







## Article

# A Hybrid Risk Factor Evaluation Scheme for Metabolic Syndrome and Stage 3 Chronic Kidney Disease Based on Multiple Machine Learning Techniques

Mao-Jhen Jhou <sup>1,†</sup> , Ming-Shu Chen <sup>2,†</sup> , Tian-Shyug Lee <sup>1,3</sup> , Chih-Te Yang <sup>4</sup> , Yen-Ling Chiu <sup>5,6,7</sup>  and Chi-Jie Lu <sup>1,3,8,\*</sup> 

- <sup>1</sup> Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City 242062, Taiwan
  - <sup>2</sup> Department of Healthcare Administration, College of Healthcare & Management, Asia Eastern University of Science and Technology, New Taipei City 220303, Taiwan
  - <sup>3</sup> Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City 242062, Taiwan
  - <sup>4</sup> Department of Business Administration, Tamkang University, New Taipei City 251301, Taiwan
  - <sup>5</sup> Department of Medical Research, Department of Medicine, Far Eastern Memorial Hospital, New Taipei City 22056, Taiwan
  - <sup>6</sup> Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei 10002, Taiwan
  - <sup>7</sup> Graduate Institute of Medicine and Graduate Program of Biomedical Informatics, Yuan Ze University, Taoyuan 32003, Taiwan
  - <sup>8</sup> Department of Information Management, Fu Jen Catholic University, New Taipei City 242062, Taiwan
- \* Correspondence: 059099@mail.fju.edu.tw; Tel.: +886-2-2905-2973
- † These authors contributed equally to this work.



**Citation:** Jhou, M.-J.; Chen, M.-S.; Lee, T.-S.; Yang, C.-T.; Chiu, Y.-L.; Lu, C.-J. A Hybrid Risk Factor Evaluation Scheme for Metabolic Syndrome and Stage 3 Chronic Kidney Disease Based on Multiple Machine Learning Techniques. *Healthcare* **2022**, *10*, 2496. <https://doi.org/10.3390/healthcare10122496>

Academic Editors: Gen-Min Lin and Daniele Giansanti

Received: 1 November 2022

Accepted: 8 December 2022

Published: 9 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** With the rapid development of medicine and technology, machine learning (ML) techniques are extensively applied to medical informatics and the suboptimal health field to identify critical predictor variables and risk factors. Metabolic syndrome (MetS) and chronic kidney disease (CKD) are important risk factors for many comorbidities and complications. Existing studies that utilize different statistical or ML algorithms to perform CKD data analysis mostly analyze the early-stage subjects directly, but few studies have discussed the predictive models and important risk factors for the stage-III CKD high-risk health screening population. The middle stages 3a and 3b of CKD indicate moderate renal failure. This study aims to construct an effective hybrid important risk factor evaluation scheme for subjects with MetS and CKD stages III based on ML predictive models. The six well-known ML techniques, namely random forest (RF), logistic regression (LGR), multivariate adaptive regression splines (MARS), extreme gradient boosting (XGBoost), gradient boosting with categorical features support (CatBoost), and a light gradient boosting machine (LightGBM), were used in the proposed scheme. The data were sourced from the Taiwan health examination indicators and the questionnaire responses of 71,108 members between 2005 and 2017. In total, 375 stage 3a CKD and 50 CKD stage 3b CKD patients were enrolled, and 33 different variables were used to evaluate potential risk factors. Based on the results, the top five important variables, namely BUN, SBP, Right Intraocular Pressure (R-IOP), RBCs, and T-Cho/HDL-C (C/H), were identified as significant variables for evaluating the subjects with MetS and CKD stage 3a or 3b.

**Keywords:** machine learning (ML); Metabolic syndrome (MetS); chronic kidney disease (CKD); end-stage kidney disease (ESKD); hybrid risk factor

## 1. Introduction

Suboptimal health status is a dynamic and intermediate bodily condition between health and disease. Various indicators of suboptimal health must be considered during the prevention of chronic diseases to achieve better health protection. Metabolic syndrome (MetS) is a collection of suboptimal health risk indicators. According to the definition provided by the Health Promotion Administration (HPA) and Ministry of Health and

Welfare (MOHW) [1], the five diagnostic criteria of MetS are excessive abdominal fat, high blood pressure, high fasting plasma glucose levels, high fasting triglycerides levels, and elevated high-density lipoprotein-cholesterol levels. A person who meets three or more of these criteria is diagnosed with MetS [1]. The accumulation of MetS risk factors increases the risk of chronic kidney disease (CKD) and other major chronic diseases [2–4].

CKD refers to the irreversible and progressive loss of kidney function caused by prolonged damage to the renal tissue for several months or years. According to the American Kidney Foundation's definition of CKD, the disease consists of the following stages: 1, 2, 3a, 3b, 4, and 5 [5,6], each of which is defined in Table 1. The middle stages, 3a and 3b, indicate moderate renal failure. Patients in CKD stage 3a are in the final stage of early CKD, while those in CKD stage 3b are in the earliest stage of end-stage kidney disease (ESKD). Thus, stage 3b is a critical point of renal function deterioration as it requires kidney dialysis. Patients in stages 3a and 3b exhibit differences in their mortality risk [7,8].

**Table 1.** Stages of CKD.

Stage	1	2	3a	3b	4	5
e-GFR value	$\geq 90$	89–60	59–45	45–30	30–15	<15 or dialysis
Description	Kidney damage with normal or e-GFR	Kidney damage with normal or mild e-GFR	Mild-moderately decreased e-GFR	Moderately-severely decreased e-GFR	Severely decreased e-GFR	Severe Renal failure

MetS and CKD are important risk factors for many comorbidities and complications [9,10]. Much research has denoted a positive correlation between MetS and CKD [11,12]. Furthermore, MetS diagnosis is an effective predictor of CKD [12]. Studies on CKD prediction have identified four major risk factors: demographic variables (e.g., age, education level), anthropometric parameters (e.g., body mass index, body fat), blood examination indicators (e.g., blood urea nitrogen, uric acid), and lifestyle habits (e.g., smoking status, alcohol consumption) [13–16]. Thus, they are often used in many studies to construct CKD analytical models through machine learning (ML)-based data analysis methods [16–20].

ML techniques are extensively applied in numerous studies on medical informatics and suboptimal health status [3,16,21–24]. They are often used to identify critical predictor variables or risk factors as they can effectively investigate the complex relationships between risk factors and outcomes, based on their promising predictive performance with vast amounts of medical data [16,22,23,25].

Because some ML techniques can identify important predictor variables, a single technique for selecting important predictor variables and risk factors may result in a localized optimal risk that generates a single ranking of the variables. A variable ensemble is often used to integrate the different variables that are selected [26]. Relevant studies have also demonstrated that using variable ensembles improves the robustness of the selected variables, compared to a single variable selection technique, and reduces the bias and variance of the results [27–30].

Existing studies that utilize ML methods to perform CKD data analysis mostly analyze the patients directly [16,17,20,31–34], and few studies have discussed the predictive models and important risk factors for CKD patients with MetS. Several studies have constructed predictive models for MetS patients, as well as their risk factors. Although patients in stages 3a and 3b of CKD vary in disease progression and mortality risk [35,36], they share highly similar clinical presentations. Thus, this study aims to examine the ML predictive models and important risk factors for CKD stages 3a and 3b patients with MetS by using ML techniques.

This study aims to construct an effective hybrid important risk factor evaluation scheme for CKD stages 3a and 3b patients with MetS, based on ML predictive models. Our study used six well-known and effective ML techniques—random forest (RF), logistic regression (LGR), multivariate adaptive regression splines (MARS), extreme gradient boost-

ing (XGBoost), gradient boosting with categorical features support (CatBoost), and light gradient boosting machine (LightGBM)—to develop ML predictive models [16,18,19,37–39]. The important risk factors identification results can provide valuable information regarding the prevention of CKD and health promotion.

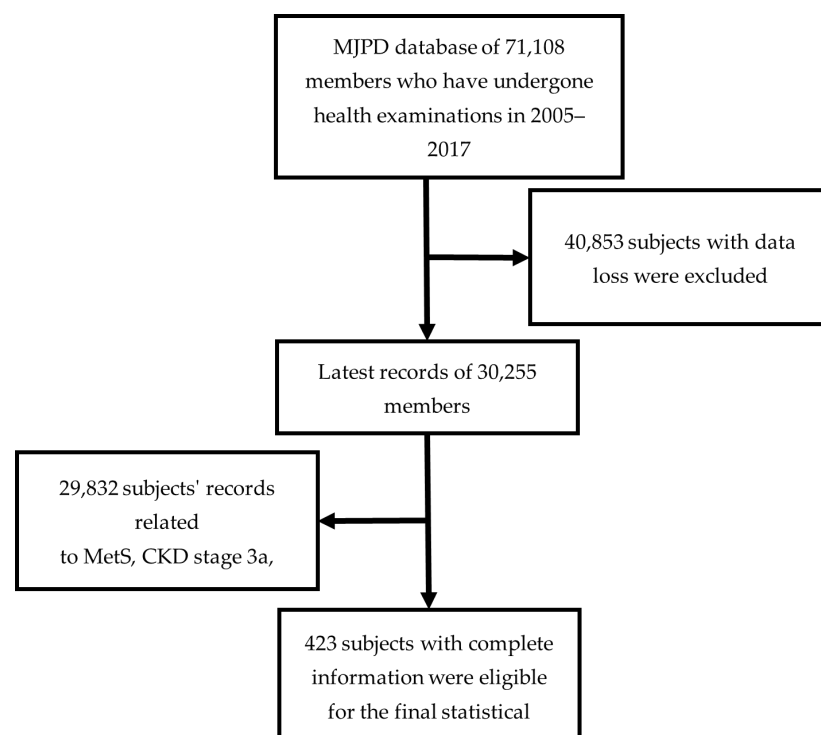
The rest of this paper is organized as follows: Section 2 describes the used materials and the proposed scheme. Section 3 presents the experiment results. Section 4 discusses the findings of the study. Finally, the study is concluded in Section 5.

## 2. Materials and Methods

### 2.1. Data

This study has selected a large database of sub-health groups in Taiwan, the MJ Health Checkup-based Population Database (MJPD, <http://www.mjhrf.org/main/page/resource/en/#resource07>, accessed on 1 August 2022). It has published more than dozens of international journal papers, including 2 JAMA and 6 Lancet journal papers. This study was approved by the institutional review board of Far Eastern Memorial Hospital (FEMH-IRB) (No: \_IRB-110027-E Approved Date: 15 February 2022) and the MJ Health Research Foundation, and registered on ClinicalTrials.gov (ID: NCT05225454).

Figure 1 shows the all-subjects identification process, and the complete data were collected from the MJPD. A total of 71,108 members from 2005 to 2017 comprised the health examination indicators and questionnaire responses. Table 2 shows the 34 health examination indicators and questionnaire variables. Among the 34 variables, CKD is the target variable and the other 33 indicators are predictor variables. Given that each member might have multiple examination records, those who had undergone multiple health examinations only had their latest records analyzed. In addition, subjects whose data had missing variables were excluded. After data processing, 30,255 subjects were eligible. We applied the MOHW's references and definitions of MetS and CKD to identify 423 MetS patients who were also diagnosed with CKD stages 3a or 3b. Table 3 presents the statistical analysis results of the participants' demographic data. A total of 375 patients (88.65%) were diagnosed with CKD stage 3a, while the remaining had CKD stage 3b.



**Figure 1.** Subject identification process.

**Table 2.** Variables for predicting the subjects with MetS and CKD stage 3a or 3b.

Abbreviation	Variables	Unit/Description
BMI	Body Mass Index	kg/m <sup>2</sup>
BF	Body Fat	% (@TANITA, DC-430MA)
WC	Waist Circumference	cm
SBP	Systolic Blood Pressure	mmHg
DBP	Diastolic Blood Pressure	mmHg
FPG	Fasting Plasma Glucose	mg/dL
L-IOP #	Left Intraocular Pressure	mmHg
R-IOP #	Right Intraocular Pressure	mmHg
r-GT	Gamma Glutamyl Transpeptidase	U/L
BUN	Blood Urea Nitrogen	mg/dL
UA	Uric Acid	mg/dL
TG	Triglyceride	mg/dL
T-Chol	Total Cholesterol	mg/dl
HDL-C	High Density Lipoprotein-Cholesterol	mg/dL
LDL-C	Low Density Lipoprotein-Cholesterol	mg/dL
C/H	T-Chol/HDL-C	the vascular risk predicts index
MS	Marital Status	(1) Single, (2) married, remarried, cohabiting, (3) divorced, (4) widowed
EL	Educational Level	(1) No formal education, (2) elementary school, (3) secondary school, (4) high school, (5) college, (6) university, (7) graduate school
FI	Yearly Family Income	(1) Unwaged, (2) NTD ≤ 200,000, (3) NTD 200,001–400,000, (4) NTD 400,001–800,000, (5) NTD 800,001–1,200,000, (6) NTD 1,200,001–1,600,000, (7) NTD 1,600,001–2,000,000; NTD: New Taiwan Dollar.
HC	Hip Circumference	cm
WHR	Waist–hip Ratio	%
LEE	Lower Extremity Edema	(1) No, (2) Yes
SGOT	Serum Glutamic-Oxaloacetic Transaminase	U/L
SGPT	Serum Glutamic-Pyruvic Transaminase	U/L
BMD	Bone Mass Density	Dual energy X-ray Absorptiometry (DEXA)
RBCs	Red Blood Cells	10 <sup>6</sup> /μl
Hb	Hemoglobin	g/dl
MCV	Mean Cell Volume	fl
UP	Urine Protein	Qualitative test, (1) non, (2) +−, (3) + (4) ++, (5) +++, (6) ++++
GU	Glucose Urine	Qualitative test, (1) non, (2) +−, (3) + (4) ++, (5) +++, (6) ++++
CS	Current Smokers	(1) Never, (2) passive smoking, (3) quit, (4) occasional, (5) addicted
AD	Alcohol Drinkers	(1) Never, (2) quit, (3) 1–2 times a week, (4) 3–4 times a week, (5) 5–6 times a week, (6) addicted
CBN	Chewing “Betel Nut”/“Areca catechu”	(1) Never, (2) quit, (3) 1–3 times a week, (4) 4–5 times a week, (5) addicted
CKD	Chronic Kidney Disease	(1) CKD stage 3a, (2) CKD stage 3b

Note: #, Intraocular pressure was measured in this database as individuals received the measurement during health check-ups.

Table 3. Subject Demographics.

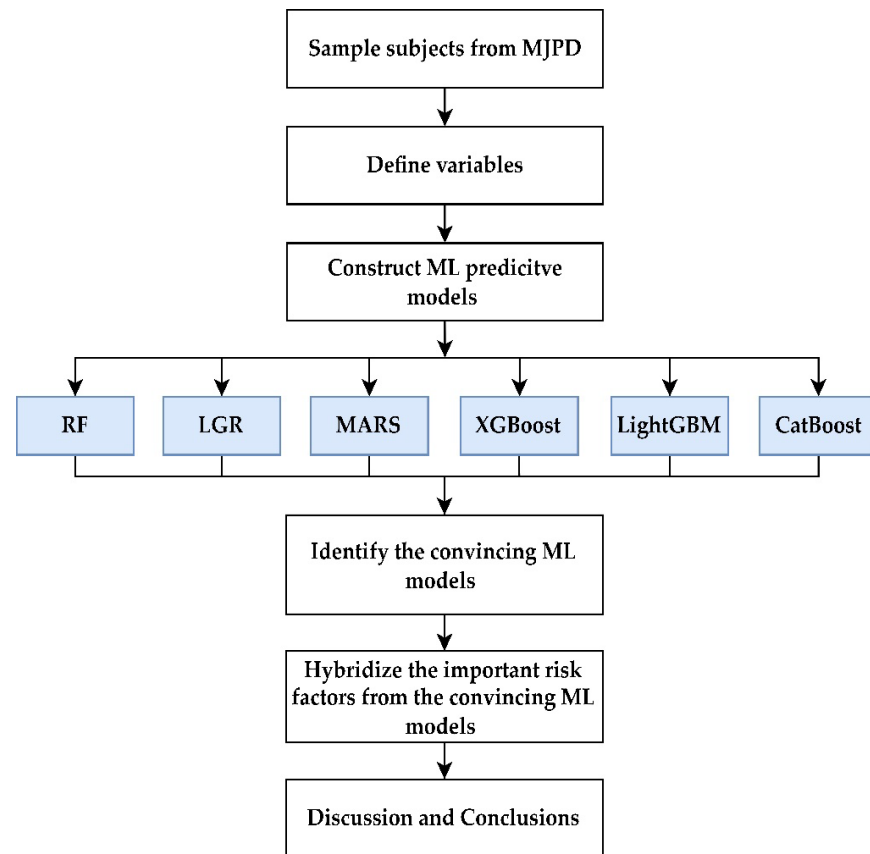
Variables		Mean $\pm$ SD		Variables	<i>n</i> (%)
	BMI	27.19 $\pm$ 3.29	FI	(1) Unwaged	78 (18.44%)
	BF	30.32 $\pm$ 6.99		(2) NTD $\leq$ 200,000	60 (14.18%)
	WC	90.21 $\pm$ 8.50		(3) NTD 200,001–400,000	87 (20.57%)
	SBP	138.18 $\pm$ 20.54		(4) NTD 400,001–800,000	82 (19.39%)
	DBP	84.30 $\pm$ 12.50		(5) NTD 800,001–1,200,000	47 (11.11%)
	FPG	123.20 $\pm$ 34.35		(6) NTD 1,200,001–1,600,000	35 (8.27%)
	L-IOP	14.60 $\pm$ 3.40		(7) NTD 1,600,001–2,000,000	34 (8.04%)
	R-IOP	14.51 $\pm$ 3.28	LEE	(1) No	418 (98.82%)
	r-GT	38.95 $\pm$ 45.14		(2) Yes	5 (1.18%)
	BUN	18.54 $\pm$ 4.74	UP	(1) non	333 (78.72%)
	UA	7.43 $\pm$ 1.64		(2) +–	34 (8.04%)
	TG	194.49 $\pm$ 78.31		(3) +	29 (6.86%)
	T-Cho	204.55 $\pm$ 37.69		(4) ++	16 (3.78%)
	HDL-C	47.36 $\pm$ 9.98		(5) +++	11 (2.60%)
	LDL-C	123.30 $\pm$ 34.39		(6) ++++	NA
C/H	4.42 $\pm$ 0.85	GU	(1) non	405 (95.74%)	
HC	99.28 $\pm$ 6.22		(2) +–	7 (1.65%)	
WHR	0.91 $\pm$ 0.06		(3) +	3 (0.71%)	
SGOT	28.61 $\pm$ 13.88		(4) ++	2 (0.47%)	
SGPT	34.19 $\pm$ 21.16		(5) +++	6 (1.42%)	
BMD	0.35 $\pm$ 1.43		(6) ++++	NA	
RBCs	4.84 $\pm$ 0.54	CS	(1) Never	292 (69.03%)	
Hb	14.48 $\pm$ 1.47		(2) Passive smoking	15 (3.55%)	
MCV	43.19 $\pm$ 4.34		(3) Quit	57 (13.48%)	
UR	1.1 $\pm$ 0.55		(4) Occasional	13 (3.07%)	
			(5) Addicted	46 (10.87%)	
Variables		<i>n</i> (%)		(1) Never	327 (77.3%)
MS	(1) Single	12 (2.84%)	AD	(2) Quit	22 (5.2%)
	(2) Married, remarried, cohabiting	332 (78.49%)		(3) 1–2 times a week	44 (10.4%)
	(3) Divorced	13 (3.07%)		(4) 3–4 times a week	17 (4.02%)
	(4) Widowed	66 (15.60%)		(5) 5–6 times a week	NA
	(1) No formal education	30 (7.09%)			(6) Addicted
EL	(2) Elementary school	101 (23.88%)	CBN	(1) Never	376 (88.89%)
	(3) Secondary school	51 (12.06%)		(2) Quit	36 (8.51%)
	(4) High school	64 (15.13%)		(3) 1–3 times a week	3 (0.71%)
	(5) College	55 (13.00%)		(4) 4–5 times a week	3 (0.71%)
	(6) University	74 (17.49%)		(5) Addicted	5 (1.18%)
		(7) Graduate school	48 (11.35%)	CKD	(1) CKD stage 3a
		(2) CKD stage 3b	48 (11.35%)		

Note: BMI, body mass index; BF, body fat; WC, waist circumference; SBP, systolic blood pressure; DBP, diastolic blood pressure; FPG, fasting plasma glucose; L-IOP, left intraocular pressure; R-IOP, right intraocular pressure; r-GT, gamma glutamyl transpeptidase; BUN, blood urea nitrogen; UA, uric acid; TG, triglyceride; T-Cho, total cholesterol; HDL-C, high density lipoprotein-cholesterol; LDL-C, low density lipoprotein-cholesterol; CH, T-Cho/HDL-C; MS, marital status; EL, educational level; FI, family income; HC, hip circumference; WHR, waist-hip ratio; LEE, lower extremity edema; SGOT, serum glutamic-oxaloacetic transaminase; SGPT, serum glutamic-pyruvic transaminase; BMD, bone mass density; RBCs, red blood cells; Hb, hemoglobin; MCV, mean cell volume; UP, urine protein; GU, glucose urine; CS, current smokers; AD, alcohol drinkers; CBN, chewing “betel nut”/“areca catechu”; CKD, chronic kidney disease.

## 2.2. Proposed Hybrid Risk Factor Evaluation Scheme

On the basis of the six ML methods, including RF, LGR, MARS, XGBoost, CatBoost, and LightGBM, this study developed a hybrid important risk factor identification scheme for the subjects with MetS and CKD stage 3a or 3b. The six ML methods used are based on different concepts and characteristics to develop the classification models [40–45]. RF,

XGBoost, CatBoost, and LightGBM are tree-based algorithms. LGR and MARS are non-parametric methods. Since they are based on different characteristics to construct effective algorithms and identify important risk factors for medical data analysis, the important variables identification results of the six methods are integrated to provide more stable and robust results. Figure 2 shows the proposed hybrid risk factor evaluation scheme.



**Figure 2.** Proposed hybrid risk factor evaluation scheme.

As shown in Figure 2, the first step was to sample the MetS subjects who were diagnosed with CKD stage 3a or 3b from through the MJPDP health examination database. Next, we defined the predictor variables and target variable. We used 33 risk factors as our predictor variables and CKD as the target variable. After consolidating the data, we built the RF, LGR, MARS, XGBoost, CatBoost, and LightGBM predictive models.

RF is a decision tree approach based on ensemble technology [40]. Its principle is to construct several unpruned decision trees, aggregate all the trees into a forest, and then generate the final model by taking the majority vote or average value of the trees. LGR is the typically most used ML method that generalizes linear models with canonical link functions [41]. Its aim is to minimize the relative cost function using a logistic function and perform model fitting using a maximum likelihood function.

MARS is a nonparametric and nonlinear statistical method in which several linear segments with different gradients are used to automatically examine the nonlinearity and dependency between multidimensional input and output variables, and then generate the final optimum nonlinear prediction model [42]. XGBoost is a decision tree-based approach that applies gradient boosting to generate multiple weak models. When each weak model is generated, the defects or shortcomings of the previous model are corrected. Finally, accuracy categorization is achieved by aggregating all the generated weak models [43].

LightGBM is a decision tree-based distributed gradient boosting framework that utilizes advanced histograms. In an iteration, it learns the approximate value of decision



tree residuals based on one-side sampling and negative gradient fitting [44]. CatBoost is a gradient-boosting decision tree technique in which sequential boosting methods are combined with gradient boosting and multiple categorical features [45]. In CatBoost, the tree combinations and categorical features generated through gradient boosting are aggregated into a sequence to generate the final model.

For constructing each ML model, we randomly divided the whole dataset into 80% training data set and 20% testing dataset. The ten-fold cross validation (CV) method was used to perform hyperparameter tuning. The selected final model is the model of the best hyperparameter configuration. This process was performed ten times.

Balanced accuracy (BA), sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC), are four well-known metrics [46–48] utilized to assess the six ML models' performance. To identify the convincing ML models, the widely used LGR was viewed as the baseline model in this study. The ML model's performance that is greater than or equal to that of the LGR model is considered as the convincing model.

To rank the importance of each predictor variable, we applied the “caret” R package of version 6.0-90 [49] to each of the six methods to produce each variable's importance value. In each model, the most important predictor variable is set as ranking 1, whereas the least important predictor variable is defined as 33, because we used 33 predictor variables in this study. Different ML methods may produce different importance rankings for each predictor variable, due to their different specific characteristics. To obtain more stable and integrable ranking results, we hybridized the importance of each variable by averaging its ranking values from the convincing ML models.

In the last step, the research findings regarding the identified important risk factors were discussed to present the conclusions of our study.

This study utilized the RStudio of version 1.1.453 and R programming language of version 3.6.2 for modeling (<http://www.R-project.org>, accessed on 1 September 2022; <https://www.rstudio.com/products/rstudio/>, accessed on 1 September 2022). In addition, each model was constructed using an R-based software package. Model construction in RF, LGR, MARS, XGBoost, CatBoost, and LightGBM was through randomForest version 4.7-1.1 [50], stats version 3.6.2, earth version 5.3.1 [51], XGBoost version 1.6.0.1 [52], catboost version 0.25.1 [53], and lightgbm version 3.3.2 [54], respectively. During the model construction process, the best hyperparameter was found using caret version 6.0-93 [49].

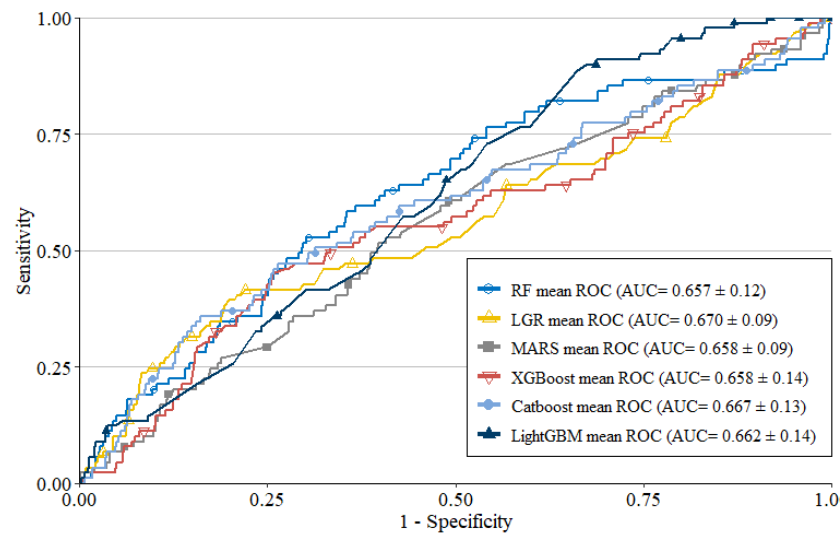
### 3. Results

This study applied six ML techniques, including RF, LGR, MARS, XGBoost, CatBoost, and LightGBM, to build predictive models for patients with MetS and CKD stage 3. Table 4 depicts the mean prediction performances of the six models after ten learning cycles, as well as the means and standard deviations (SDs) of the four performance metrics used. Figure 3 demonstrates the ROC curves of the six models. From Table 4, it can be observed that the prediction performances of the six models were similar, and the AUC of each model was greater than 0.657. The LGR has the highest AUC value of 0.670 and the RF has the lowest AUC value of 0.657.

**Table 4.** Model performance in predicting the subjects with MetS and CKD stage 3a or 3b.

Methods	Balanced Accuracy Mean (SD)	Sensitivity Mean (SD)	Specificity Mean (SD)	AUC Mean (SD)
RF	0.698 (0.09)	0.697 (0.21)	0.700 (0.17)	0.657 (0.12)
LGR	0.719 (0.06)	0.678 (0.19)	0.761 (0.25)	0.670 (0.09)
MARS	0.690 (0.07)	0.774 (0.21)	0.606 (0.26)	0.658 (0.09)
XGBoost	0.685 (0.09)	0.615 (0.16)	0.755 (0.16)	0.658 (0.14)
CatBoost	0.710 (0.19)	0.698 (0.17)	0.722 (0.18)	0.667 (0.13)
LightGBM	0.660 (0.12)	0.624 (0.26)	0.697 (0.27)	0.662 (0.14)

RF—random forest; LGR—logistic regression; MARS—multivariate adaptive regression splines; XGBoost—extreme gradient boosting; LightGBM—Light Gradient Boosting Machine; CatBoost—Gradient Boosting with Categorical Features Support.



**Figure 3.** ROC curves of the six methods.

To evaluate the performance of the six methods, DeLong's test was used since it is one of the effective tests employed to evaluate the statistically significant difference between two models' AUC values [55]. We used DeLong's test to compare AUC values between the model with the highest AUC model (i.e., LGR model in this study) to each of the remaining five ML methods. Table 5 depicts the results of DeLong's test. It can be determined from the table that the performance difference between the LGR model and each ML method is not significant, since all  $p$ -values are greater than 0.05. Therefore, the six models' prediction performances were alike and can be viewed as the convincing models. However, it is still worth noting that, from Table 4, the LGR was relatively the best ML model in this study because it can generate the highest mean balanced accuracy, specificity, and AUC values of 0.719, 0.761, and 0.670, respectively.

**Table 5.** DeLong's test between LGR and the five ML methods of this study.

	RF	MARS	XGBoost	CatBoost	LightGBM
LGR	1.258 (0.208)	1.402 (0.160)	1.615 (0.106)	0.693 (0.488)	0.184 (0.853)

Note: The numbers in parentheses are the corresponding  $p$ -value; \*\*:  $p < 0.05$ .

Because the six methods used are all considered to be convincing models, we used the variable importance generated by all six methods as the basis for our risk factor ensemble.

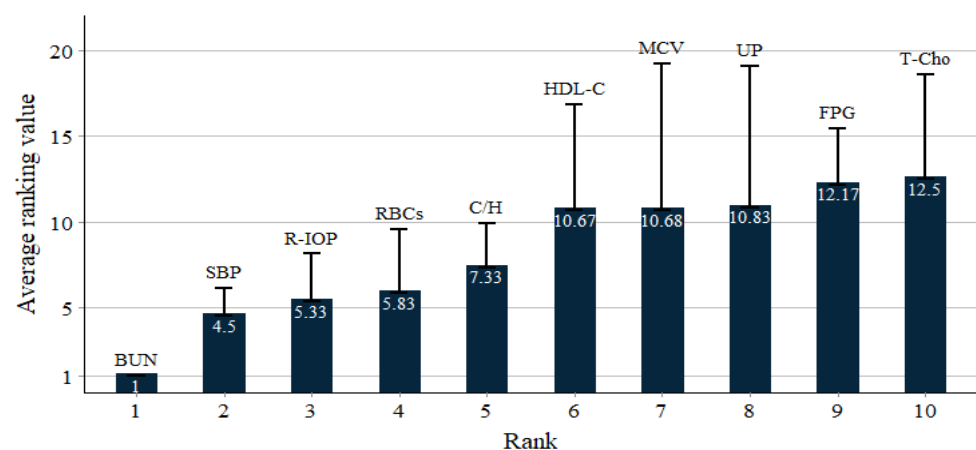
Table 6 shows the overall importance ranking of each predictor variable based on the six convincing models. Note that only the first 15 variables of Table 2 are shown. The "Average ranking of RF" to "Average ranking of LightGBM" are the average rankings, with the modeling of each of the six models repeated ten times. The different models produced different variable importance ranking results based on their modeling rules. In order to hybridize the findings of the six models, we summarized the ranking results of the six models equally in the proposed scheme. We obtained the "Average ranking of the six models" with simple averaging values from the six models.

To clarify the ranking, Figure 4 shows the ranked top ten important variables by increasing order of the average ranking values of the six models. From Figure 4, to compactly discuss the important predictor variables, based on physicians' recommendations, the top five important predictor variables, namely BUN, SBP, R-IOP, RBCs and C/H, were identified as significant variables for assessing the subjects with MetS and CKD stage 3a or 3b.



**Table 6.** Overall importance ranking of each predictor variable (only the first 15 variables of Table 2 shown).

Variables	Average Ranking of RF	Average Ranking of LGR	Average Ranking of MARS	Average Ranking of XGBoost	Average Ranking of CatBoost	Average Ranking of LightGBM	Average Ranking of the Six Models (SD)
BMI	14	12	7	12	29	5	13.17 (8.47)
BF	12	14	22	15	20	25	18.00 (5.10)
WC	24	30	26	17	15	26	23.00 (5.80)
SBP	4	4	5	5	2	7	4.50 (1.64)
DBP	25	27	14	29	33	20	24.67 (6.77)
FPG	15	6	11	14	13	14	12.17 (3.31)
L-IOP	20	19	20	23	16	13	18.50 (6.35)
R-IOP	9	2	3	4	8	6	5.33 (6.60)
r-GT	18	37	9	20	22	12	19.67 (9.81)
BUN	1	1	1	1	1	1	1 (0)
UA	22	23	27	16	25	19	22.00 (4.00)
TG	10	33	19	9	7	9	14.50 (9.99)
T-Cho	13	8	23	10	6	15	12.50 (6.09)
HDL-C	7	17	12	8	18	2	10.67 (6.19)
LDL-C	17	16	6	13	28	22	17.00 (7.54)
...	...	...	...	...	...	...	...

**Figure 4.** The ranked top ten most important variables.

#### 4. Discussion

Most of the previous academic literature has confirmed the related risk factors of CKD, including sex, age, race, obesity, smoking, unhealthy diets, family-related history, proteinuria, and anemia. It is associated with chronic diseases including metabolic syndrome, type 2 diabetes, hypertension, cardiovascular disease, hyperlipidemia, hyperuricemia, etc., and related indicators were also found to correlate with CKD. Based on the results of previous studies, many important risk factors of CKD, such as BUN, creatinine (Cr.), UA, SBP, DBP, WC, BMI, BF, FPG, T-Cho, and LDL were known [16,17,56–60], in addition to indicators such as SGPT, SGOT [16], Education [16,56], RBCs, UP (Urine Protein) [17].

In recent years, three related studies have been published that use different analytical tools and subjects to determine the risk factors for CKD in Taiwan [16,17,59]. Chang et al. (2020) consulted the Elderly Health Examination Database and used 2006–2012 data from

297,603 elderly people aged 65 years and older in Taipei City, Taiwan. Employing the non-CKD criteria with the G1 and G2 stages ( $\text{e-GFR} > 60 \text{ mL/min/1.73 m}^2$ ), their results showed a 29.7% e-GFR reduction in the likelihood of CKD diagnosis. The study found smoking to be significantly associated with an elevated risk of reduced e-GFR, and found physical exercise and healthy lifestyle habits to be significantly associated with increased e-GFR. Additionally, it found CVD, hypertension, obesity, and diabetes-related indicators to be linked to an increased risk of developing CKD [59]. Another study published in China used the same criteria ( $\text{e-GFR} > 60 \text{ mL/min/1.73 m}^2$ ) to detect CKD among 15,229 subjects (mean age: 62.8 years) from the Dongfeng–Tongji examination dataset (2008–2013). It found that BMI and MetS are potential indicators of CKD risk among elderly people [60].

Shih et al. (2020) analyzed data from an adult health examination dataset, as well as data on elderly adults they collected from three physical examination centers and 32 clinics in Taiwan (2015–2019). However, this study features a notable limitation: the G2 stage was not rigorous when it was used to represent and indicate CKD subjects. It was selected out of 14,169 non-CKD subjects ( $63.37 \pm 11.56$  years) and 5101 CKD subjects ( $69.19 \pm 10.74$  years)—a total of 19,270 subjects—with effective records, but they determined CKD by using the G1 stage ( $\text{e-GFR} \geq 90 \text{ mL/min/1.73 m}^2$ ) to indicate non-CKD. The study found the UP-Cr. ratio, proteinuria (PRO), RBCs, FPG, TG, T-Cho, age, and gender to be important risk factors for early CKD prediction [17]. Interestingly, they identified RBCs, in addition to UP, as an important factor, though they did not elaborate on it. Previous research on UP features supports data on the correlation with RBCs; in fact, some studies show that it may be a risk factor for hypertension [48].

This study is the follow-up research to Chiu et al.'s (2021) study. The datasets were collected from four major health screening centers in the northern, central, and southern parts of Taiwan (2010–2015). A total of 65,394 subjects were included in the MJPD database for the analysis of 18 risk indicators, CKD was determined by using the criteria with the G2 stages ( $\text{e-GFR} > 60 \text{ mL/min/1.73 m}^2$ ). The MJPD datasets were of the sub-health population, including more young subjects, aged around 30 to 50 years old (y/o). The study results showed that BUN and UA were identified as the first and second most important indicators, and SBP, SGPT, SGOT, and LDL-C were also related risk factors. Interestingly, socioeconomic status (SES)-related education was found to be the third important indicator in this study [16].

From the perspective of preventive medicine, the knowledge of risk factors facilitates early detection and, in turn, allows for targeting and improving relevant lifestyle habits, enabling people to avoid serious chronic diseases. In this study, we continued to use MJPD datasets [16], though notably with a younger sample. However, unlike the three most prominent previous CKD-related studies in Taiwan [16,57,59], we raised the criteria for CKD, asserting that CKD is stage 3b in the earliest stage of end-stage renal disease ( $\text{e-GFR} > 45 \text{ mL/min/1.73 m}^2$ ). At the same time, we increased the number of data-covered years (2005–2017) to increase the sample size. The MJPD dataset excluded the subjects' records related to anything but MetS, CKD stage 3a, and CKD stage 3b. Out of a total of 423 subjects, 88.65% were diagnosed with stage 3a CKD, and 11.35% were diagnosed with stage 3b CKD. BUN, SBP, R-IOP, RBCs, and C/H were identified as the five most important variables for evaluating subjects with MetS, CKD stage 3a, and CKD stage 3b.

### Limitations

In order to add the variables found in related studies and the variables that the researcher is interested in, and because analyzing too many research variables may affect the Area Under the Curve (AUC) of the algorithm, it is recommended that follow-up studies appropriately reduce variable analysis, or integrate more relevant variables, such as L-IOP and R-IOP, or T-Cho, HDL-C, LDL-C, and C/H related indicators. In addition, for a smaller number of samples, follow-up research may be able to further advance the analysis of the two risk factor values of relative importance value (RIV) or ordinal ranking value (ORV).

## 5. Conclusions

This study proposed innovative algorithms for the analysis of health-screening data pertaining to the third stage of CKD: the earliest stage of ESKD. This study contributed 33 relevant research variables, including R-IOP, RBCs, and T-CHO/HDL-C, outlining their varied associations with risk indicators identified in previous studies. This study suggested that some factorial combinations could potentially be used to separate individuals with stage 3a CKD from those with stage 3b CKD, facilitating the design of prospective studies in the future. We believe that this study has made several valuable contributions to the literature, including some that will aid in the prevention and treatment of CKD and the evaluation of high-risk groups in the third stage.

**Author Contributions:** Conceptualization, M.-S.C. and C.-J.L.; methodology, M.-J.J. and C.-J.L.; software, M.-J.J. and C.-J.L.; validation, T.-S.L., Y.-L.C. and C.-T.Y.; writing—original draft preparation, M.-J.J., M.-S.C. and C.-J.L.; writing—review and editing, M.-J.J., M.-S.C., T.-S.L., C.-T.Y., Y.-L.C. and C.-J.L.; supervision, T.-S.L. and C.-T.Y.; project administration, C.-J.L.; funding acquisition, M.-S.C. and C.-J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Science and Technology Council (NSTC 110-2221-E-030-010; NSTC-110-2221-E-161-003) and Fu Jen Catholic University (A0111181). This study did not receive a specific grant from any funding agency in the public, commercial, or not-for-profit sector.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Research Ethics Review Committee of Far Eastern Memorial Hospital (No: IRB-110027-E/Approved date; 15 February 2022), and ClinicalTrials.gov (ID: NCT05225454).

**Informed Consent Statement:** Informed consent was not required.

**Data Availability Statement:** Authorization is required for the use of all data sets collected from the MJ Health Research Foundation. The application procedures are accessed via this link. <http://www.mjhrf.org/main/page/release1/en/#release01> (accessed on 1 August 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Health Promotion Administration, Ministry of Health and Welfare. Metabolic Syndrome Criteria. 2007. Available online: <https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=639&pid=1219> (accessed on 1 September 2022).
2. Khosravi, A.; Sadeghi, M.; Barghikar, M. Which Components of Metabolic Syndrome Have a Greater Effect on Mortality, CVA and Myocardial Infarction, Hyperglycemia, High Blood Pressure or Both? *Adv. Biomed. Res.* **2017**, *6*, 121. [CrossRef]
3. DeBoer, M.D.; Filipp, S.L.; Musani, S.K.; Sims, M.; Okusa, M.D.; Gurka, M.J. Metabolic Syndrome Severity and Risk of CKD and Worsened GFR: The Jackson Heart Study. *Kidney Blood Press. Res.* **2018**, *43*, 555–567. [CrossRef] [PubMed]
4. Adair, K.E.; Ylitalo, K.R.; Forsse, J.S.; Funderburk, L.K.; Bowden, R.G. Metabolic Constellations, Clusters, and Renal Function: Findings from the 2013–2018 National Health and Nutrition Examination Surveys. *Life* **2021**, *11*, 904. [CrossRef] [PubMed]
5. Levey, A.S.; Eckardt, K.U.; Tsukamoto, Y.; Levin, A.; Coresh, J.; Rossert, J.; de Zeeuw, D.; Hostetter, T.H.; Lameire, N.; Eknoyan, G. Definition and classification of chronic kidney disease: A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* **2005**, *67*, 2089–2100. [CrossRef] [PubMed]
6. Levey, A.S.; Atkins, R.; Coresh, J.; Cohen, E.P.; Collins, A.J.; Eckardt, K.U.; Nahas, M.E.; Jaber, B.L.; Jadoul, M.; Levin, A.; et al. Chronic kidney disease as a global public health problem: Approaches and initiatives—A position statement from kidney disease improving global outcomes. *Kidney Int.* **2007**, *72*, 247–259. [CrossRef]
7. Baek, S.D.; Baek, C.H.; Kim, J.S.; Kim, S.M.; Kim, J.H.; Kim, S.B. Does stage III chronic kidney disease always progress to endstage renal disease? A ten-year follow-up study. *Scand. J. Urol. Nephrol.* **2012**, *46*, 232–238. [CrossRef]
8. Lee, J.; Oh, K.-H.; Park, S.-K. Dietary Micronutrients and Risk of Chronic Kidney Disease: A Cohort Study with 12 Year Follow-Up. *Nutrients* **2021**, *13*, 1517. [CrossRef]
9. Prasad, G.V. Metabolic syndrome and chronic kidney disease: Current status and future directions. *World J. Nephrol.* **2014**, *3*, 210–219. [CrossRef]
10. Choe, W.S.; Choi, E.K.; Han, K.D.; Lee, E.J.; Lee, S.R.; Cha, M.J.; Oh, S. Association of metabolic syndrome and chronic kidney disease with atrial fibrillation: A nationwide population-based study in Korea. *Diabetes Res. Clin. Pract.* **2019**, *148*, 14–22. [CrossRef]
11. Tozawa, M.; Iseki, C.; Tokashiki, K.; Chinen, S.; Kohagura, K.; Kinjo, K.; Takishita, S.; Iseki, K. Metabolic syndrome and risk of developing chronic kidney disease in Japanese adults. *Hypertens. Res.* **2007**, *30*, 937–943. [CrossRef]

12. Thomas, G.; Sehgal, A.R.; Kashyap, S.R.; Srinivas, T.R.; Kirwan, J.P.; Navaneethan, S.D. Metabolic syndrome and kidney disease: A systematic review and meta-analysis. *Clin. J. Am. Soc. Nephrol.* **2011**, *6*, 2364–2373. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Bonhauser, M.; Fernandez, G.; Püschel, K.; Yañez, F.; Montero, J.; Thompson, B.; Coronado, G. Improving physical fitness and emotional well-being in adolescents of low socioeconomic status in Chile: Results of a school-based controlled trial. *Health Promot. Int.* **2005**, *20*, 113–122. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Chertow, G.M.; Hsu, C.Y.; Johansen, K.L. The enlarging body of evidence: Obesity and chronic kidney disease. *J. Am. Soc. Nephrol.* **2006**, *17*, 1501–1502. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Tanner, R.M.; Brown, T.M.; Muntner, P. Epidemiology of obesity, the metabolic syndrome, and chronic kidney disease. *Curr. Hypertens. Rep.* **2012**, *14*, 152–159. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Chiu, Y.-L.; Jhou, M.-J.; Lee, T.-S.; Lu, C.-J.; Chen, M.-S. Health Data-Driven Machine Learning Algorithms Applied to Risk Indicators Assessment for Chronic Kidney Disease. *Risk Manag. Healthc. Policy* **2021**, *14*, 4401–4412. [\[CrossRef\]](#)
17. Shih, C.-C.; Lu, C.-J.; Chen, G.-D.; Chang, C.-C. Risk Prediction for Early Chronic Kidney Disease: Results from an Adult Health Examination Program of 19,270 Individuals. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4973. [\[CrossRef\]](#)
18. Chowdhury, N.H.; Reaz, M.B.I.; Haque, F.; Ahmad, S.; Ali, S.H.M.; A Bakar, A.A.; Bhuiyan, M.A.S. Performance Analysis of Conventional Machine Learning Algorithms for Identification of Chronic Kidney Disease in Type 1 Diabetes Mellitus Patients. *Diagnostics* **2021**, *11*, 2267. [\[CrossRef\]](#)
19. Wang, W.; Chakraborty, G.; Chakraborty, B. Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm. *Appl. Sci.* **2021**, *11*, 202. [\[CrossRef\]](#)
20. Krishnamurthy, S.; KS, K.; Dovgan, E.; Luštrek, M.; Gradišek Piletič, B.; Srinivasan, K.; Li, Y.-C.; Gradišek, A.; Syed-Abdul, S. Machine Learning Prediction Models for Chronic Kidney Disease Using National Health Insurance Claim Data in Taiwan. *Healthcare* **2021**, *9*, 546. [\[CrossRef\]](#)
21. Liu, Y.; Chen, P.-H.C.; Krause, J.; Peng, L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA* **2019**, *322*, 1806–1816. [\[CrossRef\]](#)
22. Triantafyllidis, A.K.; Tsanas, A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. *J. Med. Internet Res.* **2019**, *21*, e12286. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Peiffer-Smadja, N.; Rawson, T.M.; Ahmad, R.; Buchard, A.; Georgiou, P.; Lescure, F.X.; Birgand, G.; Holmes, A.H. Machine learning for clinical decision support in infectious diseases: A narrative review of current applications. *Clin. Microbiol. Infect.* **2020**, *26*, 584–595. [\[CrossRef\]](#)
24. Huang, L.-Y.; Chen, F.-Y.; Jhou, M.-J.; Kuo, C.-H.; Wu, C.-Z.; Lu, C.-H.; Chen, Y.-L.; Pei, D.; Cheng, Y.-F.; Lu, C.-J. Comparing Multiple Linear Regression and Machine Learning in Predicting Diabetic Urine Albumin–Creatinine Ratio in a 4-Year Follow-Up Study. *J. Clin. Med.* **2022**, *11*, 3661. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Chang, C.-C.; Yeh, J.-H.; Chiu, H.-C.; Chen, Y.-M.; Jhou, M.-J.; Liu, T.-C.; Lu, C.-J. Utilization of Decision Tree Algorithms for Supporting the Prediction of Intensive Care Unit Admission of Myasthenia Gravis: A Machine Learning-Based Approach. *J. Pers. Med.* **2022**, *12*, 32. [\[CrossRef\]](#)
26. Pes, B. Ensemble feature selection for high-dimensional data: A stability analysis across multiple domains. *Neural Comput. Appl.* **2020**, *32*, 5951–5973. [\[CrossRef\]](#)
27. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2012**, *42*, 463–484. [\[CrossRef\]](#)
28. Amasyali, M.F.; Ersoy, O. A comparative review of regression ensembles on drug design datasets. *Turk. J. Electr. Eng. Comput. Sci.* **2013**, *21*, 586–602. [\[CrossRef\]](#)
29. Alghamdi, M.; Al-Mallah, M.; Keteyian, S.; Brawner, C.; Ehrman, J.; Sakr, S. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. *PLoS ONE* **2017**, *12*, e0179805. [\[CrossRef\]](#)
30. Tuli, S.; Basumatary, N.; Gill, S.S.; Kahani, M.; Arya, R.C.; Wander, G.S.; Buyya, R. HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. *Futur. Gener. Comput. Syst.* **2020**, *104*, 187–200. [\[CrossRef\]](#)
31. Weber, C.; Röschke, L.; Modersohn, L.; Lohr, C.; Kolditz, T.; Hahn, U.; Ammon, D.; Betz, B.; Kiehnopf, M. Optimized Identification of Advanced Chronic Kidney Disease and Absence of Kidney Disease by Combining Different Electronic Health Data Resources and by Applying Machine Learning Strategies. *J. Clin. Med.* **2020**, *9*, 2955. [\[CrossRef\]](#)
32. Glazyrin, Y.E.; Veprintsev, D.V.; Ler, I.A.; Rossovskaya, M.L.; Varygina, S.A.; Glizer, S.L.; Zamay, T.N.; Petrova, M.M.; Minic, Z.; Berezovski, M.V.; et al. Proteomics-Based Machine Learning Approach as an Alternative to Conventional Biomarkers for Differential Diagnosis of Chronic Kidney Diseases. *Int. J. Mol. Sci.* **2020**, *21*, 4802. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Granal, M.; Slimani, L.; Florens, N.; Sens, F.; Pelletier, C.; Pszczolinski, R.; Casiez, C.; Kalbacher, E.; Jolivot, A.; Dubourg, L.; et al. Prediction Tool to Estimate Potassium Diet in Chronic Kidney Disease Patients Developed Using a Machine Learning Tool: The UniverSel Study. *Nutrients* **2022**, *14*, 2419. [\[CrossRef\]](#)
34. Lee, C.-L.; Liu, W.-J.; Tsai, S.-F. Development and Validation of an Insulin Resistance Model for a Population with Chronic Kidney Disease Using a Machine Learning Approach. *Nutrients* **2022**, *14*, 2832. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Abutaleb, N. Why we should sub-divide CKD stage 3 into early (3a) and late (3b) components. *Nephrol. Dial. Transplant.* **2007**, *22*, 2728–2729. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Zahran, A.; Shoker, A. About CKD stage-3 subdivision proposal. *Nephrol. Dial. Trans.* **2008**, *23*, 1765. [\[CrossRef\]](#)



37. Segal, Z.; Kalifa, D.; Radinsky, K.; Ehrenberg, B.; Elad, G.; Maor, G.; Lewis, M.; Tibi, M.; Korn, L.; Koren, G. Machine learning algorithm for early detection of end-stage renal disease. *BMC Nephrol.* **2020**, *21*, 518. [CrossRef]
38. Oh, T.R.; Song, S.H.; Choi, H.S.; Suh, S.H.; Kim, C.S.; Jung, J.Y.; Choi, K.H.; Oh, K.-H.; Ma, S.K.; Bae, E.H.; et al. Predictive Model for High Coronary Artery Calcium Score in Young Patients with Non-Dialysis Chronic Kidney Disease. *J. Pers. Med.* **2021**, *11*, 1372. [CrossRef] [PubMed]
39. Silveira, A.C.M.d.; Sobrinho, Á.; Silva, L.D.d.; Costa, E.d.B.; Pinheiro, M.E.; Perkusich, A. Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets. *Appl. Sci.* **2022**, *12*, 3673. [CrossRef]
40. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Peng, C.-Y.J.; Lee, K.L.; Ingersoll, G.M. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [CrossRef]
42. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
43. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
44. Ke, G.; Meng, Q.; Finley, T.W.; Wang, T.; Chen, W.; Ma, W.; Qiwei, Y.; Liu, T. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3147–3155.
45. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363. [CrossRef]
46. Meng, J.; Hao, H.; Luan, Y. Classifier ensemble selection based on affinity propagation clustering. *J. Biomed. Inform.* **2016**, *60*, 234–242. [CrossRef]
47. Woloshuk, A.; Khochare, S.; Almulhim, A.F.; McNutt, A.T.; Dean, D.; Barwinska, D.; Ferkowicz, M.J.; Eadon, M.T.; Kelly, K.J.; Dunn, K.W.; et al. In Situ Classification of Cell Types in Human Kidney Tissue Using 3D Nuclear Staining. *Cytom. Part A* **2021**, *99*, 707–721. [CrossRef]
48. Liao, P.-C.; Chen, M.-S.; Jhou, M.-J.; Chen, T.-C.; Yang, C.-T.; Lu, C.-J. Integrating Health Data-Driven Machine Learning Algorithms to Evaluate Risk Factors of Early Stage Hypertension at Different Levels of HDL and LDL Cholesterol. *Diagnostics* **2022**, *12*, 1965. [CrossRef] [PubMed]
49. Kuhn, M. Caret: Classification and Regression Training. R Package Version, 6.0-93. 2022. Available online: <https://CRAN.R-project.org/package=caret> (accessed on 1 September 2022).
50. Breiman, L.; Cutler, A.; Liaw, A.; Wiener, M. randomForest: Breiman and Cutler’s Random Forests for Classification and Regression. R Package Version, 4.7-1.1. 2022. Available online: <https://CRAN.R-project.org/package=randomForest> (accessed on 1 September 2022).
51. Milborrow, S. Derived from Mda: MARS by T. Hastie and R. Tibshirani. Earth: Multivariate Adaptive Regression Splines. R Package Version, 5.3.1. 2021. Available online: <http://CRAN.R-project.org/package=earth> (accessed on 1 September 2022).
52. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. Xgboost: Extreme Gradient Boosting. R Package Version, 1.6.0.1. 2022. Available online: <https://CRAN.R-project.org/package=xgboost> (accessed on 1 September 2022).
53. Yandex Technologies. CatBoost: Unbiased Boosting with Categorical Features. R Package Version, 1.0.6. 2022. Available online: <https://github.com/catboost/catboost/> (accessed on 1 September 2022).
54. Microsoft. LightGBM: Light Gradient Boosting Machine. R Package Version, 3.3.2. 2022. Available online: <https://github.com/microsoft/LightGBM> (accessed on 1 September 2022).
55. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **1988**, *44*, 837–845. [CrossRef] [PubMed]
56. Boulware, L.E.; Jaar, B.G.; Tarver-Carr, M.E.; Brancati, F.L.; Powe, N.R. Screening for Proteinuria in US Adults: A cost-effectiveness analysis. *JAMA* **2003**, *290*, 3101–3114. [CrossRef]
57. Wen, C.P.; Cheng, T.Y.; Tsai, M.K.; Chang, Y.C.; Chan, H.T.; Tsai, S.P.; Chiang, P.H.; Hsu, C.C.; Sung, P.K.; Hsu, Y.H.; et al. All-cause mortality attributable to chronic kidney disease: A prospective cohort study based on 462,293 adults in Taiwan. *Lancet* **2008**, *371*, 2173–2182. [CrossRef]
58. Kazancıoğlu, R. Risk factors for chronic kidney disease: An update. *Kidney Int. Suppl.* **2013**, *3*, 368–371. [CrossRef]
59. Chang, H.J.; Lin, K.R.; Lin, M.T.; Chang, J.L. Association between lifestyle factors and decreased kidney function in older adults: A community-based cross-sectional analysis of the Taipei City elderly health examination database. *BMC Nephrol.* **2020**, *21*, 169. [CrossRef]
60. Wang, Y.; Sun, B.; Sheng, L.-T.; Pan, X.-F.; Zhou, Y.; Zhu, J.; Li, X.; Yang, K.; Guo, K.; Zhang, X.; et al. Association between weight status, metabolic syndrome, and chronic kidney disease among middle-aged and elderly Chinese. *Nutr. Metab. Cardiovasc. Dis. NMCD* **2020**, *30*, 2017–2026. [CrossRef] [PubMed]