Machine learning for morbid glomerular hypertrophy

Hiroshi Kataoka (kataoka@twmu.ac.jp)
Tokyo Women's Medical University
Yusuke Ushio
Tokyo Women's Medical University
Kazuhiro Iwadoh
Tokyo Women's Medical University
Mamiko Ohara
Kameda Medical Center
Tomo Suzuki
Kameda Medical Center
Maiko Hirata
Saitama Red Cross Hospital
Shun Manabe
Tokyo Women's Medical University
Keiko Kawachi
Tokyo Women's Medical University
Taro Akihisa
Tokyo Women's Medical University
Shiho Makabe
Tokyo Women's Medical University
Masayo Sato
Tokyo Women's Medical University
Naomi Iwasa
Tokyo Women's Medical University
Rie Yoshida
Tokyo Women's Medical University
Junichi Hoshino
Tokyo Women's Medical University
Toshio Mochizuki
Tokyo Women's Medical University
Ken Tsuchiya
Tokyo Women's Medical University
Kosaku Nitta

Tokyo Women's Medical University

Research Article

Keywords: machine learning, glomerular hypertrophy, feature selection, symbolic regression, cross-validation

Posted Date: June 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1546191/v1

Abstract

Background:

A practical research method integrating machine learning with conventional statistics has been sought in medicine. Although glomerular hypertrophy (or a large renal corpuscle) on renal biopsy has pathophysiological implications, it is often misdiagnosed as adaptive/compensatory hypertrophy. We explored the factors associated with a maximal glomerular diameter of ≥242.3 μm using machine learning.

Methods:

Using the frequency-of-usage feature ranking in predictive models, we defined the machine learning scores calculated from symbolic regression models. We compared features selected by genetic programming with those selected by a point-biserial correlation coefficient using multivariable logistic and linear regression to validate discriminatory ability, goodness-of-fit, and collinearity.

Results:

Body mass index, complement component C3, serum total protein, arteriolosclerosis, C-reactive protein, and the Oxford E1 score were ranked among the top 10 features with high machine learning scores using genetic programming. The estimated glomerular filtration rate was ranked 46th among the 60 features. In multivariable analyses, the R^2 value was higher (0.61 vs. 0.45), and the corrected Akaike Information Criterion value was lower (402.7 vs. 417.2) in the model with features generated by genetic programming than in the model generated with features using point-biserial *r*. There were two features with variance inflation factors higher than 5 in the model using point-biserial *r* and none in the machine learning model.

Conclusions:

Machine learning may be useful in identifying significant and insignificant correlated factors. Our method may be generalized to other medical research due to the procedural simplicity of using top-ranked features selected by machine learning.

Introduction

Machine learning (ML) refers to computational techniques applied in many fields. ML has been proposed as a valuable and necessary tool for modern health care systems [1–4]. Traditional statistics focus on inferring linear relationships among variables, whereas ML extracts both linear and non-linear relationships between explanatory variables (features) and a target variable (feature), emphasizing the building of accurate predictive models [2–4]. Although clinicians are interested in applying ML techniques to healthcare fields [1,5], the data-driven approach of ML seems at odds with the traditional model-driven

statistical approach in medical research [3,4]. Therefore, a practical method that integrates ML with conventional statistics is needed.

However, although ML models are expressed in an implicit form, some ML models are transparent enough to identify novel related factors. ML methods, such as symbolic regression (SR) techniques, have transparency in the created models, allowing easy understanding of the prediction mechanisms. Furthermore, ML methods can be applied to wide-ranged data (high-dimensional context) because there are no restrictions regarding the number of input variables and their parameters [2,3,6]. Since biological data are often characterized by multitudes of variables and few available samples, clinicians want to ascertain the dominant features used in the predictive model rather than the predictive model function itself.

Glomerular hypertrophy (GH) (or a large renal corpuscle) reflects the renal injury state [7]. We first reported that GH, defined by maximal glomerular diameter (MaxGD), implies poor renal prognosis in IgA nephropathy [8]. We confirmed MaxGD as an effectual pathological factor predicting renal IgA nephropathy prognosis in another cohort [9] and a follow-up study [10]. However, in clinical settings, GH is often underestimated as compensatory GH. A few studies have investigated the pathophysiological factors associated with GH. There is also insufficient research demonstrating that MaxGD, which is quantified histologically by needle biopsy, is not an indicator of compensatory GH but of renal damage itself. Due to difficulties enrolling an adequate number of cases in renal biopsy studies, a technique is needed to examine small cohorts. Although ML is excellent for analyzing “big data,” techniques for analyzing small-size data have also been developed [3,11]. Herein, we integrated traditional statistical methods with ML by first using an ML technique for feature selection from many candidate risk factors and subsequently using traditional statistical methods for their validation.

Materials And Methods

Study design

The primary outcome of this study was GH, defined as a MaxGD of ≥ 242.3 μm. This study was approved by the Ethics Committee of the Kameda Medical Center (No. 17–170) and followed the principles of the 1964 Helsinki Declaration. This retrospective study used the same cohort as our previous study (43 patients with IgA nephropathy) [8]. We used passive informed consent (opt-out) for patients. Data were analyzed anonymously.

Variables examined for GH presence

The 60 surveyed predictive variables for GH (56 baseline variables, 4 follow-up variables) included patient baseline data, laboratory test data, comorbidities, pathological findings, and medical history. The Supplemental Methods describe the comorbidity definitions and histological kidney biopsy assessment. Table 1 shows the survey items.

Table 1
Patient characteristics according to the presence of MaxGD ≥ 242.3 μm: Clinical, laboratory, and pathological findings

Variables	Entire cohort	Cohort with MaxGD ≥ 242.3 μm	Cohort without MaxGD ≥ 242.3 μm	*P*-value
	n = 43	n = 15	n = 28	
Clinical Findings				
Height (m)	1.63 (1.47–1.79)	1.64 (1.47–1.79)	1.63 (1.50–1.78)	0.9797
BW (kg)	67.8 (41.6–114.0)	72.9 (54.5–114.0)	62.2 (41.6–90.1)	0.0226
BMI (kg/m^2)	25.1 (16.8–37.5)	27.8 (21.7–37.5)	23.8 (16.8–32.3)	0.0005
Sex (Men; %)	39.5/ 60.5	26.7/ 73.3	46.4/ 53.6	0.3274
Age (years)	41 (19–59)	41 (26–59)	41 (19–53)	0.6742
Interval from onset (months)	42 (1–325)	37 (1–287)	42 (2–325)	0.6193
SBP (mmHg)	143.1 ± 22.2	145.3 ± 21.2	142.0 ± 23.0	0.6464
DBP (mmHg)	82.3 ± 15.4	84.5 ± 14.8	81.2 ± 15.9	0.3266
MBP (mmHg)	102.6 ± 16.8	104.8 ± 15.9	101.4 ± 17.5	0.4677
PP (mmHg)	60.8 ± 13.4	60.7 ± 14.1	60.8 ± 13.3	0.8285
Edema (%)	86.1/ 14.0	100.0/ 0.0	78.6/ 21.4	0.0764
Acute kidney injury (%)	88.4/ 11.6	86.7/ 13.3	89.3/ 10.7	1.0000

Continuous variables are expressed as means ± standard deviation or median (minimum–maximum). Count data are expressed as %. For the variables with missing data, the number of patients with non-missing data are shown in []. Abbreviations: MaxGD, maximal glomerular diameter; n, number; BW, body weight, BMI, body mass index; %, percentages; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; PP, pulse pressure; WBC, white blood cells; eGFR, estimated glomerular filtration rate; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; C3, complement component 3; C4, complement component 4; U-Prot, Urinary protein excretion; U-RBC, urinary red blood cells; HPF, high-power field; IBW, ideal body weight; GBM, glomerular basement membrane; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; C, cellular/fibrocellular crescents.

Variables	Entire cohort	Cohort with MaxGD ≧ 242.3 μm	Cohort without MaxGD ≧ 242.3 μm	*P*-value
	n = 43	n = 15	n = 28	
Family history (%)	86.1/ 14.0	86.7/ 13.3	85.7/ 14.3	1.0000
Laboratory Findings				
WBC (10^3/μL)	6.78 ± 1.28	7.05 ± 1.23	6.64 ± 1.31	0.3725
Hemoglobin (g/dL)	14.2 ± 2.0	14.6 ± 2.5	14.0 ± 1.7	0.2788
Platelet (10^3/μL)	21.6 ± 6.1	22.3 ± 7.9	21.2 ± 5.0	0.9898
Total protein (g/dL)	6.48 ± 0.59	6.71 ± 0.51	6.36 ± 0.59	0.1084
Serum albumin (g/dL)	3.79 ± 0.39	3.91 ± 0.40	3.73 ± 0.38	0.1360
eGFR (mL/min/1.73m^2)	78.6 ± 17.5	72.2 ± 12.9	81.9 ± 18.9	0.1141
C-reactive protein (mg/dL)	0.08 (0.01–0.92)	0.13 (0.01–0.77)	0.06 (0.01–0.92)	0.0313
IgG (mg/dL)	1141.8 ± 321.0	1173.8 ± 401.4	1124.6 ± 275.4	0.9797
IgA (mg/dL)	337.1 ± 139.7	362.1 ± 162.2	323.7 ± 127.2	0.4755
IgM (mg/dL)	166.0 ± 92.0	160.6 ± 84.3	168.9 ± 97.3	0.7832
C3 (mg/dL)	90.4 ± 19.2	103.6 ± 21.6	83.9 ± 14.2	0.0059
C4 (mg/dL)	36.9 ± 10.9	38.7 ± 11.3	36.0 ± 10.8	0.4599
U-Prot (g/day)	1.4 (0.0–7.0)	1.5 (0.6–5.7)	1.1 (0.0–7.0)	0.5493

Continuous variables are expressed as means ± standard deviation or median (minimum–maximum). Count data are expressed as %. For the variables with missing data, the number of patients with non-missing data are shown in []. Abbreviations: MaxGD, maximal glomerular diameter; n, number; BW, body weight, BMI, body mass index; %, percentages; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; PP, pulse pressure; WBC, white blood cells; eGFR, estimated glomerular filtration rate; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; C3, complement component 3; C4, complement component 4; U-Prot, Urinary protein excretion; U-RBC, urinary red blood cells; HPF, high-power field; IBW, ideal body weight; GBM, glomerular basement membrane; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; C, cellular/fibrocellular crescents.

Variables	Entire cohort	Cohort with MaxGD ≥ 242.3 μm	Cohort without MaxGD ≥ 242.3 μm	*P*-value
	n = 43	n = 15	n = 28	
U-RBC (counts/HPF)	10 (0–200)	1 (0–100)	10 (0–200)	0.0926
U-WBC (counts/HPF)	1.8 (0–100)	1 (0–20)	1 (0–100)	0.9898
Salt intake (g/day)	8.57 ± 3.06	9.04 ± 3.76	8.31 ± 2.64	0.2961
Protein intake (g/kg□IBW/day)	1.05 ± 0.26	1.13 ± 0.31	1.01 ± 0.22	0.0977
Follow-up Findings				
U-Prot during 10-year follow-up (grades 0–4)	1.1 (0.2–3.4)	1.7 (0.3–3.4)	1.0 (0.2–2.7)	0.0059
Dipstick hematuria during 10-year follow-up (grades 0–4)	1.4 (0.0–3.0)	1.4 (0.1–3.0)	1.4 (0.0–3.0)	0.9289
Increase in U-Prot during 10-year follow-up (%)	62.8/ 37.2	40.0/ 60.0	75.0/ 25.0	0.0236
Increase in dipstick hematuria during 10-year follow-up (%)	60.5/ 39.5	33.3/ 66.7	75.0/ 25.0	0.0077
Comorbidities				
Hypertension (%)	23.3/ 76.7	20.0/ 80.0	25.0/ 75.0	1.0000
Hyperuricemia (%)	72.1/ 27.9	53.3/ 46.7	82.1/ 17.9	0.0447
Hypertriglyceridemia (%)	62.8/ 37.2	46.7/ 53.3	71.4/ 28.6	0.1094
Hypercholesterolemia (%)	76.7/ 23.3	80.0/ 20.0	75.0/ 25.0	1.0000
Nephrotic Syndrome (%)	86.1/ 14.0	100.0/ 0.0	78.6/ 21.4	0.0764
Histological findings				
Number of glomeruli	13.4 ± 5.5	12.1 ± 4.5	14.1 ± 5.9	0.1810

Continuous variables are expressed as means ± standard deviation or median (minimum–maximum). Count data are expressed as %. For the variables with missing data, the number of patients with non-missing data are shown in []. Abbreviations: MaxGD, maximal glomerular diameter; n, number; BW, body weight, BMI, body mass index; %, percentages; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; PP, pulse pressure; WBC, white blood cells; eGFR, estimated glomerular filtration rate; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; C3, complement component 3; C4, complement component 4; U-Prot, Urinary protein excretion; U-RBC, urinary red blood cells; HPF, high-power field; IBW, ideal body weight; GBM, glomerular basement membrane; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; C, cellular/fibrocellular crescents.

Variables	Entire cohort	Cohort with MaxGD ≥ 242.3 μm	Cohort without MaxGD ≥ 242.3 μm	*P*-value
	n = 43	n = 15	n = 28	
Global sclerosis (%)	14.3 (0.0–57.1)	12.5 (0.0–55.6)	14.5 (0.0–57.1)	0.9695
Segmental sclerosis (%)	14.7 (0.0–69.2)	11.1 (0.0–55.6)	17.7 (0.0–69.2)	0.2515
Perihilar hyalinosis (%) [42]	57.1/ 42.9	33.3/ 66.7	70.4/ 29.6	0.0269
Crescent (%)	6.3 (0.0–83.3)	6.3 (0.0–33.3)	5.9 (0.0–83.3)	0.6465
Cellular crescent (%)	5.9 (0.0–83.3)	0.0 (0.0–22.2)	5.9 (0.0–83.3)	0.4148
Fibrous crescent (%)	0.0 (0.0–12.5)	0.0 (0.0–12.5)	0.0 (0.0–8.3)	0.3101
Focal adhesions to Bowman's capsule (%) [42]	90.5/ 9.5	80.0/ 20.0	96.3/ 3.7	0.1220
Thickness of GBM (%) [42]	90.5/ 9.5	86.7/ 13.3	92.6/ 7.4	0.6080
Mesangial cell proliferation (grades 1–3) 1/ 2/ 3 (%)	20.9/ 32.6/ 46.5	13.3/ 26.7/ 60.0	25.0/ 35.7/ 39.3	0.5353
Number of mesangial cells per mesangial lesion	7 (4–10)	8 (4–10)	6.5 (4–9)	0.3061
Interstitial fibrosis (%)	5.0 (0.0–60.0)	5.0 (0.0–60.0)	5.0 (0.0–30.0)	0.7766
Interstitial inflammation (%)	5.0 (0.0–30.0)	5.0 (0.0–25.0)	5.0 (0.0–30.0)	0.5248
Intimal thickening of interlobular artery (grades 0–3) 0/ 1/ 2/ 3 (%)	48.8/ 34.9/ 16.3/ 0.0	20.0/ 53.3/ 26.7/ 0.0	64.3/ 25.0/ 10.7/ 0.0	0.0201

Continuous variables are expressed as means ± standard deviation or median (minimum–maximum). Count data are expressed as %. For the variables with missing data, the number of patients with non-missing data are shown in []. Abbreviations: MaxGD, maximal glomerular diameter; n, number; BW, body weight, BMI, body mass index; %, percentages; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; PP, pulse pressure; WBC, white blood cells; eGFR, estimated glomerular filtration rate; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; C3, complement component 3; C4, complement component 4; U-Prot, Urinary protein excretion; U-RBC, urinary red blood cells; HPF, high-power field; IBW, ideal body weight; GBM, glomerular basement membrane; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; C, cellular/fibrocellular crescents.

Variables	Entire cohort	Cohort with MaxGD ≥ 242.3 μm	Cohort without MaxGD ≥ 242.3 μm	*P*-value
	n = 43	n = 15	n = 28	
Arteriolosclerosis (grades 0–3) 0/ 1/ 2/ 3 (%)	25.6/ 53.5/ 16.3/ 4.7	6.7/ 46.7/ 33.3/ 13.3	35.7/ 57.1/ 7.1/ 0.0	0.0073
Japanese histological grades (grades 1–4) 1/ 2/ 3/ 4 (%)	4.7/ 4.7/ 86.1/ 4.7	6.7/ 6.7/ 80.0/ 6.7	3.6/ 3.6/ 89.3/ 3.6	1.0000
Oxford M1 (%)	18.6/ 81.4	20.0/ 80.0	17.9/ 82.1	1.0000
Oxford E1 (%)	46.5/ 53.5	40.0/ 60.0	50.0/ 50.0	0.5309
Oxford S1 (%)	18.6/ 81.4	26.7/ 73.3	14.3/ 85.7	0.4188
Oxford T 0/ 1/ 2 (%)	95.4/ 2.3/ 2.3	93.3/ 0.0/ 6.7	96.4/ 3.6/ 0.0	0.5814
Oxford C 0/ 1/ 2 (%)	44.2/ 46.5/ 9.3	46.7/ 53.3/ 0.0	42.9/ 42.9/ 14.3	0.4128
Localization of glomerulus examined for MaxGD in renal cortex; inner/middle/outer (%)	23.3/ 55.8/ 20.9	20.0/ 60.0/ 20.0	25.0/ 53.6/ 21.4	1.0000
MaxGD (μm)	221.7 ± 30.8	253.1 ± 10.9	204.9 ± 24.0	< 0.0001

Continuous variables are expressed as means ± standard deviation or median (minimum–maximum). Count data are expressed as %. For the variables with missing data, the number of patients with non-missing data are shown in []. Abbreviations: MaxGD, maximal glomerular diameter; n, number; BW, body weight, BMI, body mass index; %, percentages; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; PP, pulse pressure; WBC, white blood cells; eGFR, estimated glomerular filtration rate; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; C3, complement component 3; C4, complement component 4; U-Prot, Urinary protein excretion; U-RBC, urinary red blood cells; HPF, high-power field; IBW, ideal body weight; GBM, glomerular basement membrane; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; C, cellular/fibrocellular crescents.

Statistical analyses

Continuous variables are reported as means and standard deviations or as medians (minimum–maximum). Categorical variables are reported as percentages unless stated otherwise. Group differences were evaluated using the unpaired t-test, Mann–Whitney U test, chi-square test, or Fisher's exact test, as appropriate. Based on a previous report [8], we classified patients into two groups as follows: patients with a MaxGD ≥ 242.3 μm and patients with a MaxGD < 242.3 μm. The correlations between MaxGD ≥ 242.3 μm and the other variables were assessed using the point-biserial correlation coefficient (point-biserial *r*). Multivariable linear regression analyses were also performed for MaxGD (μm).

ML

To create an ML-based predictive model (based on MaxGD ≥ 242.3 μm presence), a dichotomous response variable (1 or 0) and 60 explanatory features were used. First, as a preprocessing test for the dataset to control Type-I errors, and to verify meaningful relationships between explanatory features and the target feature [12], we performed a permutation test [13,14] using the Python scikit-learn library [15] (Supplemental Methods) for the entire dataset. We used DataModeler® version 9.3 (Evolved Analytics LLC, Midland, MI, USA) that runs on Mathematica® version 12.1 (Wolfram Research Incorporated, Champaign, IL, USA) to create a dichotomous discriminant function for predicting MaxGD ≥ 242.3 μm. DataModeler® performs SR via genetic programming (SR via GP), a type of ML, which was evaluated by the LOO-CV procedure based on the study sample size [11,13,14]. Detailed procedures for generating prognostic predictive functions and ML scores for GH presence are described in the Supplemental Methods.

Statistical assessment

Using the ML scores associated with MaxGD ≥ 242.3 μm derived from SR via GP, we compared features selected by point-biserial *r* and from ML scores. Multivariable logistic regression analyses and multivariable linear regression analyses related to MaxGD were used to validate discriminatory ability, goodness-of-fit, and collinearity. The model's goodness-of-fit was assessed using R-squared (R^2), McFadden's pseudo-R-squared (pseudo-R^2) [16], and the corrected Akaike Information Criterion (AIC) [17]. The model's discriminatory ability was evaluated using the concordance statistic (C-statistic) [18,19], and multicollinearity was assessed using the variance inflation factor (VIF) [20]. We used standard methods to estimate the sample sizes needed for the multivariate logistic and linear regression analyses; at least five positive outcomes were needed for each included independent variable [21–23]. Two-tailed statistical significance was set at $P < 0.05$. Analyses were performed using JMP Pro version 15.1.0 (SAS Institute, Cary, NC, USA), and the remaining analyses were performed using Mathematica® version 12.1 (Wolfram Research, Inc., Champaign, IL, USA).

Results

Patient characteristics and the permutation test for the dataset

Patients characteristics (26 men, 17 women) were analyzed (Table 1). Figure 1 illustrates a histogram of MaxGD. Fifteen patients (34.9%) had a MaxGD of ≥ 242.3 μm. Figure 2 shows the accuracy scores for the permuted data distribution. The accuracy score value with the dataset (MaxGD dataset) was 0.84, significantly higher than that obtained using permuted data, and the permutation P-value was 0.001. This indicated a low likelihood of this score being obtained by chance alone. It provides evidence that the

MaxGD dataset contains a dependency between explanatory variables and target variables, with reliable classification performance and sufficient information to analyze.

Correlation coefficients (point-biserial Rs) between predictive variables and MaxGD ≥ 242.3 µm

All correlation coefficients (denoted as point-biserial R) of prognostic predictive variables in the presence of MaxGD are shown (Supplemental Fig. 1). Further, the point-biserial R values of the top 20 variables with the strongest positive or negative correlations are listed (Supplemental Table 1).

Analysis of predictive features for MaxGD ≥ 242.3 µm using ML

We generated 19,437 functions using GP, and Fig. 3 shows their distribution in the function space. We selected 1,819 functions with lower complexities and lower $1 - R^2$ in each epoch to limit the number of models to 8–20% of all generated models. We ranked the appearance frequencies of all predictive features in the 1,819 limited functions (Supplemental Fig. 2). Among these, the top 15 features of appearance frequencies are shown (Fig. 4).

ML score

We created an ML score via GP [ML score (GP)] by dividing these frequencies by 100. Table 2 shows the top 15 features with high ML scores (GP) in association with MaxGD ≥ 242.3 µm. The ML score (GP) values, ML score (GP) ranks, $/R/$ values, and $|R|$ ranks for each feature are listed (Table 2). The top 8 features, with a higher ML score (GP) were body mass index (BMI), complement C3, serum total protein, arteriolosclerosis, urinary protein excretion during the 10-year follow-up, edema, C-reactive protein, and the Oxford E1 score. The estimated glomerular filtration rate (eGFR) (15th in the $/R/$ rank) was not in the top 15 rankings in the ML scores (GP) (46th rank), indicating that MaxGD is not affected by nephron disappearance but by another mechanism. Inflammatory indicators, such as C-reactive protein (16th in the $/R/$ rank), Oxford E1 (41st in the $/R/$ rank), and the number of mesangial cells per mesangial lesion (25th in the $/R/$ rank), ranked in the top 15 ML scores (GP), indicating the association of MaxGD with inflammation.

Table 2
Top 15 variables in machine learning score via genetic programming [ML score (GP)] in association with the MaxGD ≥ 242.3 μm

Variables	ML score (GP)	Rank of ML score (GP)	$\|R\|$	Rank of $\|R\|$
BMI (kg/m^2)	0.829	**1**	0.530	**1**
Complement C3 (mg/dL)	0.722	**2**	0.474	**3**
Total protein (g/dL)	0.673	**3**	0.293	**13**
Arteriolosclerosis (grades 0–3)	0.482	**4**	0.502	**2**
U-Prot during 10-year follow-up (grades 0–4)	0.237	**5**	0.460	**4**
Edema (vs. no)	0.148	**6**	0.295	**11**
C-reactive protein (mg/dL)	0.071	**7**	0.263	**16**
Oxford E1 (vs. no)	0.068	**8**	0.096	**41**
Body weight (kg)	0.034	**9**	0.411	**5**
Increase in dipstick hematuria during 10-year follow-up (vs. no)	0.032	**10**	0.406	**6**
Intimal thickening of the interlobular artery (grades 0–3)	0.032	**11**	0.389	**7**
Increase in U-Prot during 10-year follow-up (vs. no)	0.030	**12**	0.345	**9**
Perihilar hyalinosis (vs. no)	0.027	**13**	0.368	**8**
Number of mesangial cells per mesangial lesion	0.020	**14**	0.154	**25**
Localization of glomerulus examined for MaxGD (inner/middle/outer)	0.019	**15**	0.026	**52**

ML score (GP) is calculated by dividing the frequency of the predictive variables utilized in 1,819 predictive functions (Supplemental Fig. 2) by 100. / *R*/ is the absolute value of point-biserial *r* (Supplemental Fig. 1). Abbreviations: ML, machine learning; GP, genetic programming; MaxGD, maximal glomerular diameter; point-biserial *r*, point-biserial correlation coefficient; BMI, body mass index; C3, component 3; U-Prot, Urinary protein excretion; Oxford E1, the presence of endocapillary hypercellularity.

Validation using a conventional statistical technique

Multivariable logistic regression analyses and multivariable linear regression analyses related to MaxGD were performed to validate discriminatory ability, goodness-of-fit, and collinearity for comparing the features selected from the ML score (GP) with those selected from point-biserial *r*.

Comparison of features selected from the point-biserial r and ML scores

Table 3 shows the results of multivariable logistic regression analyses for MaxGD ≥ 242.3 μm using the top 3 features based on the ML score (GP) or |*R*|. Although the model based on the top 3 features of the ML score (GP) (Model ML top 3) detected two significant features, BMI and serum total protein, the model based on the top 3 features of the |*R*| (Model |*R*| top 3) detected no significant features. Furthermore, the values of pseudo-R^2 and C-statistic were higher (0.50 vs. 0.42 and 0.93 vs. 0.87), and the AIC value was lower (36.9 vs. 41.5) in the Model ML top 3 than in the Model |*R*| top 3.

Table 3
Multivariable logistic regression analyses for the MaxGD ≥ 242.3 μm using top 3 variables based on ML score (GP) or |*R*|

Model based on top 3 variables of ML score (GP) **pseudo-R^2 = 0.50, AICc = 36.9, C-statistic = 0.93**	**Odds Ratio (95% CI)**	***P*-value**
BMI (1 kg/m^2 increase)	1.57 (1.11–2.20)	0.0100*
Complement C3 (1 mg/dL increase)	1.05 (0.98–1.13)	0.1393
Total protein (1 g/dL increase)	12.94 (1.68–99.71)	0.0140*
Model based on top 3 variables of |R| **pseudo-R^2 = 0.42, AICc = 41.5, C-statistic = 0.87**	**Odds Ratio (95% CI)**	**P-value**
BMI (1 kg/m^2 increase)	1.25 (0.92–1.68)	0.1516
Arteriolosclerosis (grades 0–3)	3.51 (0.93–13.21)	0.0631
Complement C3 (1 mg/dL increase)	1.06 (1.00–1.13)	0.0681

*P < 0.05. Variables of top 3 variables based on ML score (GP) or |R| were included in the multivariate model. Abbreviations: MaxGD, maximal glomerular diameter; ML, machine learning; GP, genetic programming; *|R|*, the absolute value of point-biserial correlation coefficient; pseudo-R^2, McFadden's pseudo-*R*-squared; AICc, small-sample corrected Akaike Information Criterion; BMI, body mass index; C3, component 3.

Comparison of features selected from the point-biserial r and the ML scores

Table 4 shows the results of the multivariable linear regression analyses for MaxGD ≥ 242.3 μm using the top 8 features based on the ML score (GP) or |*R*|. Although the Model ML top 8 detected four significant features, i.e., BMI, serum total protein, C-reactive protein, and the Oxford E1 score, the Model |*R*| top 8 detected no significant features (Table 4). In the Model ML top 8, the R^2 value was higher (0.61 vs. 0.45), and the AIC value was lower (402.7 vs. 417.2) than that in the Model |*R*| top 8. Furthermore, in Model |*R*| top 8, there were two features with VIF > 5 and four features with VIF > 1.81 [calculated by 1/(1-overall model R^2)], [24] indicating multicollinearity in the model. Contrastingly, in the Model ML top 8, there were no features with VIF > 5 or VIF > 2.54 [calculated by 1/(1-overall model R^2], indicating reduced multicollinearity in the model.

Table 4
Multivariable linear regression analyses for the MaxGD using top 8 variables based on ML score (GP) or |*R*|

Model based on top 8 variables of ML score (GP) R^2 **= 0.61, AICc = 402.7**	***β***	**VIF**	***t*-ratio**	***P*-value**
BMI (1 kg/m^2 increase)	0.47	2.04	3.06	0.0043*
Complement C3 (1 mg/dL increase)	0.09	1.42	0.69	0.4956
Total protein (1 g/dL increase)	0.30	1.73	2.13	0.0401*
Arteriolosclerosis (grades 0–3)	-0.05	1.76	-0.36	0.7221
U-Prot during 10-year follow-up (grades 0–4)	0.25	1.58	1.84	0.0747
Edema (vs. no)	-0.24	1.88	-1.62	0.1150
C-reactive protein (1 mg/dL increase)	0.23	1.10	2.07	0.0464*
Oxford E1 (vs. no)	0.27	1.23	2.23	0.0327*
Model based on top 8 variables of |R| R^2 **= 0.45, AICc = 417.2**	β	**VIF**	t**-ratio**	P**-value**
BMI (1 kg/m^2 increase)	0.60	6.53	1.84	0.0745
Arteriolosclerosis (grades 0–3)	0.11	2.33	0.55	0.5875
Complement C3 (1 mg/dL increase)	0.11	1.47	0.70	0.4887
U-Prot during 10-year follow-up (grades 0–4)	0.22	1.76	1.27	0.2118
Body weight (1 kg increase)	-0.39	5.74	-1.28	0.2096
Increase in dipstick hematuria during 10-year (vs. no)	-0.01	1.63	-0.04	0.9688
Intimal thickening of the interlobular artery (grades 0–3)	0.14	1.85	0.82	0.4172
Perihilar hyalinosis (vs. no)	0.15	1.39	0.97	0.3413

*P < 0.05. Top 8 variables based on ML score (GP) or /point-biserial *r*/ were included in the multivariate model. Abbreviations: MaxGD, maximal glomerular diameter; ML, machine learning; GP, genetic programming; /*R*/, the absolute value of point-biserial correlation coefficient; R^2, *R*-squared; AICc, small-sample corrected Akaike Information Criterion; β, standardized partial regression coefficient; VIF, variance inflation factor; BMI, body mass index; C3, component 3; U-Prot, Urinary protein excretion; Oxford E1, the presence of endocapillary hypercellularity.

Discussion

We applied the advantages of ML to research using traditional statistical methods and had two main findings. First, the increased R^2, pseudo-R^2, C-statistic values, and the decreased AIC values observed in feature selection by ML compared with that observed in feature selection by point-biserial *r* suggest that ML techniques can handle large variable numbers and are useful for elucidating novel factors related to multifactorial diseases. Second, the weak association between MaxGD and eGFR demonstrates that MaxGD represents morbid GH rather than compensatory GH.

Traditional statistics based on *a priori* knowledge with a hypothesized model may delay clinical research progress, as few novel prognostic features are addressed in each study [3]. Contrastingly, ML has the power to reduce the dimensionality of variables [3,4], which is based on its intrinsic ability to formulate a predictive data-based model. Performing ML first allows us to account for potential linear and non-linear predictors without unconscious bias, avoiding *a priori* choice among potential predictors. Nonetheless, ML methods neither assess statistical inference nor offer valid inferences on feature importance; therefore, statistical testing should be performed. Considering that the advantages of ML techniques are flexibility and lack of *a priori* assumptions and that the advantages of traditional statistical approaches are simplicity and transparency of understanding [3,4], the order of our analysis in the present study was to first narrow down the number of features by ML and subsequently validate the features by traditional statistics, such as R^2, pseudo-R^2, AIC, and C-statistic. To overcome the overfitting problem [11] and generalize the ML results, we focused on the ability of ML to select features instead of focusing on its predictive models.

Furthermore, while traditional regression analyses suffer from multicollinearity in the presence of many variables, ML techniques alleviate collinearity limitations by leveraging penalization approaches [25,26]. Interestingly, the features selected by the ML method showed lower VIFs than those selected by point-biserial *r*, indicating that SR via GP is effective in avoiding the introduction of collinear variables. SR via GP [27,28] is a versatile and heuristic model that allows detailed analyses, even in datasets with a small sample size, partly because it can control the complexity of models to prevent overfitting of the training data. We confirmed that SR via GP is useful for identifying new risk factors and discriminating unrelated factors. As we routinely perform SR as the ML method, we used this technique. SR showed excellent results in the ML methods examined (Fig. 5, Table 5).

Table 5
AUC and other statistical metrics in nine machine learning models

Machine learning model	AUC	Accuracy	Precision	Recall	F-score
Linear Regression	0.536	0.488	0.333	0.467	0.389
Lasso Regression	0.736	0.698	0.563	0.600	0.581
Ridge Regression	0.695	0.651	0.500	0.533	0.714
Logistic Regression	0.662	0.721	0.636	0.467	0.857
Naïve Bayes	0.548	0.512	0.385	0.667	0.488
SVMs	0.615	0.721	0.800	0.267	0.964
Random Forrest	0.631	0.721	0.714	0.333	0.929
XG Boost	0.826	0.698	0.563	0.600	0.750
Symbolic regression via GP	0.774	0.767	0.727	0.533	0.615
Abbreviations: AUC, area under the curve; GP, genetic programming.					

Although GH is multifactorial and occurs in different pathophysiological conditions [7], it is technically difficult to distinguish true valuable risk factors from the many risk factors using a conventional statistical method. However, the significant power of factor selection via ML makes such discrimination possible by creating rankings for predictive features of patients with multifactorial chronic diseases. Here, the ML score (GP) for MaxGD ≥ 242.3 μm identified the top 8 features, including BMI, complement C3, serum total protein, arteriolosclerosis, **urinary protein excretion** (U-Prot) during the 10-year follow-up, edema, C-reactive protein, and the Oxford E1 score. Furthermore, eGFR was ranked 46th in the ML scores (GP), indicating that MaxGD is more relevant for current injuries, such as vascular damage, inflammation, and obesity, rather than past injuries represented by nephron disappearance.

Aging is associated with nephron loss/low eGFR; however, it remains controversial whether these findings are pathological or not [29–32]. Therefore, it is desirable to distinguish compensatory hypertrophy due to nephron loss/low GFR from morbid hypertrophy due to disease activity [7]. The key to understanding this question is the rightward shift [33] in the glomerular size distribution caused by nephron loss/low GFR and the threshold for morbid GH [7,33]. The five-sixths nephrectomized (or subtotal nephrectomized) model is a frequently used animal model of progressive kidney failure by nephron loss/low GFR. The glomerular diameter in subtotal nephrectomized rats increased to approximately 1.5 times the glomerular diameter of the control group, while that of the non-nephrectomized rats increased to approximately 1.1 times the glomerular diameter of the control group [34]. Therefore, GH > 1.5 times its original diameter (2.25 times area and 3.38 times volume, assuming glomeruli are spherical) may be morbid [7]. In a recent study examining individual glomerular size using magnetic resonance imaging-based glomerular morphology in a mouse model [35], the increase in glomerular volume due to aging showed a rightward shift in the glomerular size distribution in the range of < 3 times the volume. Similar results have been observed in

humans [30]. Furthermore, while American women with low nephron numbers did not demonstrate GH, American men with low nephron numbers showed marked GH [36]. Thus, compared with nephron loss, glomerular size could be a more direct indicator of disease severity.

To the best of our knowledge, our study is the first to identify predictive features for MaxGD using ML. The methodological novelty of our research included combining the exploratory strengths of ML with the validation strengths of conventional statistical methods, and our findings will contribute significantly to future clinical research. Furthermore, our findings on the discrimination between compensatory GH caused by nephron loss and pathological GH have significant clinical importance in nephrology.

One limitation of this study is that this study was observational; thus, any observed association does not prove causality. Although countermeasures for small sample sizes [13,14] were adopted, such as permutation tests and LOO-CV using SR via GP, the small sample size should be noted. Furthermore, since the comparison of superiority and inferiority between ML methods was not the main focus, we did not adopt the approach of a study on ML comparing the hit rates of predictive models created by ML. Lastly, although one of the strengths of ML is that it allows exploration of non-linear relationships, complete verification may not be possible with conventional statistical methods when non-linear factors are selected by ML. However, the study results, such as the increase of pseudo-R^2 and R^2 in the model based on factors selected by SR, indicate that the factor selection ability of ML is also excellent for linear factors.

In conclusion, ML demonstrated a weak association between MaxGD and eGFR, indicating that MaxGD represents morbid GH rather than compensatory GH. Furthermore, in a comparative validation using a conventional statistical technique, feature selection by ML avoided collinearity and increased pseudo-R^2 and R^2 values to a greater degree than feature selection using point-biserial *r*. Moreover, ML may be useful for identifying unknown risk factors and unrelated factors. Our method may be generalized to other types of medical research because of the procedural simplicity of using top-ranked features selected by ML.

Declarations

ETHICAL APPROVAL

This study was approved by the Ethics Committee of the Kameda Medical Center (No. 17-170) and followed the principles of the 1964 Helsinki Declaration.

INFORMED CONSENT

We used passive informed consent (opt-out) for patients, and all data were analyzed anonymously.

ACKNOWLEDGEMENTS

We express our gratitude to Dr. Takahiro Mochizuki (Deceased 25 June 2017) for his advice on this work and his contribution to medical care and research in Japan.

CONFLICT OF INTEREST STATEMENT

Toshio Mochizuki received honoraria for lectures from Otsuka Pharmaceutical Co. Toshio Mochizuki and Hiroshi Kataoka belong to an endowed department sponsored by Otsuka Pharmaceutical Co., Chugai Pharmaceutical Co., Kyowa Hakko Kirin Co., and JMS Co. All other authors have no conflicts of interest to declare.

FUNDING

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

DATA AVAILABILITY STATEMENT

The dataset analyzed and the programs developed in this study are available on GitHub as supplementary files (https://github.com/kiwindow/SymbolicRegression). The codes to perform SR via GP are written in Mathematica® (Wolfram Research Inc., Champaign, IL), and DataModeler® PRO or EDU (Evolved Analytics LLC, Rancho Santa Fe, CA) is needed. The program for SR via GP is uploaded as BuildModel.m along with the init.m file needed to import user-defined functions. The dataset is saved as newMaxGD2.xlsx and newMaxGD2.csv. Usage instructions are provided in README.md and ReadMeFirst.nb. The minimum commands explained in README.md are written on QuickStarte.nb file.

References

1. Beam, A. L. & Kohane, I. S. Big Data and Machine Learning in Health Care. Jama. **319**, 1317–1318 (2018).
2. Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. Nat Methods. **15**, 233–234 (2018).
3. Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N. & Fanos, V. Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. Medicina (Kaunas). **56** (2020).
4. Bzdok, D. Classical Statistics and Statistical Learning in Imaging Neuroscience. Front Neurosci. **11**, 543 (2017).
5. Deo, R. C. Machine Learning in Medicine. Circulation. **132**, 1920–1930 (2015).
6. Fabris, A. *et al.* Proteomic-based research strategy identified laminin subunit alpha 2 as a potential urinary-specific biomarker for the medullary sponge kidney disease. Kidney international. **91**, 459–468 (2017).

7. Kataoka, H., Mochizuki, T. & Nitta, K. Large Renal Corpuscle: Clinical Significance of Evaluation of the Largest Renal Corpuscle in Kidney Biopsy Specimens. Contrib Nephrol. **195**, 20–30 (2018).
8. Kataoka, H., Ohara, M., Honda, K., Mochizuki, T. & Nitta, K. Maximal glomerular diameter as a 10-year prognostic indicator for IgA nephropathy. Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association. **26**, 3937–3943 (2011).
9. Kataoka, H. *et al.* Maximum Glomerular Diameter and Oxford MEST-C Score in IgA Nephropathy: The Significance of Time-Series Changes in Pseudo-R(2) Values in Relation to Renal Outcomes. Journal of clinical medicine. **8** (2019).
10. Kataoka, H. *et al.* Time series changes in pseudo-R2 values regarding maximum glomerular diameter and the Oxford MEST-C score in patients with IgA nephropathy: A long-term follow-up study. PloS one. **15**, e0232885 (2020).
11. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. PloS one. **14**, e0224365 (2019).
12. Potter, D. M. A permutation test for inference in logistic regression with small- and moderate-sized data sets. Statistics in medicine. **24**, 693–708 (2005).
13. Valente, G., Castellanos, A. L., Hausfeld, L., De Martino, F. & Formisano, E. Cross-validation and permutations in MVPA: Validity of permutation strategies and power of cross-validation schemes. Neuroimage. **238**, 118145 (2021).
14. Ball, T. M., Squeglia, L. M., Tapert, S. F. & Paulus, M. P. Double Dipping in Machine Learning: Problems and Solutions. Biol Psychiatry Cogn Neurosci Neuroimaging. **5**, 261–263 (2020).
15. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. Front Neuroinform. **8**, 14 (2014).
16. Hauber, A. B. *et al.* Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health. **19**, 300–315 (2016).
17. Hurvich, C. M. & Tsai, C. L. Model selection for extended quasi-likelihood models in small samples. Biometrics. **51**, 1077–1084 (1995).
18. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. **143**, 29–36 (1982).
19. Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. Jama. **247**, 2543–2546 (1982).
20. Yoo, W. *et al.* A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int J Appl Sci Technol. **4**, 9–19 (2014).
21. Vittinghoff, E. & McCulloch, C. E. Relaxing the rule of ten events per variable in logistic and Cox regression. American journal of epidemiology. **165**, 710–718 (2007).
22. Austin, P. C. & Steyerberg, E. W. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. **68**, 627–636 (2015).

23. Curtis, M. J. *et al.* Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol. **172**, 3461–3471 (2015).
24. Vatcheva, K. P., Lee, M., McCormick, J. B. & Rahbar, M. H. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiology (Sunnyvale). **6** (2016).
25. Doupe, P., Faghmous, J. & Basu, S. Machine Learning for Health Services Researchers. Value Health. **22**, 808–815 (2019).
26. La Cava, W. & Moore, J. H. Learning feature spaces for regression with genetic programming. Genet Program Evolvable Mach. **21**, 433–467 (2020).
27. Murari, A., Lungaroni, M., Peluso, E., Craciunescu, T. & Gelfusa, M. A Model Falsification Approach to Learning in Non-Stationary Environments for Experimental Design. Scientific reports. **9**, 17880 (2019).
28. Tan, K. C., Yu, Q., Heng, C. M. & Lee, T. H. Evolutionary computing for knowledge discovery in medical diagnosis. Artif Intell Med. **27**, 129–154 (2003).
29. Hughson, M. D., Hoy, W. E. & Bertram, J. F. Progressive Nephron Loss in Aging Kidneys: Clinical-Structural Associations Investigated by Two Anatomical Methods. Anat Rec (Hoboken). **303**, 2526–2536 (2020).
30. Denic, A. *et al.* The Substantial Loss of Nephrons in Healthy Human Kidneys with Aging. Journal of the American Society of Nephrology: JASN. **28**, 313–320 (2017).
31. Denic, A., Glassock, R. J. & Rule, A. D. Structural and Functional Changes With the Aging Kidney. Advances in chronic kidney disease. **23**, 19–28 (2016).
32. Hommos, M. S., Glassock, R. J. & Rule, A. D. Structural and Functional Changes in Human Kidneys with Healthy Aging. Journal of the American Society of Nephrology: JASN. **28**, 2838–2844 (2017).
33. 33.. How to read clinical journals: II. To learn about a diagnostic test. Canadian Medical Association journal. **124**, 703–710 (1981).
34. Santos, L. S., Chin, E. W., Ioshii, S. O. & Tambara Filho, R. Surgical reduction of the renal mass in rats: morphologic and functional analysis on the remnant kidney. Acta cirurgica brasileira. **21**, 252–257 (2006).
35. Geraci, S. *et al.* Combining new tools to assess renal function and morphology: a holistic approach to study the effects of aging and a congenital nephron deficit. American journal of physiology. Renal physiology. **313**, F576-F584 (2017).
36. Puelles, V. G. *et al.* Glomerular hypertrophy in subjects with low nephron number: contributions of sex, body size and race. Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association. **29**, 1686–1695 (2014).

Figures

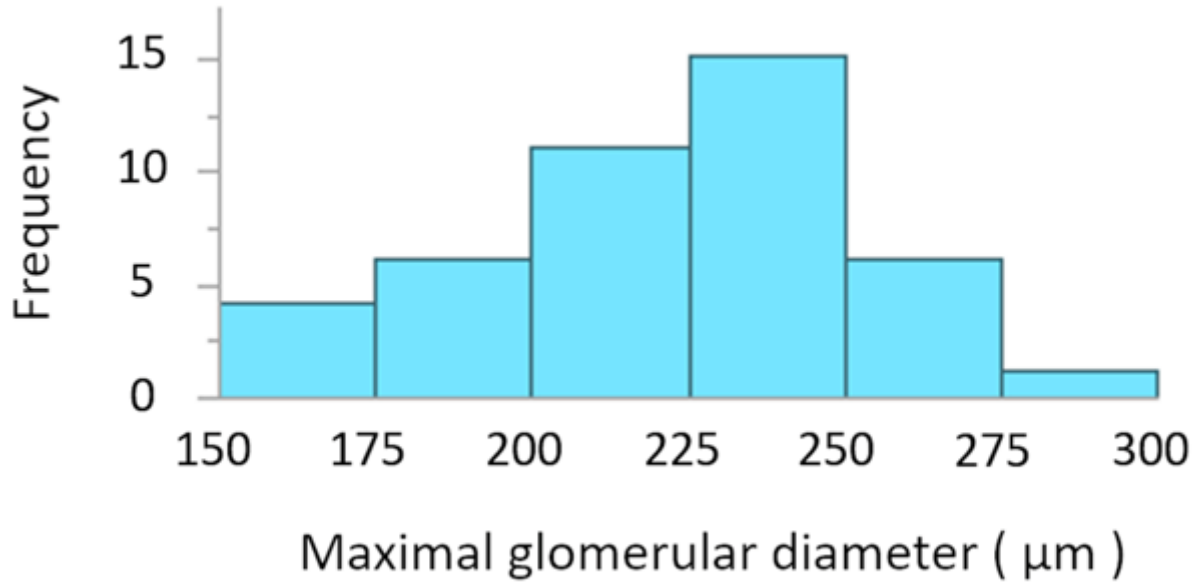

Figure 1

Histogram of MaxGD

The distribution of MaxGD is illustrated as light blue histograms. Abbreviation: MaxGD, maximal glomerular diameter.

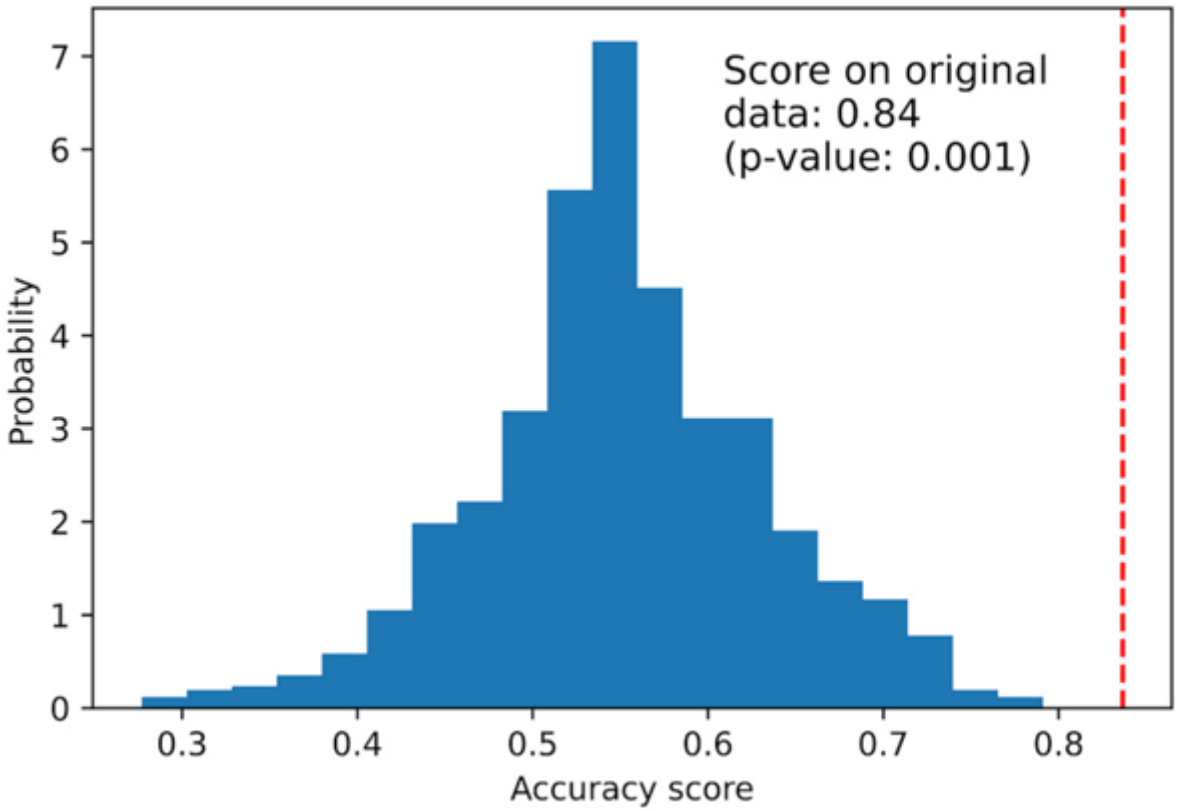

Figure 2

Permutation test results with the original dataset (permutation test scores for the classifier of MaxGD≥242.3 μm)

The distribution of accuracy score for the permuted data is illustrated as blue histograms. It represents the result of 5,000 permutation tests for assessing classifier performance when selecting the 60 most discriminative variables. The red dotted line indicates the accuracy score value (0.84) obtained by the classifier in the original dataset (permutation P-value, 0.001). Abbreviation: GD, glomerular diameter.

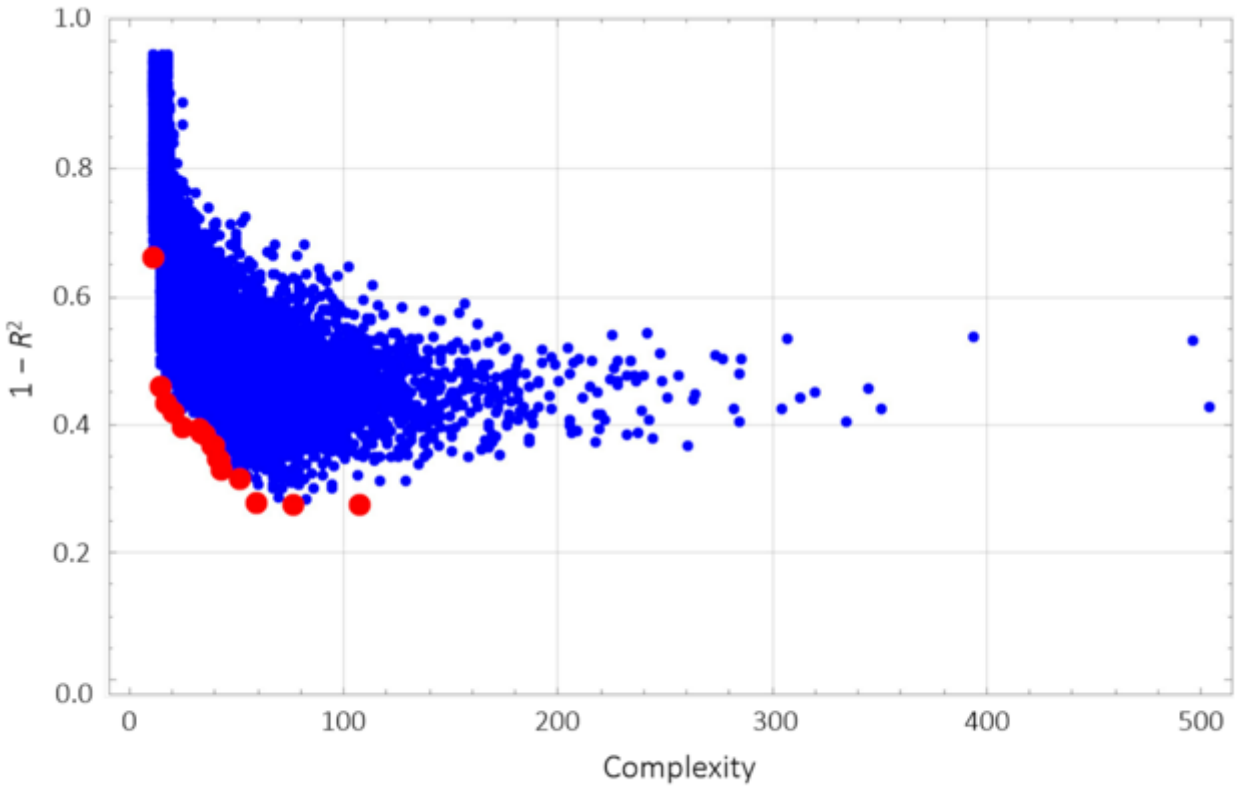

Figure 3

Distribution of functions generated with symbolic regression via genetic programming Generated functions are plotted on the function space, where the horizontal axis represents the complexity of a function, and the vertical axis represents $1-R^2$ or error. In total, 19,437 predictive functions are generated with symbolic regression via genetic programming. Each dot represents one function, and the red dots represent functions on the Pareto front that are candidates for optimized functions with ensemble learning.

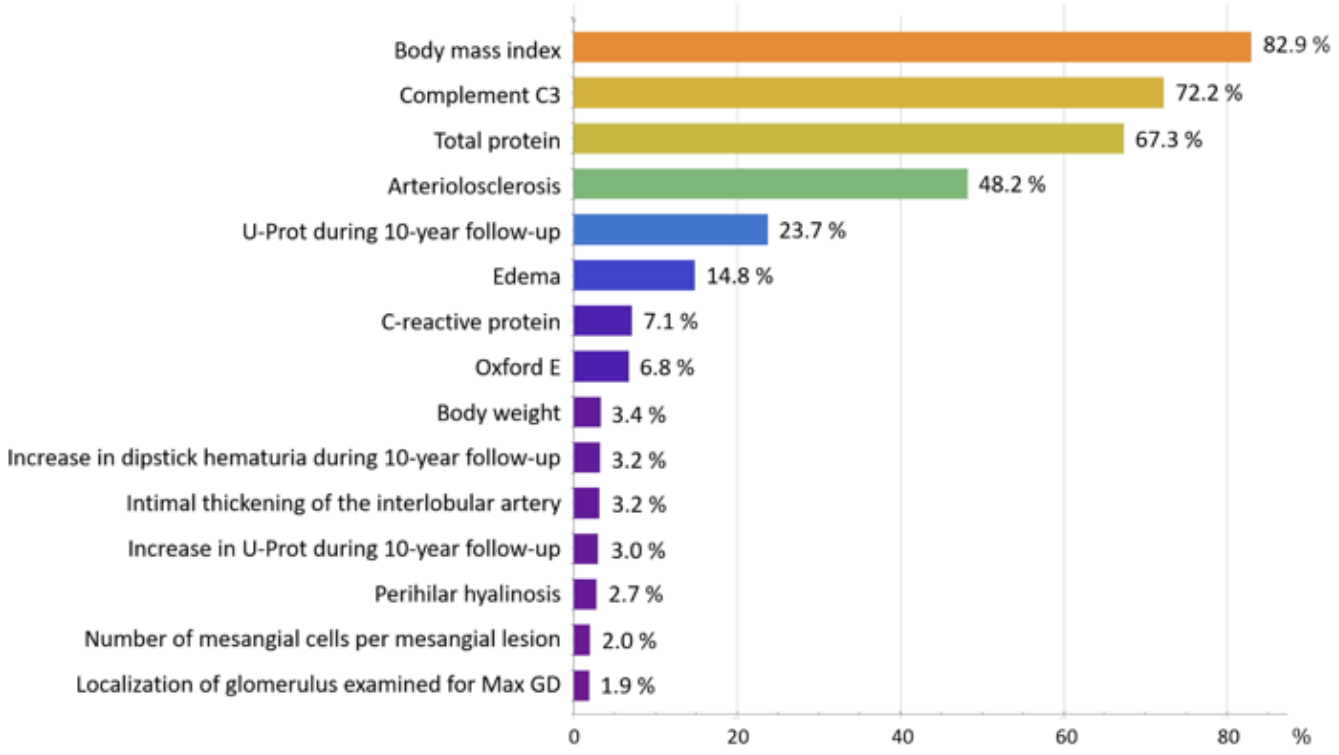

Figure 4

Frequently utilized predictive variables in selected models via machine learning

The 15 most frequently utilized predictive variables in 1,819 predictive functions, which are selected among 19,437 models generated in the leave-one-out cross-validation using symbolic regression via genetic programming, are listed in descending order. The horizontal axis represents the percentage at which each predictive variable is utilized in all 1,819 predictive functions. Abbreviations: C3, component 3; U-Prot, urinary protein excretion; Oxford E1, the presence of endocapillary hypercellularity; MaxGD, maximal glomerular diameter.

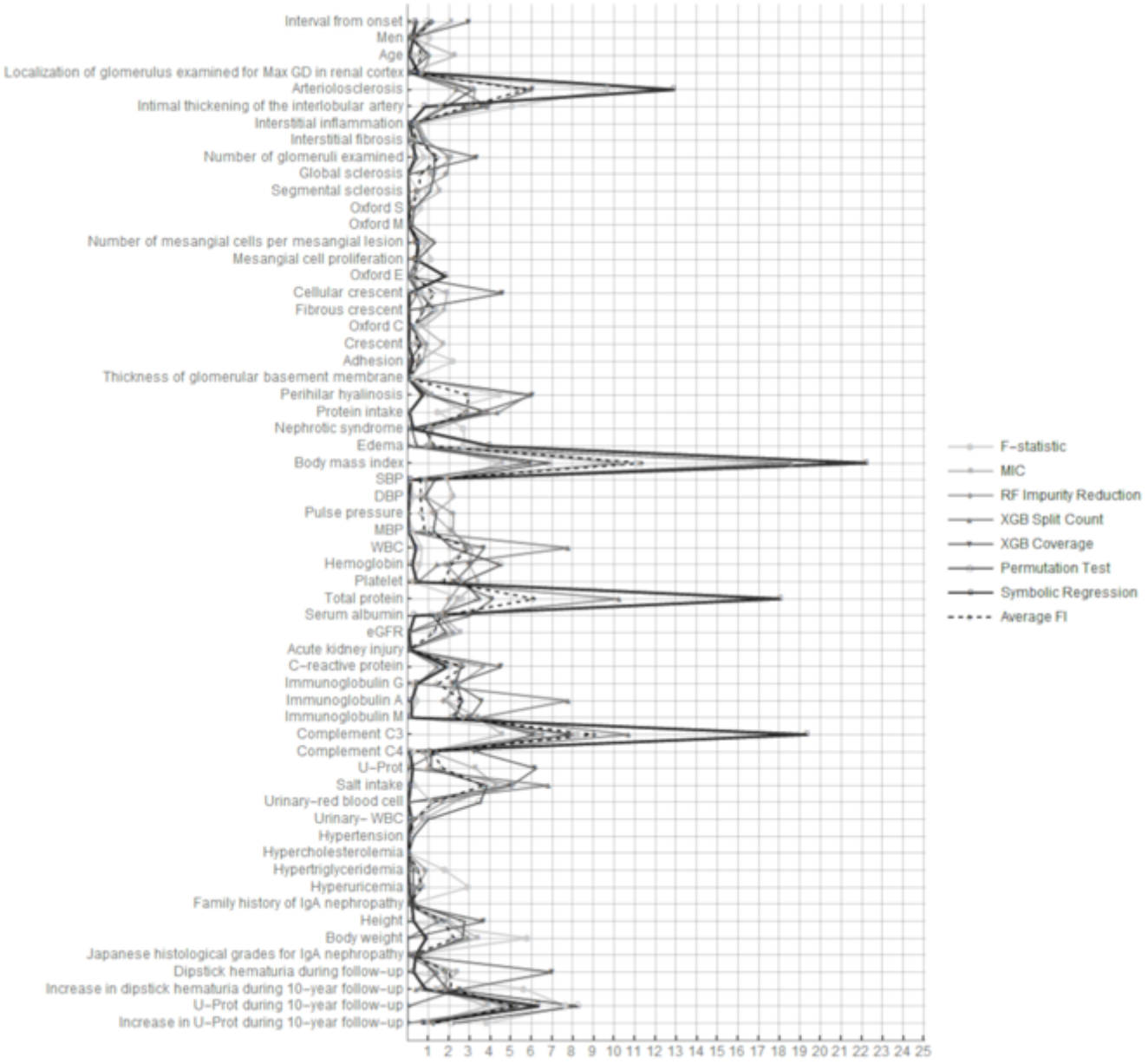

Figure 5

Feature importance scores using eight machine learning models

Feature importance score of ML score (GP) is calculated as 100 times the ML score and expressed as Symbolic Regression. Abbreviations: MaxGD, maximal glomerular diameter; eGFR, estimated glomerular filtration rate; WBC, white blood cell; U-Prot, Urinary protein excretion; SBP, systolic blood pressure; MBP, mean blood pressure; DBP, diastolic blood pressure; Complement C4, complement component 4; Complement C3, complement component 3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTALMATERIAL20220411GPKamedaIgANMaxGDSRMaxGD.docx