

# Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques

Hiroshi Kanegae BSc<sup>1,2</sup> | Kenji Suzuki BA<sup>3</sup> | Kyohei Fukatani MBA<sup>4</sup> | Tetsuya Ito MEng<sup>4</sup> | Nakahiro Harada BEng<sup>4</sup> | Kazuomi Kario MD, PhD<sup>1</sup> 

<sup>1</sup>Department of Medicine, Division of Cardiovascular Medicine, Jichi Medical University School of Medicine, Tochigi, Japan

<sup>2</sup>Genki Plaza Medical Center for Health Care, Tokyo, Japan

<sup>3</sup>The Japan Health Promotion Foundation, Tokyo, Japan

<sup>4</sup>Fukuda Denshi Co., Ltd., Tokyo, Japan

## Correspondence

Kazuomi Kario, MD, PhD, Department of Medicine, Division of Cardiovascular Medicine, Jichi Medical University School of Medicine, 3311-1 Yakusiji, Shimotsuke, Tochigi 329-0498, Japan.  
Email: kkario@jichi.ac.jp

## Funding information

This study was funded by a research grant from Fukuda Denshi Co.

## Abstract

Hypertension is a significant public health issue. The ability to predict the risk of developing hypertension could contribute to disease prevention strategies. This study used machine learning techniques to develop and validate a new risk prediction model for new-onset hypertension. In Japan, Industrial Safety and Health Law requires employers to provide annual health checkups to their employees. We used 2005–2016 health checkup data from 18 258 individuals, at the time of hypertension diagnosis [Year (0)] and in the two previous annual visits [Year (-1) and Year (-2)]. Data were entered into models based on machine learning methods (XGBoost and ensemble) or traditional statistical methods (logistic regression). Data were randomly split into a derivation set (75%, n = 13 694) used for model construction and development, and a validation set (25%, n = 4564) used to test performance of the derived models. The best predictor in the XGBoost model was systolic blood pressure during cardio-ankle vascular index measurement at Year (-1). Area under the receiver operator characteristic curve values in the validation cohort were 0.877, 0.881, and 0.859 for the XGBoost, ensemble, and logistic regression models, respectively. We have developed a highly precise prediction model for future hypertension using machine learning methods in a general normotensive population. This could be used to identify at-risk individuals and facilitate earlier non-pharmacological intervention to prevent the future development of hypertension.

## 1 | INTRODUCTION

It has been estimated that hypertension is the cause of approximately 13% of all deaths worldwide each year.<sup>1</sup> In light of the world's population growth and aging demographic, hypertension is a global burden, along with other cardiovascular and age-related diseases.<sup>2,3</sup>

Early intervention with lifestyle modifications and treatment of "prehypertension" may reduce the incidence and long-term consequences of clinical hypertension.<sup>4–7</sup> Recent guidelines lowered the recommended thresholds for diagnosing hypertension or abnormal "elevated blood pressure (BP)" and the BP goal during antihypertensive therapy.<sup>8–10</sup> Therefore, the ability to predict an

individual's risk of developing hypertension would be helpful for clinicians. They could then plan and prescribe personalized lifestyle modifications or make therapeutic decisions designed to prevent or postpone the development of hypertension. There are several models available to predict the risk of new-onset hypertension; these have been developed in Western and Asian countries using traditional statistical methods (eg, Cox regression or logistic regression).<sup>11,12</sup>

Arterial stiffness is increasingly being recognized as making an important contribution to increases in systolic BP (SBP) and the development of hypertension in general populations, independent of traditional hypertension risk factors.<sup>13–16</sup> In addition, arterial

stiffness has been associated with increased risk of cardiovascular disease, cardiovascular events, and all-cause mortality.<sup>17–22</sup>

The cardio-ankle vascular index (CAVI) is an indicator of arterial stiffness and has been associated with cardiovascular risk factors and cardiovascular disease.<sup>23,24</sup> CAVI is one of the vascular measures of the systemic hemodynamic atherothrombotic syndrome (SHATS), which is characterized by a vicious cycle of BP variability and vascular disease contributing to cardiovascular events.<sup>25,26</sup> In addition, CAVI has been shown to predict future development of hypertension independently of risk factors in a general normotensive population.<sup>27</sup>

Artificial intelligence and machine learning (ML) are poised to influence nearly every aspect of the human condition, and cardiology is no exception.<sup>28</sup> ML algorithms are typically used without making many assumptions about the underlying data.<sup>28</sup> This study describes the development of a model involving CAVI for the prediction of future hypertension development using ML methods in a general population.

## 2 | METHODS

### 2.1 | Study subjects

In Japan, Industrial Safety and Health Law requires employers to provide annual health checkups to their employees. This study included individuals who underwent health checkups at the Japan Health Promotion Foundation in at least three successive years from 2005 to 2016, were not being treated with antihypertensive medication, and had office BP < 140/90 mm Hg at the two checkup visits prior to being diagnosed with hypertension [defined as Year (-2) and Year (-1)]. Hypertension was diagnosed at the checkup visit in Year (0).

The study was conducted according to the principles of the Declaration of Helsinki. The study protocol was approved by an ethics committee of the Jichi Medical University School of Medicine (Approval No. RIN A17-HEN 119). It was not necessary to obtain informed consent from subjects because identifying information (eg, names and addresses) was not collected. Subjects had the right to opt out of the study. The fact that Jichi Medical University was using health checkup data for this study was disclosed to the public on the Japan Health Promotion Foundation website.

### 2.2 | Outcomes

The primary end point was new-onset hypertension (defined as SBP/diastolic BP [DBP] ≥ 140/90 mm Hg or the initiation of antihypertensive medication with self-reported hypertension) at Year (0).

### 2.3 | Assessments

Annual health checkup visits included recording of an individual's medical history, lifestyle factors, anthropometric measurements,

and biochemical measurements. Full details have been described previously.<sup>27</sup>

### 2.4 | Statistical analyses

The prediction model for new-onset hypertension was constructed to predict an individual's hypertension risk at Year (0) based on variables at Year (-1), Year (-2), and changes from Year (-2) to Year (-1). The last observation carried forward method was used if a subject did not undergo a health checkup at Year (-1) or Year (-2). For missing variables, mean imputation was used for continuous variables and mode imputation was used for categorical variables. Variables with a skewed distribution were log-transformed to obtain a normal distribution. The data were randomly split into a derivation set (75%, n = 13 694), used for model construction and development, and a validation set (25%, n = 4564), used to test performance of the derived model (Figure 1).

We used a scalable end-to-end tree boosting system called XGBoost model, which is widely used by data scientists to achieve state-of-the-art results on many ML challenges.<sup>28</sup> We also used an ensemble model, which is a supervised learning technique for combining multiple weak models to produce a strong model, and a logistic regression model (a traditional method). The bagging

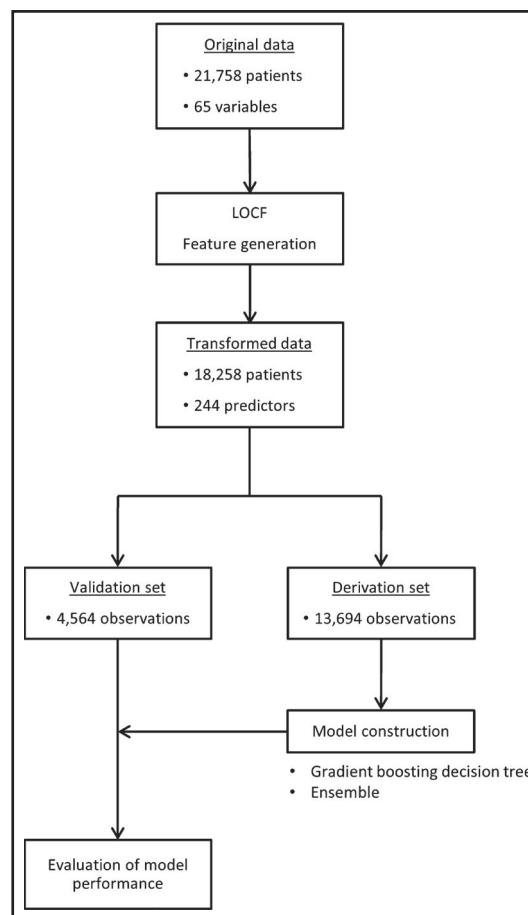


FIGURE 1 Study flow chart

**TABLE 1** Subject demographic and clinical characteristics at Year (-2)

Variables	Subjects (n = 18 258)
Age, years	46.4 ± 12.1
Men, %	44.6
Body mass index, kg/m <sup>2</sup>	22.3 ± 3.2
Waist, cm	79.1 ± 7.4
Clinic SBP, mm Hg	118.7 ± 11.2
Clinic DBP, mm Hg	70.0 ± 8.7
CAVI	7.5 ± 0.9
SBP at CAVI measurement, mm Hg	116.1 ± 12.0
DBP at CAVI measurement, mm Hg	72.1 ± 8.9
High-density lipoprotein cholesterol, mg/dL	69.5 ± 18.1
Low-density lipoprotein cholesterol, mg/dL	126.2 ± 26.7
Uric acid, mg/dL	5.0 ± 1.3
Fasting glucose, mg/dL	87.0 ± 11.8
Diabetes mellitus, %	1.6
Chronic kidney disease, %	0.5
Smoking status, %	
Non-smoker	71.0
Past smoker	11.8
Current smoker	17.2
Alcohol use, %	
0 d/wk	35.0
1-2 d/wk	19.2
3-4 d/wk	7.0
5-6 d/wk	6.1
7 d/wk	11.5

Note: Values are expressed as the mean ± SD or proportion of patients (%).

Abbreviations: CAVI, cardio-ankle vascular index; DBP, diastolic blood pressure; SBP, systolic blood pressure.

method<sup>29</sup> was used to combine three models—regularized logistic regression model, random forest model, and XGBoost model. The receiver operating characteristic (ROC) curve and validated area under the curve (AUC) value were derived to evaluate the performance of the derived prediction model. All analyses were performed with R version 3.4.1 (The R Foundation for Statistical Computing).

## 3 | RESULTS

### 3.1 | Subjects

A total of 18 258 subjects were included (mean age 46 years, 45% men, low prevalence of diabetes and chronic kidney disease) (Table 1). The number of cases of new-onset hypertension identified was 2672.

**TABLE 2** The top 20 predictors in the XGBoost model

Rank	Variable	Relative importance (%)
1	SBP at Year (-1) CAVI measurement	100.0
2	Clinic SBP at Year (-1)	57.3
3	DBP at Year (-1) CAVI measurement	47.8
4	SBP at Year (-2) CAVI measurement	40.0
5	Clinic SBP at Year (-2)	26.4
6	Clinic DBP at Year (-2)	23.3
7	DBP at Year (-2) CAVI measurement	23.2
8	Clinic DBP at Year (-1)	12.4
9	Body mass index at Year (-1)	10.6
10	Age at Year (-2)	10.3
11	Body mass index at Year (-2)	8.5
12	Age at Year (-1)	7.3
13	CAVI at Year (-2)	7.2
14	Clinic SBP by SBP at CAVI measurement at Year (-1)	7.2
15	Waist at Year (-1)	7.0
16	Triglycerides at Year (-2)	6.7
17	Clinic DBP by DBP at CAVI measurement at Year (-1)	6.6
18	CAVI at Year (-1)	6.6
19	ALP at Year (-1)	6.6
20	Fasting glucose at Year (-2)	6.1

Abbreviations: ALP, alkaline phosphatase; CAVI, cardio-ankle vascular index; DBP, diastolic blood pressure; SBP, systolic blood pressure.

### 3.2 | Predictors of new-onset hypertension

As expected, increasing values of clinic SBP and DBP were important predictors of new-onset hypertension, but SBP during CAVI measurement in the year before hypertension onset was the most important predictor of future new-onset hypertension (Table 2). Different BP and BP during CAVI measures comprised the top eight predictors of future hypertension. These were followed by increasing body mass index (BMI), age, CAVI, waist circumference, triglyceride levels, alkaline phosphatase levels, and fasting glucose (Table 2).

### 3.3 | Model performance

Table 3 and Figure 2 show the AUC, precision, recall, and ROC curves for the XGBoost, ensemble, and logistic regression models in both the derivation and validation sets. The prediction model using XGBoost achieved a fitted AUC of 0.976 in the derivation set. This model also performed well when applied to the validation set (AUC 0.876). The ensemble method-based prediction model achieved the best predictive performance: AUC of 0.992 and 0.881 in the derivation and validation sets, respectively. The prediction model using logistic regression had

**TABLE 3** Prediction results

Model		AUC	Precision (PPV)	Recall (True positive rate)
XGBoost	Derivation	0.976	0.944	0.667
	Validation	0.877	0.601	0.317
Ensemble	Derivation	0.992	0.976	0.670
	Validation	0.881	0.635	0.253
Logistic	Derivation	0.855	0.604	0.265
	Validation	0.859	0.638	0.290

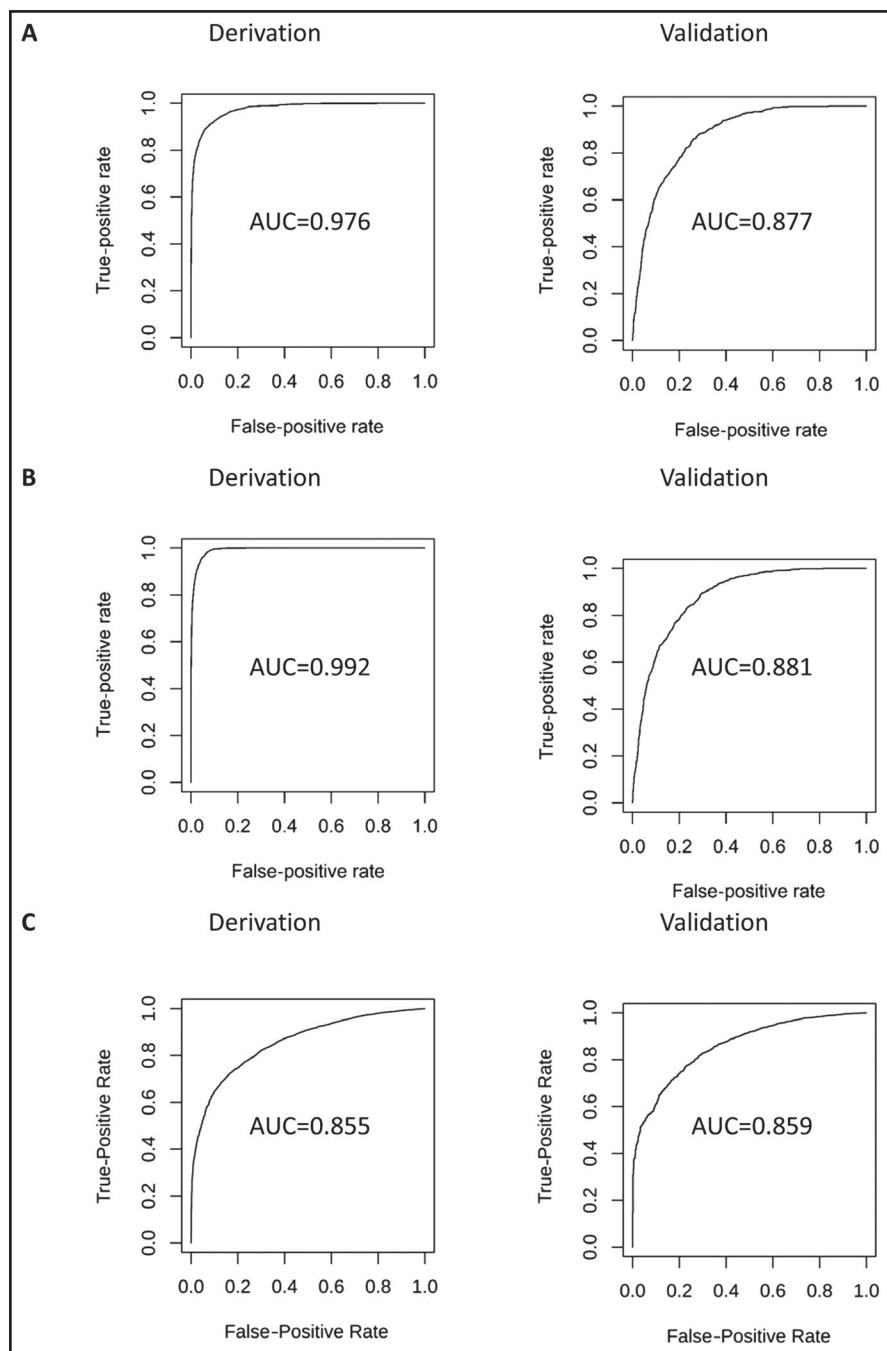
Abbreviations: AUC, area under the curve; PPV, positive predictive value.

the lowest predictive performance of all three models: AUC of 0.855 and 0.859 in the derivation and validation sets, respectively.

## 4 | DISCUSSION

We used ML to develop a highly precise prediction model for future hypertension in a general population. The performance of the ensemble model for new-onset hypertension was better than that of the XGBoost and logistic regression models.

Traditional statistical models, such as logistic regression or Cox regression models, require a number of important assumptions to



**FIGURE 2** Receiver operating characteristic curves for each model: (A) XGBoost model; (B) ensemble model; and (C) logistic regression model

be met (eg independence of observations and no multicollinearity among variables).<sup>28</sup> In contrast, ML algorithms typically make fewer assumptions about the underlying data.<sup>28</sup> This results in algorithms that are generally more accurate for prediction and classification.<sup>28</sup> Although our prediction model based on logistic regression performed well in both the derivation and validation sets (AUC 0.855 and 0.859, respectively), its performance was below those of models generated using ML methods.

In this study, BP during CAVI measurement and clinic BP at health checkups in the year or two prior to hypertension diagnosis were the top eight predictors of new-onset hypertension. Although sitting clinic BP is traditionally used to diagnose hypertension, we found that SBP at CAVI measurement in the supine position at Year (-1) was the strongest predictor of future hypertension. This suggests that assessment of BP in different settings is important to allow precise prediction of new-onset hypertension. Based on the findings of this study, BP measurement in the supine position at rest may be more useful for predicting future hypertension compared with BP measurement in the sitting position.

The ML-based analysis was able to incorporate all BP measures into the same model. In contrast, the traditional regression model could not enter the sitting and supine BP measures with high collinearity into the same model. Many previous prediction models included age and BMI<sup>12</sup> because these have been strongly associated with new-onset hypertension. Age and BMI were also important predictors of hypertension in our model, after BP measurements. CAVI, which has been directly associated with new-onset hypertension,<sup>27</sup> was another important predictor in our model. Obesity-related metabolic risk factors, such as glucose and triglycerides, were lower ranked but significant predictors of hypertension in the top 20 predictors in the XGBoost model. Thus, both metabolic and vascular components predict the future development of hypertension.

Ye et al<sup>30</sup> used a ML algorithm (XGBoost) to construct and prospectively validated a risk prediction model for future 1-year risk of incident essential hypertension using electronic health record-derived data from more than 1.5 million people. The model achieved predictive accuracy of 0.917 and 0.870 in retrospective and prospective (validation) cohorts, respectively. Similar predictive performance values from our prediction model using the XGBoost were achieved in our study (0.976 and 0.876, respectively, in the derivation and validation sets). This shows that our model based on variables at 2 years was better than the previous model based on variables at 1 year.

The strengths of this study are its sample size, and the uniform and standardized approach to data collection. However, there were also several limitations, largely due to the characteristics of the population, and the methods used to measure the study parameters. We only measured clinic BP, which is unable to identify masked hypertension and white-coat hypertension. In addition, we developed the prediction model using only health checkup data for two successive years. Finally, the findings are only applicable to Japanese patients. Therefore, our model needs to be validated in other populations.

## ACKNOWLEDGMENTS

Statistical analyses using ML methods were independently conducted by Naoya Fukawa and Hiroaki Shibata, Fujitsu Research Institute, Tokyo, Japan. We thank Masaichi Suzuki and Yoshiomi Uba, Fukuda Denshi Co., Tokyo, Japan for the coordination of this study. Medical writing assistance was provided by Nicola Ryan, independent medical writer.

## CONFLICT OF INTEREST

Kario Kazuomi received a research grant from Fukuda Denshi Co., Ltd.. The other authors have no conflicts of interest to declare.

## AUTHOR CONTRIBUTIONS

HK conducted the study and data analysis, and had the primary responsibility of writing this paper. KS collected the data. KF reviewed/edited the manuscript and contributed to the Discussion section. TI advised on the data analysis and reviewed/edited the manuscript. NH advised on the data analysis and reviewed/edited the manuscript. KK supervised the conduct of the study and data analysis.

## PERSPECTIVES

We have developed a robust ML-based prediction model for hypertension. Although the accuracy of this model needs to be tested in additional population samples, we hope that it is used to identify at-risk individuals and facilitate earlier non-pharmacological intervention to prevent the future development of hypertension.

## ORCID

Kazuomi Kario  <https://orcid.org/0000-0002-8251-4480>

## REFERENCES

1. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. [http://www.who.int/healthinfo/global\\_burden\\_disease/global\\_health\\_risks/en/](http://www.who.int/healthinfo/global_burden_disease/global_health_risks/en/). Accessed May 2, 2019.
2. Turana Y, Tengkawan J, Chia YC, et al. Hypertension and dementia: a comprehensive review from the HOPE Asia Network. *J Clin Hypertens (Greenwich)*. 2019;21:1091-1098.
3. World Health Organization. A global brief on hypertension. [http://www.who.int/cardiovascular\\_diseases/publications/global\\_brief\\_hypertension/en/](http://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension/en/). Accessed May 2, 2019.
4. The hypertension prevention trial: three-year effects of dietary changes on blood pressure. Hypertension Prevention Trial Research Group. *Arch Intern Med*. 1990;150:153-162.
5. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high-normal blood pressure. The trials of hypertension prevention, phase II. The trials of Hypertension Prevention Collaborative Research Group. *Arch Intern Med*. 1997;157:657-667.
6. Julius S, Nesbitt SD, Egan BM, et al. Feasibility of treating pre-hypertension with an angiotensin-receptor blocker. *N Engl J Med*. 2006;354:1685-1697.
7. Skov K, Eiskjaer H, Hansen HE, Madsen JK, Kvist S, Mulvany MJ. Treatment of young subjects at high familial risk of future hypertension with an angiotensin-receptor blocker. *Hypertension*. 2007;50:89-95.
8. Umemura S, Arima H, Arima S, et al. The Japanese Society of Hypertension Guidelines for the Management of Hypertension (JSH 2019). *Hypertens Res*. 2019;42:1235-1481.

9. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol.* 2018;71:2199-2269.
10. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension. *Eur Heart J.* 2018;39:3021-3104.
11. Kanegae H, Oikawa T, Suzuki K, Okawara Y, Kario K. Developing and validating a new precise risk-prediction model for new-onset hypertension: the Jichi Genki hypertension prediction model (JG model). *J Clin Hypertens (Greenwich).* 2018;20:880-890.
12. Sun D, Liu J, Xiao L, et al. Recent development of risk-prediction models for incident hypertension: an updated systematic review. *PLoS ONE ONE.* 2017;12:e0187240.
13. Dernellis J, Panaretou M. Aortic stiffness is an independent predictor of progression to hypertension in nonhypertensive subjects. *Hypertension.* 2005;45:426-431.
14. Liao D, Arnett DK, Tyroler HA, et al. Arterial stiffness and the development of hypertension. The ARIC study. *Hypertension.* 1999;34:201-206.
15. Najjar SS, Scuteri A, Shetty V, et al. Pulse wave velocity is an independent predictor of the longitudinal increase in systolic blood pressure and of incident hypertension in the Baltimore Longitudinal Study of Aging. *J Am Coll Cardiol.* 2008;51:1377-1383.
16. Takase H, Dohi Y, Toriyama T, et al. Brachial-ankle pulse wave velocity predicts increase in blood pressure and onset of hypertension. *Am J Hypertens.* 2011;24:667-673.
17. Boutouyrie P, Tropeano AJ, Asmar R, et al. Aortic stiffness is an independent predictor of primary coronary events in hypertensive patients: a longitudinal study. *Hypertension.* 2002;39:10-15.
18. Laurent S, Boutouyrie P, Asmar R, et al. Aortic stiffness is an independent predictor of all-cause and cardiovascular mortality in hypertensive patients. *Hypertension.* 2001;37:1236-1241.
19. Park K-H, Park WJ, Kim M-K, et al. Noninvasive brachial-ankle pulse wave velocity in hypertensive patients with left ventricular hypertrophy. *Am J Hypertens.* 2010;23:269-274.
20. Stefanadis C, Dernellis J, Tsiamis E, et al. Aortic stiffness as a risk factor for recurrent acute coronary events in patients with ischaemic heart disease. *Eur Heart J.* 2000;21:390-396.
21. Vlachopoulos C, Aznaouridis K, Stefanadis C. Prediction of cardiovascular events and all-cause mortality with arterial stiffness: a systematic review and meta-analysis. *J Am Coll Cardiol.* 2010;55:1318-1327.
22. Yamashina A, Tomiyama H, Arai T, et al. Brachial-ankle pulse wave velocity as a marker of atherosclerotic vascular damage and cardiovascular risk. *Hypertens Res.* 2003;26:615-622.
23. Matsushita K, Ding N, Kim ED, et al. Cardio-ankle vascular index and cardiovascular disease: systematic review and meta-analysis of prospective and cross-sectional studies. *J Clin Hypertens (Greenwich).* 2019;21:16-24.
24. Tanaka A, Tomiyama H, Maruhashi T, et al. Physiological diagnostic criteria for vascular failure. *Hypertension.* 2018;72:1060-1071.
25. Kario K. Systemic hemodynamic atherothrombotic syndrome (SHATS): Diagnosis and severity assessment score. *J Clin Hypertens (Greenwich).* 2019;21:1011-1015.
26. Kario K, Chirinos JA, Townsend R, et al. Systemic hemodynamic atherothrombotic syndrome (SHATS)—coupling vascular disease and blood pressure variability: proposed concept from Pulse of Asia. *Prog Cardiovasc Dis.* 2019; in press.
27. Kario K, Kanegae H, Oikawa T, Suzuki K. Hypertension is predicted by both large and small artery disease. *Hypertension.* 2019;73:75-83.
28. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *arXiv.* 2016.
29. Breiman L. Bagging predictors. *Mach Learn.* 1996;24:123-140.
30. Ye C, Fu T, Hao S, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. *J Med Internet Res.* 2018;20:e22.

**How to cite this article:** Kanegae H, Suzuki K, Fukatani K, Ito T, Harada N, Kario K. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. *J Clin Hypertens.* 2020;22:445-450. <https://doi.org/10.1111/jch.13759>