis and prognosis of chronic kidney disease. 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET); 2022: IEEE.

24. Mennuni S, Rubattu S, Pierelli G, Tocci G, Fofi C, Volpe M. Hypertension and kidneys: unraveling complex molecular mechanisms underlying hypertensive renal damage. J Hum Hypertens. 2014;28(2):74-9.

25. Wilund KR, Thompson S, Viana JL, Wang AY. Physical activity and health in chronic kidney disease. Contrib Nephrol. 2021;199:43-55.

26. Kelly MS, Lewis J, Huntsberry AM, Dea L, Portillo I. Efficacy and renal outcomes of SGLT2 inhibitors in patients with type 2 diabetes and chronic kidney disease. Postgrad Med. 2019;131(1):31-42.

27. Ruilope LM, Ortiz A, Lucia A, Miranda B, Alvarez-Llamas G, Barderas MG, et al. Prevention of cardiorenal damage: importance of albuminuria. Eur Heart J. 2023;44(13):1112-23.

28. Yoon SY, Park HW, Kim HJ, Kronbichler A, Koyanagi A, Smith L, et al. National trends in the prevalence of chronic kidney disease among Korean adults, 2007-2020. Sci Rep. 2023;13(1):5831.

29. Kwon R, Lee H, Kim MS, Lee J, Yon DK. Machine learning-based prediction of suicidality in adolescents during the COVID-19 pandemic (2020-2021): Derivation and validation in two independent nationwide cohorts. Asian J Psychiatr. 2023;88:103704.

30. James MT, Hemmelgarn BR, Tonelli M. Early recognition and prevention of chronic kidney disease. Lancet. 2010;375(9722):1296-309.