**Supplementary Appendix**

**Supplementary Methods**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | ***Topic*** | **Page no.** |
| **1** | ***Dataset and features*** | **2–3** |
| ***1.1*** | ***Physical examinations*** | **3** |
| ***1.2*** | ***Biochemical method*** | **3** |
| ***1.3*** | ***History*** | **4–5** |
| ***1.4*** | ***Treatment information*** | **5** |
| ***1.5*** | ***Assay methods of plasma aldosterone concentration and plasma renin activity*** | **5–6** |
| **2** | ***Process to develop the machine learning algorithms*** | **6** |
| ***2.1*** | ***Data preparation*** | **6–7** |
| **2.2** | ***General process of model construction and evaluation*** | **7–8** |
| **2.3** | ***Imputation of missing values by MissForest*** | **9** |
| **2.4** | ***Oversampling by SMOTE*** | **9–10** |
| **2.5** | ***Machine-learning algorithms*** | **10** |
| **3** | ***References*** | **11–13** |

1. ***Dataset and features***

The handling of datasets and features is shown in detail. First, we created two datasets from the JRAS database: a screening dataset and confirmatory dataset. The screening dataset contained 46 variables, including age, sex, body mass index (BMI), age at hypertension onset, period of time with hypertension, family history of hypertension, smoking, drinking, systolic (SBP) and diastolic blood pressure (DBP) at first visit, SBP with antihypertensive treatment, DBP with antihypertensive treatment, and daily defined doses (DDD) of antihypertensive agents based on the ATC/DDD index of 2020 (1), a scale that can quantify the type of antihypertensive drug and its potency, and published by the WHO Collaborating Centre for Drug Statistics Methodology annually. Biochemical examination variables were the following: serum potassium level (s-K) at first visit; s-K after medication; daily potassium supplementation dose; serum levels of sodium (s-Na), chloride (s-Cl), creatinine (s-Cr), and uric acid (s-UA); blood urea nitrogen (BUN); estimated glomerular filtration rate (eGFR); urine protein; fasting blood sugar (FBA); HbA1c; total cholesterol (TC); triglycerides (TG); and LDL and HDL cholesterol levels. Endocrinological examination variables were PAC, plasma renin activity (PRA), and ARR. Urine protein was semi-quantitatively examined using five levels: “negative,” “trace,” “1+,” “2+,” and “3+.”

The JRAS database has information related to the medical history and treatment of diseases, such as stroke, ischemic heart disease (IHD), heart failure, atrial fibrillation, chronic kidney disease (CKD), hyperuricemia, diabetes, dyslipidemia, sleep apnea syndrome (SAS), depression, periodic paralysis, and pregnancy-induced hypertension. The diagnostic criteria for these diseases are shown in the Appendix. Additionally, we collected information on therapeutic drugs used in the treatment of diabetes and dyslipidemia as one of the features.

The confirmatory test dataset, which contained 59 features, included the variables of the screening test dataset, and CCT and FUT parameters. Blood samples were collected and PRA, PAC, and ARR were measured before and 60 or 90 min after administering 50 mg of captopril in the CCT, and before and at 120 min in an upright posture after an intravenous bolus injection of 40 mg furosemide in the FUT.

***1.1 Physical examinations***

・The BMI was calculated as weight (kg) divided by height (m) squared.

・The BP was measured by employing an established procedure using a sphygmomanometer with an appropriately sized cuff positioned at the level of the heart (2–4).

***1.2 Biochemistry method***

・The fasting blood sample was obtained early in the morning for the measurement of levels of fasting blood glucose, serum cholesterol, and the other biochemistry parameters, including electrolytes and creatinine levels, which were determined by employing standard laboratory procedures, as described previously (5).

・The glomerular filtration rate was measured using an abbreviated equation, which included serum creatinine (Cre), age, and sex for easy computation: eGFR (mL×min−1×1.73 m−2)=186×(Cre)−1.094×(age)−0.287×(0.742 for women) (6).

・Proteinuria was semi-quantitatively diagnosed across 5 levels: “negative”, “trace”, “1+”, “2+” and “3+”.

***1.3 Medical history***

・Stroke included a diagnosis of cerebral infarction, cerebral hemorrhage, and subarachnoid hemorrhage, confirmed by neurologists.

・Ischemic heart disease (IHD) included myocardial infarction (MI) and angina pectoris, which were confirmed by cardiologists; atrial fibrillation was diagnosed based on an electrocardiogram.

・The diagnostic criteria for diabetes were in accordance with the Japan Diabetes Guidelines (7). The diagnostic criteria for dyslipidemia were in accordance with the Japan Atherosclerosis Society Guidelines (8).

・Hyperuricemia, sleep apnea syndrome, depression, periodic paralysis, and pregnancy-induced hypertension data were registered in the JRAS database based on the patient’s self-reported declarations.

***1.4 Treatment information***

・The antihypertensive medications were converted into standardized DDD according to the World Health Organization Anatomical Therapeutic Chemical /DDD Index 2010 (1).

・Daily amounts of potassium (mEq/day) were standardized as daily dose of potassium L-Aspartate; 411 patients were treated for hypokalemia using potassium L-Aspartate (n = 43), [potassium](https://www.weblio.jp/content/Potassium) [glutamate](https://www.weblio.jp/content/glutamate) (n = 5), and potassium chloride (n = 367) among the 1,897 examined patients. We calculated amounts of potassium contained in potassium L-aspartate as being twice that of potassium chloride based on the commonly used dose because of the difference in potassium internalization between potassium L-aspartate and potassium chloride.

***1.5 Assay methods of plasma aldosterone concentration and plasma renin activity***

Plasma aldosterone concentrations were measured using commercially available radioimmunoassays (SPAC-S Aldosterone kits, Fuji Rebio, Co., Ltd, Tokyo, Japan) at 38 centers or Chemiluminescent Enzyme Immunoassay (CLEIA; Accuraseed Aldosterone, FUJIFILM Wako Pure Chemical, Co., Tokyo, Japan) at 3 centers. The PRA was measured using a radioimmunoassay or enzyme immunoassay (EIA). The reference range for PRA with patients in the supine position was 0.3–2.9 ng/mL/hr (PRA-FR RIA kits, Fuji Rebio, Co., Ltd, Tokyo, Japan) at 24 centers, 0.2–2.3 ng/mL/hr (PRA EIA kits, Yamasa, Co., Ltd, Choshi, Japan) at 13 centers, and 0.2–2.7 ng/mL/hr (PRA RIA kits, Yamasa, Co., Ltd, Choshi, 122 Japan) at 3 centers.

1. ***Process to develop the machine-learning algorithms***

***2.1 Data preparation***

For supervised learning, the output variable adopted the PA laterality subtypes: APA as “1” (favorable) and BAH as “0” (unfavorable). Numerical variables were treated as continuous variables. Categorical variables were employed as binary variables, except urinary protein, which comprised five levels.

Input variables included the 46 and 59 variables of the screening and confirmatory test datasets, respectively. The entire process is described in Figure 1. First, manual variable extraction was performed as described in the Dataset and variables subsection. Subsequently, the database was divided into training (70%), validation (10%), and test (20%) datasets, stratified according to the proportion of PA laterality subtypes. The training dataset was used to train prediction models, the validation dataset was used for parameter tuning, and the test dataset was used to measure generalization performance. Lastly, the constructed models were evaluated using area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and positive and negative predictive values. Specificity and sensitivity were evaluated using Youden’s index and the AUC. Youden’s index is the point existing at the maximum vertical distance of the receiver operating characteristic curve from the diagonal line and can be calculated by combining the sensitivity and specificity and subtracting 1 (9). The entire process was repeated 50 times, and each model’s performance was evaluated according to the mean value of the replicates.

***2.2 General process of model construction and evaluation***

This study adopted the ensemble learning model (ELM) and employed seven machine learning algorithms (MLA): Random Forest (RF), Multilayer Perceptron (MLP), Light Gradient Boosting Machine (LightGBM), Support Vector Machine (SVM), Logistic Regression (LR), k-nearest neighbor algorithm (KNN), and Naive Bayes (NB). The ELM was constructed using multiple different types of MLA and combined the prediction of each model (10–12). The integrated output of the ELM was calculated by averaging the class probability of individual MLAs. The entire ELM was built and compared following five steps. In the first step of model construction, the missing values in each dataset were imputed using MissForest (13), based on training data. Subsequently, we employed feature selection based on feature importance ranking in RF and identified variables using the best-performing prediction models. Each machine-learning model was trained with a feature subset consisting of 1–46 and 1–59 variables in the screening and confirmatory test datasets, respectively. Additionally, a third dataset was developed using the five most important variables according to the importance ranking on RF.

In the third step, Synthetic Minority Oversampling Technique (SMOTE) (14) was used for oversampling until the minority class ratio was equal to the majority class ratio on training data.

All models had hyperparameters that contributed to their performance. The best parameters of the model learned with the training dataset were searched using validation data. In the fourth step, six different algorithms except for NB were optimized for the following parameters in accordance with methods that were previously reported (15–21).

・RF: the number and the depth of trees.

・MLP: the number of hidden layers and neurons, the penalty parameter.

・LightGBM: the number and the depth of trees, the regularization strength.

・SVM: the regularization strength and a coefficient for radial basis function (RBF) kernel.

・LR: the regularization strength.

・KNN: the number of nearest neighbors.

All other hyperparameters were set at the default values in Scikit-learn 0.22.1.

In the final step, the prediction models underwent external validation employing a test dataset that had not been utilized in previous steps. These evaluated models were refit by employing the training dataset, the tuned hyperparameters, and the optimized number of features. The ELM’s prediction for the test dataset was calculated by combining the class probability output by the seven MLAs.

Additionally, the generalization performance of each model was evaluated in this step with the measurements of AUC, sensitivity, specificity, and positive and negative predictive values.

***2.3 Imputation of missing values using MissForest***

We used MissForest to impute missing value in six algorithms (13). MissForest creates a random forest model for each variable with missing values, using the rest of the variables in the dataset. MissForest then predicts the missing values for that variable. The process is repeated until a terminating criterion is met or the set number of attempts is reached.

***2.4 Oversampling using SMOTE***

We used SMOTE, which is an oversampling technique, to increase the minority class to the same number as the majority class to manage handling of imbalanced data (14). SMOTE generates synthetic data using interpolation with the data detected using the k-neighbor method. The model is constructed in the following steps:

1. Choose a minority sample () and one of k minority neighborhood of as



1. Using random number ranged from 0 to 1(gap), difference in attribute values of and (diff)



Synthetic data () is



1. Repeat until the same number as the majority class

***2.5 Machine-learning algorithms***

We adopted seven machine-learning algorithms: RF, MLP, LightGBM, SVM, LR, KNN, and NB. These algorithms can be described briefly as follows. RF22 is a kind of ensemble learning that creates multiple models (called weak learners) and combines them to make a decision. RF comprises decision trees constructed with a different subset of training data, where each tree is slightly different from the others. It can calculate feature importance based on the impurity. MLP23 is the principal artificial neural network algorithm. MLP consists of an input layer, a hidden layer, and an output layer with non-linear activation function, and it learns by updating the weights until the error between the output model data and the teacher data is reduced. The advantage of this technique is that more complex network architectures are used to solve more complex problems. LightGBM24 is a model based on the Gradient Boosting Decision Tree and a kind of ensemble learning. LightGBM can build the tree sequentially, which makes it an effective method for processing large-scale data and features that can handle missing values as they are. SVM25 constructs a separating decision boundary, called hyperplane, and discriminates by maximizing the margin between classes. The algorithm can map the input feature space to a multi-dimensional space using kernel functions. LR26 is a traditional method for binary classification, which is the extension of a linear model. The model is explainable as the effect of the coefficient of the dependent variable on the outcome. KNN27 is non-parametric algorithm which performs a classification based on distance metric. The prediction for the test data is assigned by a majority vote of *k* samples that are close in distance to the training data. NB28, 29 is a probabilistic model based on the Bayes’ theorem with the strong assumption of conditional independence between input features. An NB classifier is easy to build because it does not require iterative parameter estimation.

***3. REFERENCES***

1. WHO Collaborating Centre for Drug Statistics Methodology: ATC/DDD Index 2020. Oslo, Norway: Norwegian Institute of Public Health (<https://www.whocc.no/atc_ddd_index/>).
2. Umemura S, Arima H, Arima S, et al. [The Japanese Society of Hypertension Guidelines for the Management of Hypertension (JSH 2019).](https://pubmed.ncbi.nlm.nih.gov/31375757/) Hypertens Res. 2019;42:1235-1481.

# [Shimamoto](https://pubmed.ncbi.nlm.nih.gov/?term=Shimamoto+K&cauthor_id=24705419) K, [Ando](https://pubmed.ncbi.nlm.nih.gov/?term=Ando+K&cauthor_id=24705419) K, [Fujita](https://pubmed.ncbi.nlm.nih.gov/?term=Fujita+T&cauthor_id=24705419) T, et al. The Japanese Society of Hypertension Guidelines for the Management of Hypertension (JSH 2014). Hypertens Res. 2014;37:253-390.

1. [Ogihara](https://pubmed.ncbi.nlm.nih.gov/?term=Ogihara+T&cauthor_id=19300436) T, [Kikuchi](https://pubmed.ncbi.nlm.nih.gov/?term=Kikuchi+K&cauthor_id=19300436) K, [Matsuoka](https://pubmed.ncbi.nlm.nih.gov/?term=Matsuoka+H&cauthor_id=19300436) H, et al. The Japanese Society of Hypertension Guidelines for the Management of Hypertension (JSH 2009). Hypertens Res. 2009;32:3-107.
2. Ohno Y, Sone M, Inagaki N, et al. Prevalence of cardiovascular disease and its risk factors in primary aldosteronism: a multicenter study in Japan. Hypertension. 2018;71:530-537.
3. Matsuo S, Imai E, Horio M, et al. [Revised equations for estimated GFR from serum creatinine in Japan.](https://pubmed.ncbi.nlm.nih.gov/19339088/) Am J Kidney Dis. 2009;53:982-992.
4. Seino Y, Nanjo K, Tajima N, et al. Report of the committee on the classification and diagnostic criteria of diabetes mellitus. J Diabetes Investig. 2010;1:212-228.
5. Teramoto T, Sasaki J, Ishibashi S, et al. Diagnostic criteria for dyslipidemia. J Atheroscler Thromb. 2013;20:655-660.
6. [Hilden](https://pubmed.ncbi.nlm.nih.gov/?term=Hilden+J&cauthor_id=8783436) J, [Glasziou](https://pubmed.ncbi.nlm.nih.gov/?term=Glasziou+P&cauthor_id=8783436) P. Regret graphs, diagnostic uncertainty and Youden's Index. Stat Med. 1996;15:969-986.
7. Alghamdi M, Al-Mallah M, Keteyian S, et al. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford ExercIse Testing (FIT) project. PLoS One. 2017;12:e0179805.
8. Lin E, Lin CH, Lane HY. Prediction of functional outcomes of schizophrenia with genetic biomarkers using a bagging ensemble machine learning method with feature selection. Sci Rep. 2021;11:1-8.
9. Park DJ, Park MW, Lee H, Kim YJ, Kim Y, Park YH. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep. 2021;11:7567.
10. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112-118.
11. Chawla N, Bowyer K, Hall L, Kegelmeyer W.: SMOTE: Synthetic Minority Over-Sampling Technique. J Artif Intell Res. 2002;16:321-357.
12. Czarnecki WM, Podlewska S, Bojarski AJ. Robust optimization of SVM hyperparameters in the classification of bioactive compounds. J Cheminform. 2015;7;38.
13. Lee HC, Yoon SB, Yang SM et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. J. Clin. Med. 2018;7:428.
14. Wong J, Manderson T, Abrahamowicz M, Buckeridge DL, Tamblyn R. Can Hyperparameter Tuning Improve the Performance of a Super Learner?: A Case Study. Epidemiology. 2019;30:521-531.
15. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:2623-2631.
16. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. WIREs Data Mining Knowl Discov. 2019;9:e1301.
17. Zhang J, Mucs D, Norinder U, Svensson, F. LightGBM: an effective and scalable algorithm for prediction of chemical toxicity - application to the Tox21 and mutagenicity datasets. J Chem Inf Model. 2019;59:4150-4158.
18. F Khan, S Kanwal, S Alamri et al. Hyper-Parameter Optimization of Classifiers, Using an Artificial Immune Network and Its Application to Software Bug Prediction. IEEE Access. 2020;8:20954-20964.
19. Breiman L. Random forests. Machine Learning. 2001;45:5-32.
20. Rumelhart DE. Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533-536.
21. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree, in: 31st Conference Neural Information Processing Systems NIPS, 2017;3146-3154.
22. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20:273-297.
23. DW Hosmer, S Lemeshow, RX Sturdivant. Applied logistic regression 3rd edition. 2013; John Wiley & Sons, Inc.
24. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory. 1967;13:21-27.
25. Langley P, Iba W, Thompson K. An analysis of Bayesian classifiers. Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI). 1992;223-228.
26. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Machine Learning. 1997;29:131-163.