

Health improvement framework
for actionable treatment planning
using a surrogate Bayesian model

階層ベイズモデルを利用した
実行可能な健康改善プランを
提案する AI 技術の開発

中村 和貴

ARTICLE



<https://doi.org/10.1038/s41467-021-23319-1>

OPEN

Health improvement framework for actionable treatment planning using a surrogate Bayesian model

Kazuki Nakamura ^{1,2}, Ryosuke Kojima ², Eiichiro Uchino ², Koh Ono ³, Motoko Yanagita ^{4,5}, Koichi Murashita⁶, Ken Itoh ⁷, Shigeyuki Nakaji ⁸ & Yasushi Okuno ^{2✉}

Clinical decision-making regarding treatments based on personal characteristics leads to effective health improvements. Machine learning (ML) has been the primary concern of diagnosis support according to comprehensive patient information. A prominent issue is the development of objective treatment processes in clinical situations. This study proposes a framework to plan treatment processes in a data-driven manner. A key point of the framework is the evaluation of the actionability for personal health improvements by using a surrogate Bayesian model in addition to a high-performance nonlinear ML model. We first evaluate the framework from the viewpoint of its methodology using a synthetic dataset. Subsequently, the framework is applied to an actual health checkup dataset comprising data from 3132 participants, to lower systolic blood pressure and risk of chronic kidney disease at the individual level. We confirm that the computed treatment processes are actionable and consistent with clinical knowledge for improving these values. We also show that the improvement processes presented by the framework can be clinically informative. These results demonstrate that our framework can contribute toward decision-making in the medical field, providing clinicians with deeper insights.

¹ Research and Business Development Department, Kyowa Hakko Bio Co., Ltd., Tokyo, Japan. ² Department of Biomedical Data Intelligence, Graduate School of Medicine, Kyoto University, Kyoto, Japan. ³ Department of Cardiovascular Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. ⁴ Department of Nephrology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. ⁵ Institute for the Advanced Study of Human Biology, Kyoto University, Kyoto, Japan. ⁶ Center of Innovation Research Initiatives Organization, Hirosaki University, Hirosaki, Japan. ⁷ Department of Stress Response Science, Hirosaki University Graduate School of Medicine, Hirosaki, Japan. ⁸ Department of Social Health, Hirosaki University Graduate School of Medicine, Hirosaki, Japan. ✉email: okuno.yasushi.4c@kyoto-u.ac.jp

Medically appropriate and patient-acceptable decision making is beneficial in enhancing the quality of care^{1–3}. Considerable evidence has been accumulated from many studies on the assessment of health statuses and risk profiles. Accordingly, standardized care has been provided in the form of guidelines. Conversely, the uniform application of standardized care is undesirable in real clinical situations whereby diversity exists in relation to individual preferences, feasibility, and acceptability⁴. Personalized medicine, also known as precision medicine, has become popular and presents new opportunities in clinical situations in recent years^{5–7}. In addition to the understanding of individual, unique health-related factors, their consideration has been essential for shared decision making between patients and clinicians. Realistic and appropriate health improvement plans for the patient's health conditions have been regarded as a crucial component of shared decision making^{8,9}. However, feasible planning largely depends on the empirical judgment of the clinicians in clinical situations, and decision making based on objective health improvement plans for both patients and clinicians is difficult. Major challenges remain in providing clinicians with tools to support clinical decision making in a data-driven manner^{10,11}.

Machine learning (ML) technology has been extensively used in the medical field, especially for diagnosis support and disease prediction based on comprehensive patient information^{12–15}. Unlike interpretable techniques, such as classical statistical analysis and linear models, the black-box nature of highly predictive ML models, including ensemble learning and deep learning, is often a barrier to clinical decision-making applications¹⁶. Explainable artificial intelligence (XAI) has been receiving increasing attention recently in the field of ML¹⁷. XAI is a research field on techniques that explain black-box ML predictions, and it has been applied to medical ML models where interpretability is often required^{18,19}. A successful application of XAI in the medical field is the identification of individual health-related factors that contribute to disease prediction using local interpretable model-agnostic explanations (LIME) and Shapley additive explanations (SHAP)^{20–25}. These methods achieve both predictive performance and individual interpretation by using an additional individual model referred to as the surrogate model. The remaining important issue in clinical decision making is the development of personalized treatment plans for rational treatment²⁶. However, these conventional methods merely explain the prediction reasons but cannot provide effective treatment processes. For example, for the prediction of hypertension, it is unclear what type of actions will improve effectively the individual blood pressure in a set of candidate actions related to blood test data and body composition despite the fact that we can understand important features in the prediction.

This study proposes a framework for planning an actionable path for personalized treatment based on the predictions of an ML model. A key idea of our framework is to use a hierarchical Bayesian model as a surrogate model of a specified ML model. We refer to this surrogate model as the stochastic surrogate model. Unlike the conventional surrogate model, the stochastic surrogate model based on the hierarchical Bayesian model enables the calculation of the probability of being the state of a given variable set. Our framework can evaluate the actionability of the treatment processes that was not considered by conventional methods^{20,21}, by computing a path probability with the use of the stochastic surrogate model. The combined use of the ML model and the stochastic surrogate model achieves both a high-prediction performance and actionability evaluation. The simultaneous computation of ML model prediction and actionability evaluation yields an actionable treatment process that leads to clinical applications to improve personal health.

This study also presents experiments conducted to address two different aspects. First, we evaluated our proposed framework from the viewpoint of the proposed methodology using a synthetic dataset. Subsequently, we assessed our framework on a clinical application using an actual health checkup dataset. In this experiment, we calculated personal health improvement paths and confirmed the consistency with clinical knowledge based on the assumption of two kinds of scenarios, wherein we aimed to lower blood pressure in individuals with systolic blood pressure (SBP) ≥ 140 mmHg and intervene individuals with a risk of chronic kidney disease (CKD), as examples for regression and classification problems, respectively. Also, we performed clinician assessments of the improvement process presented by the framework. To the best of our knowledge, this is the first study that made it possible to present effective improvement paths based on the ML model using an actual health checkup dataset.

Results

Path planning framework using surrogate Bayesian model.

This section describes the proposed framework (Fig. 1, Supplementary Fig. 1). Our framework consisted of three steps. Step 1: we built a prediction model using ML methods. Step 2: the surrogate model of the prediction model was constructed using hierarchical Bayesian modeling. Step 3: path planning was conducted to identify an actionable path for the treatment performed using the surrogate model. The actionability of the path was defined as the product of probabilities of taking a series of variable states on a specified path (detailed in “Methods”). Thus, the combined framework of the prediction model and the stochastic surrogate model achieves both high prediction performance and actionability evaluation, which leads to an actionable treatment process for personal health improvement.

In Step 1, a prediction model was built from a dataset expressed in the form of a table format, wherein columns consist of multiple explanatory variables and a response variable, and rows represent instances (Fig. 1a). The output of the constructed model would correspond to a clinician's assessment of individual health status or future predictions. For example, a regression model for blood pressure could be used to estimate the value of the response variable, i.e., blood pressure, from the value of explanatory variables, such as the body composition and blood test data. In our framework, we can use arbitrary ML algorithms, such as high-performance nonlinear algorithms, to construct the prediction model.

Based on the original explanatory variable values and the predicted values of the prediction model in Step 1, a stochastic surrogate model was constructed in Step 2 using hierarchical Bayesian modeling (Figs. 1b and 2). Items that cannot be easily measured in clinical situations, or future values, are available as the response variables given that the predicted values of the prediction model were applied in our framework. Clinically, this stochastic surrogate model represents a set of realistic health conditions for patients. The model was used to compute the probability of given values of explanatory and response variables to evaluate actionability in the next step. Note that this stochastic surrogate model represents the probability density for all possible states of variables.

In Step 3, a health-improvement treatment path was calculated for each instance using the stochastic surrogate model (Figs. 1c, d, and 3, as detailed in “Methods”), and would be equivalent to the consideration of an appropriate health improvement plan for individual patients in clinical situations. The explanatory variables of an instance, such as body composition and blood test data, were hypothetically changed to improve the response variable predicted using the prediction model. By using the

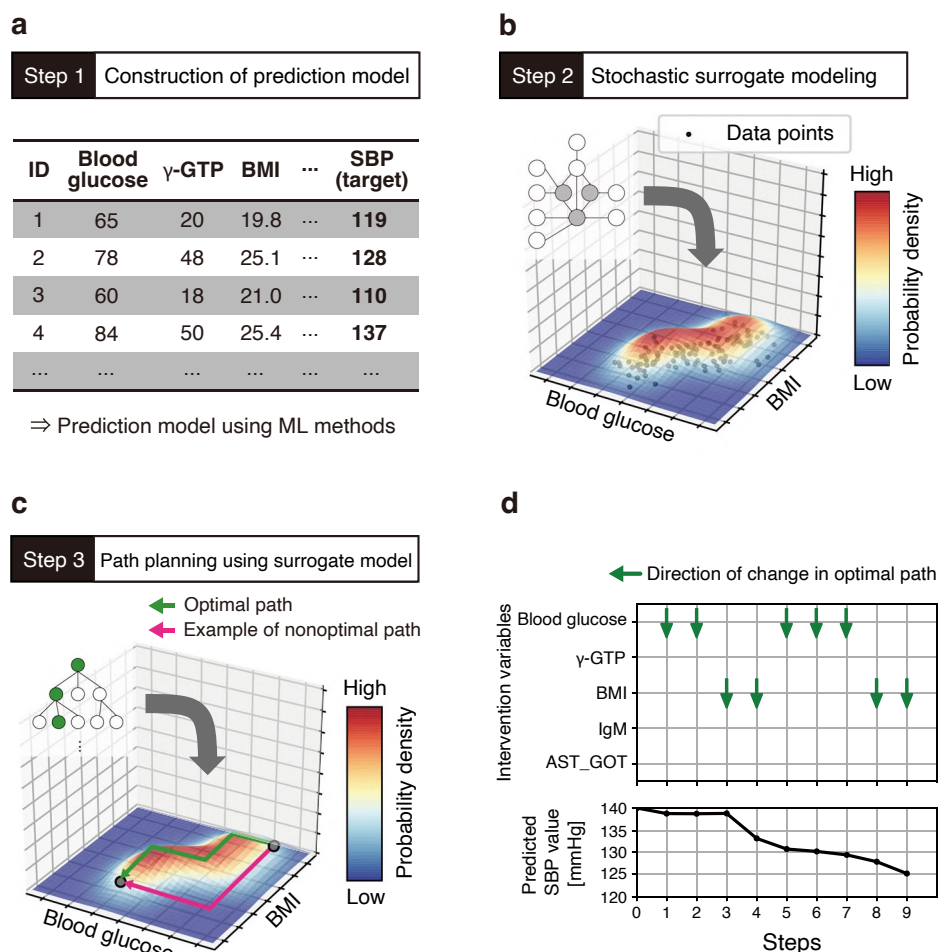


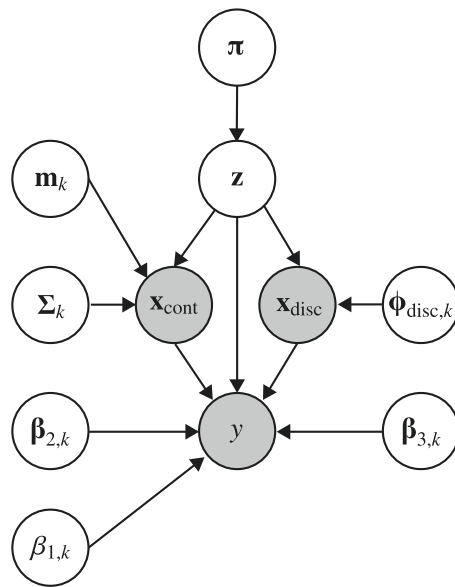
Fig. 1 Schematic representation of the framework for planning actionable paths for treatment using hierarchical Bayesian modeling. The framework consists of three steps. A schematic is given as an example in which a path is planned to improve the systolic blood pressure (SBP) owing to changes in blood data and body composition data. **a** Construction of a prediction model from the dataset. A variable SBP is set as the response variable in this case. **b** Construction of a stochastic surrogate model based on the original dataset and the predicted values of the prediction model. This figure shows a schematic representation of a two-variable space regarding blood glucose and body mass index (BMI). The heatmap and vertical axis represent the existence probability of data in the variable space, which is expressed by the stochastic surrogate model. **c, d** Actionable path planning is applied to improve the response variable. The path is represented as a set of multistep transitions on explanatory variables. In our framework, the optimal path (green line in **c**) is planned on the grid graph with high probabilities in the variable space based on the stochastic surrogate model. Conversely, the nonoptimal path (red line in **c**) could pass through nodes with low probability. γ -GTP gamma glutamyl transferase, IgM immunoglobulin M, AST_GOT aspartate transaminase.

stochastic surrogate model constructed in Step 2, we could calculate how easy it was to take the state of the combination of variables. The framework output the most actionable (optimal) path for a state that improved the prediction value, which was a sequence of the changed values with high probability in the surrogate model for the given ML model (detailed in “Methods”). Therefore, this model can avoid the nonoptimal health-improving paths (shown by the red line in Fig. 1c) where intermediate situations could be unrealistic.

Considering real applications, explanatory variables contain variables that cannot be changed by interventions such as age. The appropriate subset of variables for the application should be determined. The explanatory variables for setting the virtually changed values were called intervention variables in this study. In our path planning setting, the intervention variables were regarded as a grid graph, and a path was defined by connecting the grid points (nodes). We defined a probability of the node as the probability of taking the node calculated using the surrogate model. Based on a breadth-first search²⁷, we obtained the optimal

path to the destination node that achieves the most improved predictive value within the search iteration count L (detailed in “Methods”).

Validation of framework on synthetic dataset. We evaluated our framework with a synthetic dataset to confirm that improvement paths with high actionability could be planned. This synthetic dataset was generated from three, three-dimensional (3D) normal distributions (Supplementary Fig. 2). We built a regression model and subsequently constructed a stochastic surrogate model using hierarchical Bayesian modeling (Supplementary Fig. 3a, b). The lowest widely applicable Bayesian information criterion (WBIC) value²⁸ was obtained when the number of mixture components in the hierarchical Bayesian model was equal to two. Details of the setting related to this model are given in “Methods”. Subsequently, we planned paths to decrease the value of the response variable using this stochastic surrogate model. All explanatory variables were selected for intervention variables, and planned paths were more actionable than the baseline path



$$k \sim \text{Categorical}(\pi)$$

$$\mathbf{x}_{\text{cont}} \sim N(\mathbf{m}_k, \Sigma_k)$$

$$\mathbf{x}_{\text{disc}} \sim \text{Categorical}(\phi_{\text{disc},k})$$

Regression

$$y \sim N(\mu_k, \sigma)$$

$$\mu_k = \beta_{1,k} + \beta_{2,k}^T \mathbf{x}_{\text{cont}} + \beta_{3,k}^T \mathbf{x}_{\text{disc}}$$

$$\sigma = \frac{\text{RMSE}_{\text{test}}}{2}$$

$$\beta_{1,k} \sim N(y_{\text{mean}}, 5y_{\text{std}})$$

Classification

$$y \sim \text{Bernoulli}(\theta_k)$$

$$\theta_k = \frac{1}{1 + e^{-\mu_k}}$$

$$\mu_k = \beta_{1,k} + \beta_{2,k}^T \mathbf{x}_{\text{cont}} + \beta_{3,k}^T \mathbf{x}_{\text{disc}}$$

$$\beta_{1,k} \sim N(0, 5)$$

$$\beta_{2,k} \sim \text{DoubleExponential}(0, 1)$$

$$\beta_{3,k} \sim \text{DoubleExponential}(0, 1)$$

$$\pi \sim \text{Dirichlet}(1)$$

$$\mathbf{m}_k \sim N(0, 5\mathbf{I})$$

$$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 2.5))$$

$$\phi_{\text{disc},k} \sim \text{Dirichlet}(1)$$

Fig. 2 Graphical model representation of stochastic surrogate model. Nodes in the graphical model are represented as follows: $\mathbf{x}_{\text{cont}}$, continuous explanatory variables; $\mathbf{x}_{\text{disc}}$, discrete explanatory variables; y , response variable predicted by the ML model; $\mathbf{z}$, the parameter of the mixture components; and all the others, prior distributions. The formulation for y differs between regression and classification tasks. k represents each mixture component, and Σ_k is a diagonal matrix with elements according to the Cauchy distribution. The symbol $\text{RMSE}_{\text{test}}$ in the equation represents a root-mean-squared error of the regression model, and y_{mean} and y_{std} represent the mean and standard deviation values of the predicted response variable, respectively.

Algorithm 1 Path search algorithm

function PATH_SEARCH

```

1: initial_node ← data of instance
2: initial_node.cost ← 0
3: All other node.cost ← ∞
4: initial_node.visited ← False
5: All other node.visited ← False
6: current_node ← initial_node
7: for  $i = 1$  to  $L$  do
8:   neighbors ← list of node adjacent to current_node in grid graph
9:   for neighbor_node in neighbors do
10:     $\text{cost} \leftarrow \text{current\_node.cost} + \text{neighbor\_node.neg\_Log\_prob}$ 
      (Negative logarithm of node probability)
11:    if neighbor_node.cost > cost then
12:      neighbor_node.cost ← cost
13:    end if
14:  end for
15:  current_node.visited ← True
16:  current_node ← node which is not visited and has smallest cost
17: end for
18: destination_node ← node with the best response variable
19: return destination_node

```

Fig. 3 The intervention variable space was regarded as a grid graph, and the grid points (nodes) were connected to plan a path. The nodal probability was calculated using the surrogate model. The actionability was defined as the product of nodal probabilities on a specified path. The most actionable (optimal) path for each node was calculated, and the output path was the optimal path to the node with the most improved predictive value within the search iteration count, L . Pseudocode of path search algorithm.

(Supplementary Fig. 3c). This baseline method for the baseline path is defined in “Methods”. The planned paths of two randomly selected instances are shown in Fig. 4. We successfully demonstrated that our framework could plan paths to improve response variable values with high probabilities in our framework. Another

experiment on a more complicated five-dimensional dataset also demonstrated successful path planning (Supplementary Notes). These results showed that the planned paths with high probabilities were discovered rather than the naïve straight path that connected the initial node to the destination node.

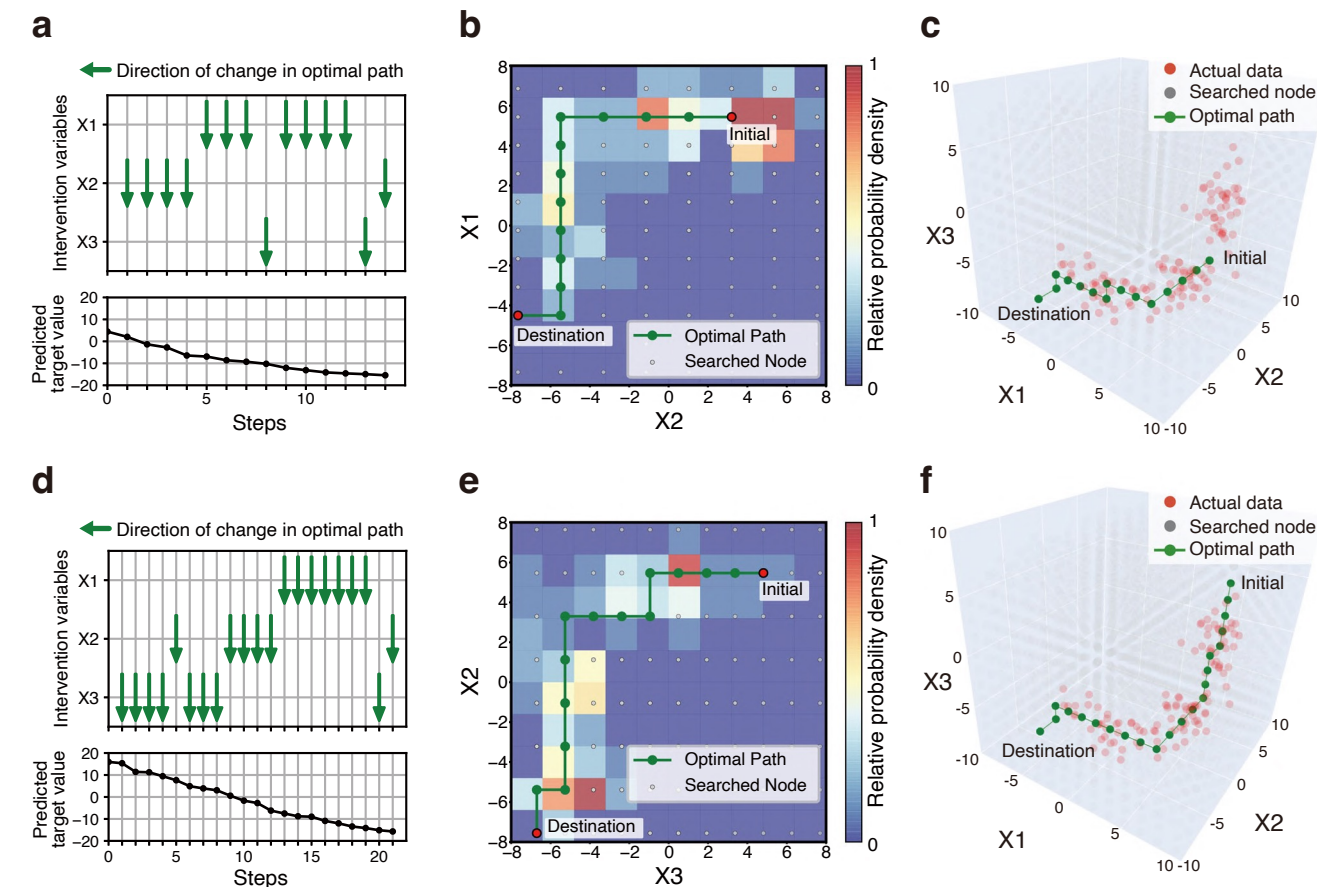


Fig. 4 Examples of actionable paths planned on synthetic dataset. The optimal paths for improving the response variable predicted by the ML model are represented for randomly selected two examples: instance A (**a–c**) and instance B (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, e** Two-dimensional (2D) plots of the path. The 2D plots are shown regarding the selected two variables: X1 and X2 (**b**), and X2 and X3 (**e**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** Three-dimensional (3D) plots of the path.

Table 1 Subject characteristics during first-time participation.	
Participant characteristics	(n = 3132)
Age (years)	51.3 ± 16.0
BMI (kg/m ²)	23.0 ± 3.5
SBP (mmHg)	125.9 ± 19.1
DBP (mmHg)	74.8 ± 11.9
eGFR (mL/min/1.73 m ²)	82.1 ± 16.5
Sex	
Male	1234 (39.4%)
Female	1898 (60.6%)
History of hypertension	
No history	2446 (78.1%)
Undertreatment	650 (20.8%)
Past history	36 (1.1%)

BMI body mass index, SBP systolic blood pressure, DBP diastolic blood pressure, eGFR estimated glomerular filtration rate.

Application on actual health checkup dataset. We used the Iwaki health promotion project (IHPP) dataset, an actual health checkup dataset, to demonstrate that the proposed framework can plan actionable paths for treatment. The IHPP has acquired (on an annual basis) a wide range of health checkup data that describe the molecular biology, physiology, biochemistry, lifestyle,

and socio-environment of participants. We considered two kinds of scenarios to lower SBP and risks for CKD, whereby planned paths could be interpreted from a clinical perspective. CKD risk was defined as estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m². Table 1 shows an overview of the IHPP dataset. Because the dataset comprised more than 2000 measurement items, we reduced the explanatory variables before we built prediction models in our framework. We excluded diastolic blood pressure (DBP) and limb blood pressures from the explanatory variables for SBP prediction and creatinine for CKD risk prediction. Our framework assumes the intervention in explanatory variables to improve response variables. Hence, actions such as intervening in blood pressure (such as DBP) to lower SBP are not reasonable. Furthermore, ambiguous items, such as answers to questionnaires and items with ≥25% missing values, were excluded. Subsequently, XGBoost-based²⁹ recursive feature elimination (RFE)³⁰ was performed to reduce the explanatory variables with the training data (Supplementary Data 1 and Supplementary Table 1). We applied one-hot encoding for categorical variables and replaced the missing values with the median of the training data for simplicity. RFE was performed with five-fold cross-validation split by participants, and the explanatory variables were reduced to 25, which had little impact on predictive scores on validation data (Supplementary Fig. 4). In the following sections, applications of our framework on the regression task for SBP and classification task for CKD risk are described.

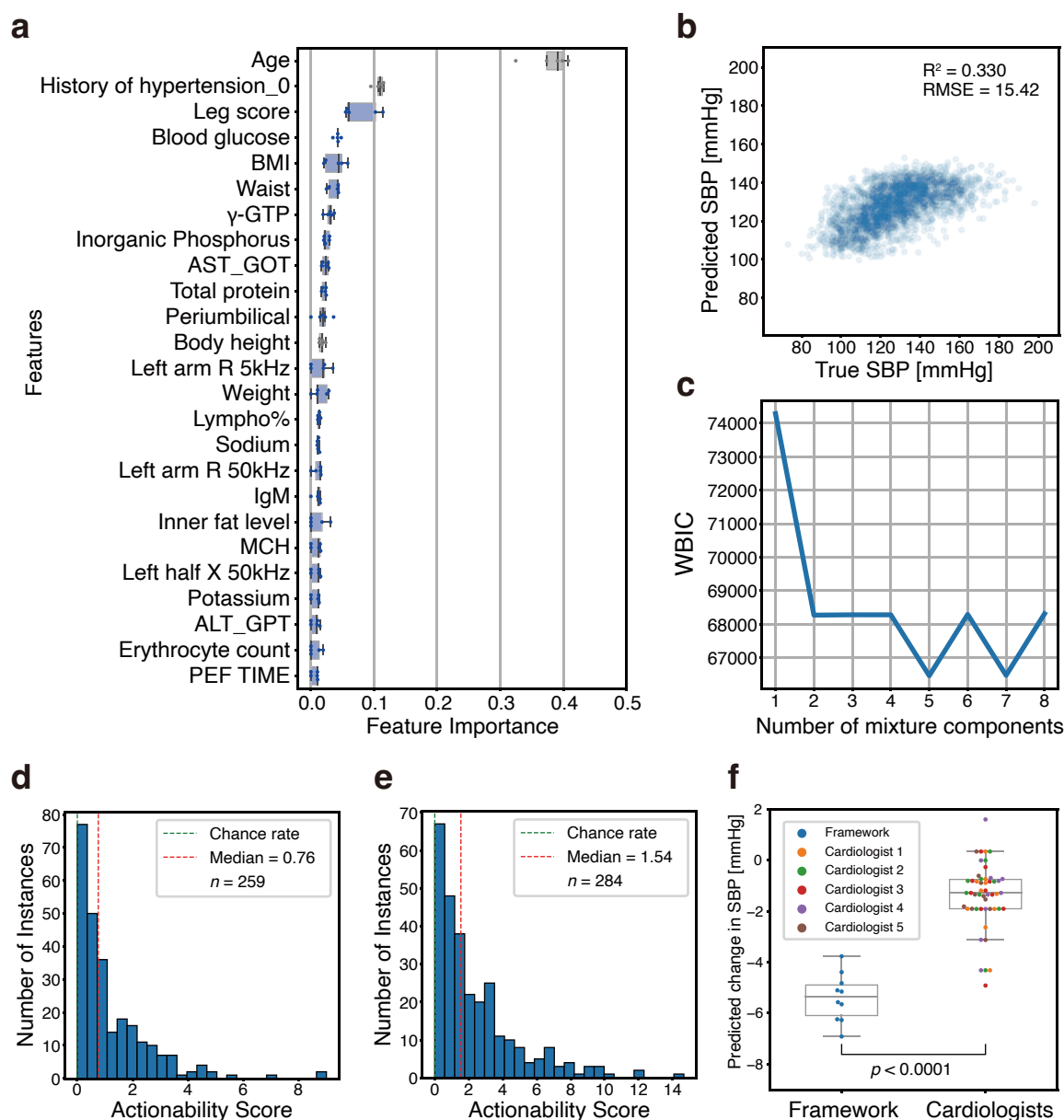
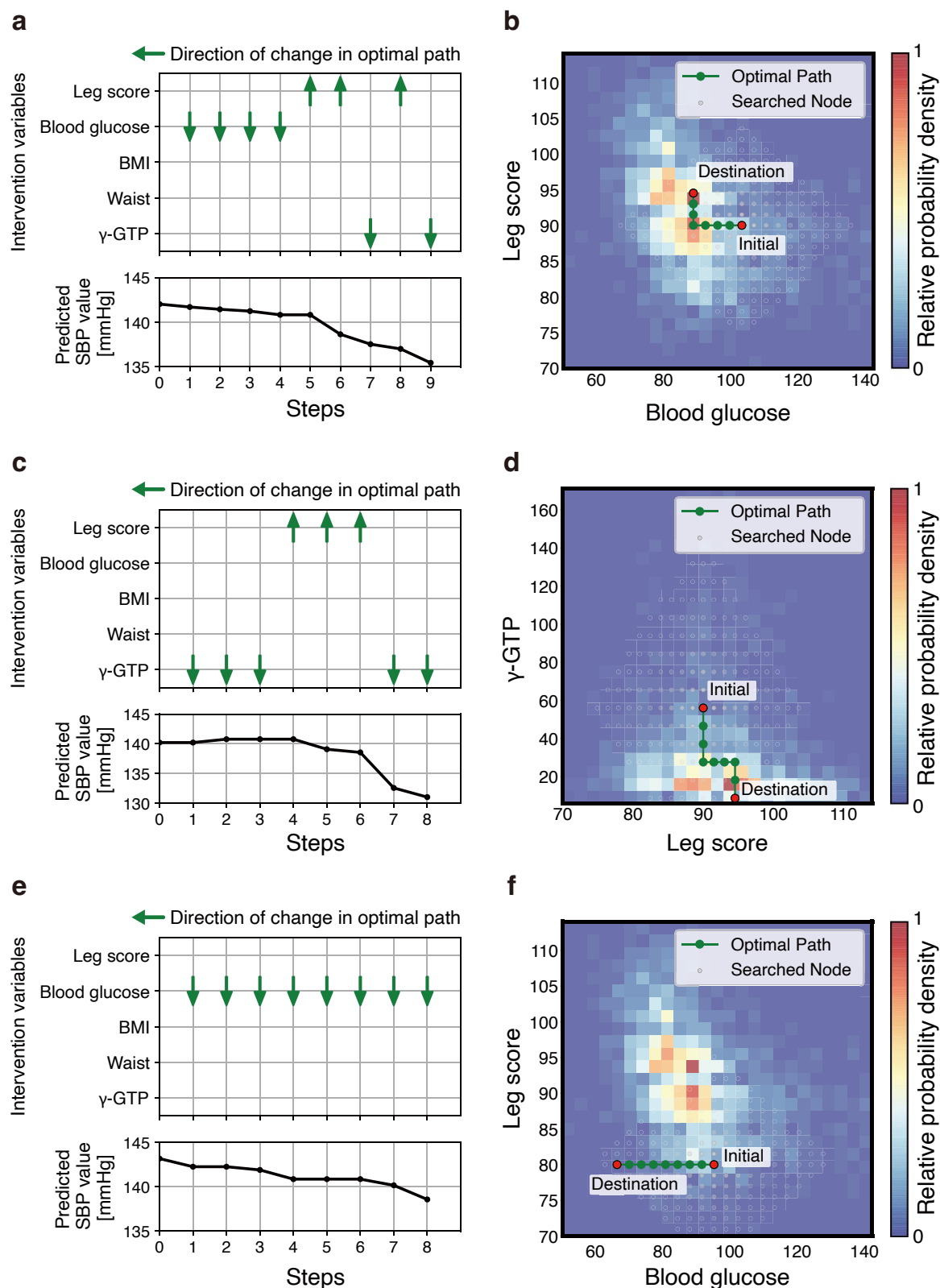


Fig. 5 Application of proposed framework on systolic blood pressure (SBP) regression task using the Iwaki health promotion project (IHPP) dataset.

a Feature importance: these 25 features were selected by recursive feature elimination (RFE) to predict the SBP. RFE was performed with fivefold cross-validation, and the feature importance when 25 variables remained is shown for each fold ($n = 5$). The plot color represents the following: gray: variables which cannot be intervened, and blue: intervenable variables. Details of features are described in Supplementary Data 1. **b** Plot for prediction vs. true response variable. **c** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **d, e** Histogram of actionability scores with intervention variables based on data-driven selection (**d**) or hypothesis-driven selection (**e**) at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path. **f** Comparison of predicted SBP reduction between framework-proposed paths and cardiologist-selected paths. Health improvement paths constructed based on hypothesis-driven intervention variables using our framework were compared with the cardiologist-selected paths among framework-proposed paths and random paths. Each cardiologist evaluated the same randomly selected instances ($n = 10$). Statistical significance was calculated using the Welch's *t*-test (two-sided). In box-plot, center line represents median; box limits, upper and lower quartiles; whiskers, $1.5 \times$ interquartile range.

Application of framework on SBP regression task. We applied our framework to the SBP regression task. Important features selected by RFE comprised items related to hypertension, such as age, body composition (leg score, body mass index [BMI], and waist), blood glucose, gamma glutamyl transferase (γ -GTP), and serum sodium^{31–40} (Fig. 5a, the details of the variables are described in Supplementary Data 1). Therefore, the selected explanatory variables were considered to be reasonable for SBP predictive models from the clinical perspective.

Following our framework, a regression model was built after the replacement of missing values with the use of multiple imputations. This is a more precise imputation method for missing values. To estimate multiple imputations, Bayesian ridge and random forest⁴¹ were used for continuous and discrete variables, respectively. Consequently, the regression model yielded a root-mean-squared error (RMSE) equal to 15.42 and an *R*-squared value equal to 0.330 (Fig. 5b). Subsequently, hierarchical Bayesian modeling was performed to construct the



stochastic surrogate model based on multiple imputed RFE-selected features and predicted SBP. The lowest WBIC value was obtained when the number of mixture components was five (Fig. 5c). Path planning was performed with this stochastic surrogate model. Regarding the intervention variables in path planning, we examined two patterns for selection methods: data-driven selection and hypothesis-driven selection.

First, the top-five variables that could be intervened were selected based on feature importance in a data-driven selection case: leg score, blood glucose, BMI, waist, and γ -GTP (Fig. 5a). Variables that could be changed by direct or indirect means, unlike such as body height and medical history, were handled as intervenable. The unit cell size of the grid was set to 0.2σ for each explanatory variable, where σ represents the standard deviation of

Fig. 6 Examples of personal actionable paths for treatment with intervention variables based on data-driven selection in systolic blood pressure (SBP) regression task. The optimal paths for improving the response variable predicted by the ML model are represented for randomly selected three examples: instance 1 (**a**) and (**b**), instance 2 (**c**) and (**d**), and instance 3 (**e**) and (**f**). **a, c, e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d, f** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: blood glucose and leg score (**b**), leg score and gamma glutamyl transferase (γ -GTP) (**d**), and blood glucose and leg score (**f**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. 3D plots of the path are shown in Supplementary Fig. 5. BMI body mass index.

the training data. Assuming a scenario wherein the task is to lower the SBP in participants with higher values, relevant instances were selected according to the following criteria: predictive SBP ≥ 140 mmHg⁴² and no missing values in the intervention variables. The number of applicable instances was 259. We executed the path-search algorithm with $L = 20,000$ for each instance and acquired a path to the node with the lowest predictive SBP value in count L . For quantitative evaluations, we introduce the actionability score for each instance that indicates how actionable the planned path was compared with the baseline path (detailed in “Methods”). The histogram of the actionability score is shown in Fig. 5d. The actionability scores were greater than zero, that is, planned paths were more actionable than baseline paths in 227/259 instances, and the median was 0.76. This result suggested that even if the response variable value after improvement was the same, the path planned by our framework had a higher actionability than the baseline path in most cases. The paths of the three randomly selected instances are shown in Fig. 6 and Supplementary Fig. 5. The actual data scatters also support the fact that the path was planned to pass through areas with high-nodal probabilities. In instance 1, it was shown that the path that improved the values of the variables in the following order was more effective: blood glucose, leg score, and γ -GTP (Fig. 6a, b). These variables are related to each other, and the path in which multiple variables fluctuated to improve the blood pressure was reasonable^{43–45}. In instance 2, the path that improved the values in the following order was more valid: γ -GTP, leg score, and again γ -GTP (Fig. 6c, d). The SBP values predicted by the regression model increased temporarily compared with the original SBP value. In instance 3, the optimal path was planned only based on the improvement of one explanatory variable, namely, the blood glucose (Fig. 6e, f). The actionability score yielded a value of zero because the optimal path was identical to the baseline path in such cases.

Second, we also planned paths with another intervention variables set considering a more realistic hypothesis-driven selection. Assuming actual instruction, the following four variables that would fluctuate with the clinical guideline-recommended treatments⁴² were selected by cardiologists as intervention variables: blood glucose, BMI, γ -GTP, and serum sodium (Supplementary Table 2). Instance selection method and other settings in path planning were the same as in the case of data-driven selection. The number of applicable instances was 284. Actionability scores were greater than zero in 265/284 instances, and the median was 1.54 (Fig. 5e). Furthermore, we performed two steps of cardiologist assessments of the improvement paths proposed by the framework. First, we compared the reduction in the predictive SBP of the framework-proposed improvement paths with the cardiologist interventions. As an alternative expression for the cardiologist’s intervention, the most suitable path was selected from ten improvement paths, which included a framework-proposed path and nine randomly intervened paths (Supplementary Fig. 6, detailed in “Methods”). According to the guideline⁴², the direction of improvement was set to decrease for all intervention variables when generating random paths.

Intervention-induced changes in predictive SBP were not displayed to the cardiologists at the time of selection. The framework-proposed paths exhibited a significant decrease in predictive SBP compared to the paths selected by the cardiologists (Fig. 5f). Subsequently, the utility of the planned paths was evaluated. The cardiologists assessed the practicality and informativeness of a framework-proposed path and random path, respectively (Supplementary Fig. 7). The ratio of the paths that the cardiologists evaluated as informative in the framework-proposed paths was 0.42 ± 0.30 (mean $\pm$ standard deviation) (Supplementary Table 3). Details of the cardiologists’ assessment of suggestive individual instances are provided in the Supplementary Notes.

Application of framework on CKD risk classification task. We also applied our framework to the CKD risk classification task. Instead of improving the predicted value in the regression task, we aimed to reduce the predicted probability of the prediction model in the classification task using our framework. According to the guideline for CKD, the eGFR cutoff value for the classification task was set to 60 mL/min/1.73 m²⁴⁶. RFE-selected features comprised items related to renal function, such as age, uric acid, blood urea nitrogen (BUN), triglyceride, anemia-related factors (hemoglobin [Hb], hematocrit [Ht], and erythrocyte count), and blood pressure-related factor (right ankle DBP [RADIA])^{47–52} (Fig. 7a). Hence, the selected explanatory variables were considered to be reasonable for CKD risk prediction.

Following our framework, a classification model and stochastic surrogate model were constructed after the replacement of missing values with the use of the precise imputation method, the multiple imputation method. The classification model yielded an area under the curve (AUC) equal to 0.844 (Fig. 7b). Subsequently, we constructed a stochastic surrogate model based on hierarchical Bayesian modeling. The lowest WBIC was obtained when the number of mixture components was five (Fig. 7c). Using this stochastic surrogate model, we performed path planning in two sets of intervention variables: data-driven selection and hypothesis-driven selection. Assuming a scenario wherein the task is to intervene the participants with decreased renal functions, relevant instances were selected according to the following criteria: predictive class of eGFR < 60 and no missing values in the intervention variables. We executed the path-search algorithm with $L = 20,000$ for each instance and acquired a path to the node with the lowest predicted probabilities of eGFR < 60 in count L . Other settings in path planning were the same as in the case of SBP.

First, the top-five variables that could be intervened were selected based on feature importance in a data-driven selection case: uric acid, BUN, Hb, immunoglobulin A (IgA), and triglyceride (Fig. 7a). The number of applicable instances was 33. Actionability scores were greater than zero in 29/33 instances, and the median was 3.51 (Fig. 7d). The paths of the three randomly selected instances are shown in Fig. 8 and Supplementary Fig. 8. In instance 4, it is shown that the path of improvement in the order of BUN and Hb is effective and succeeds in reducing

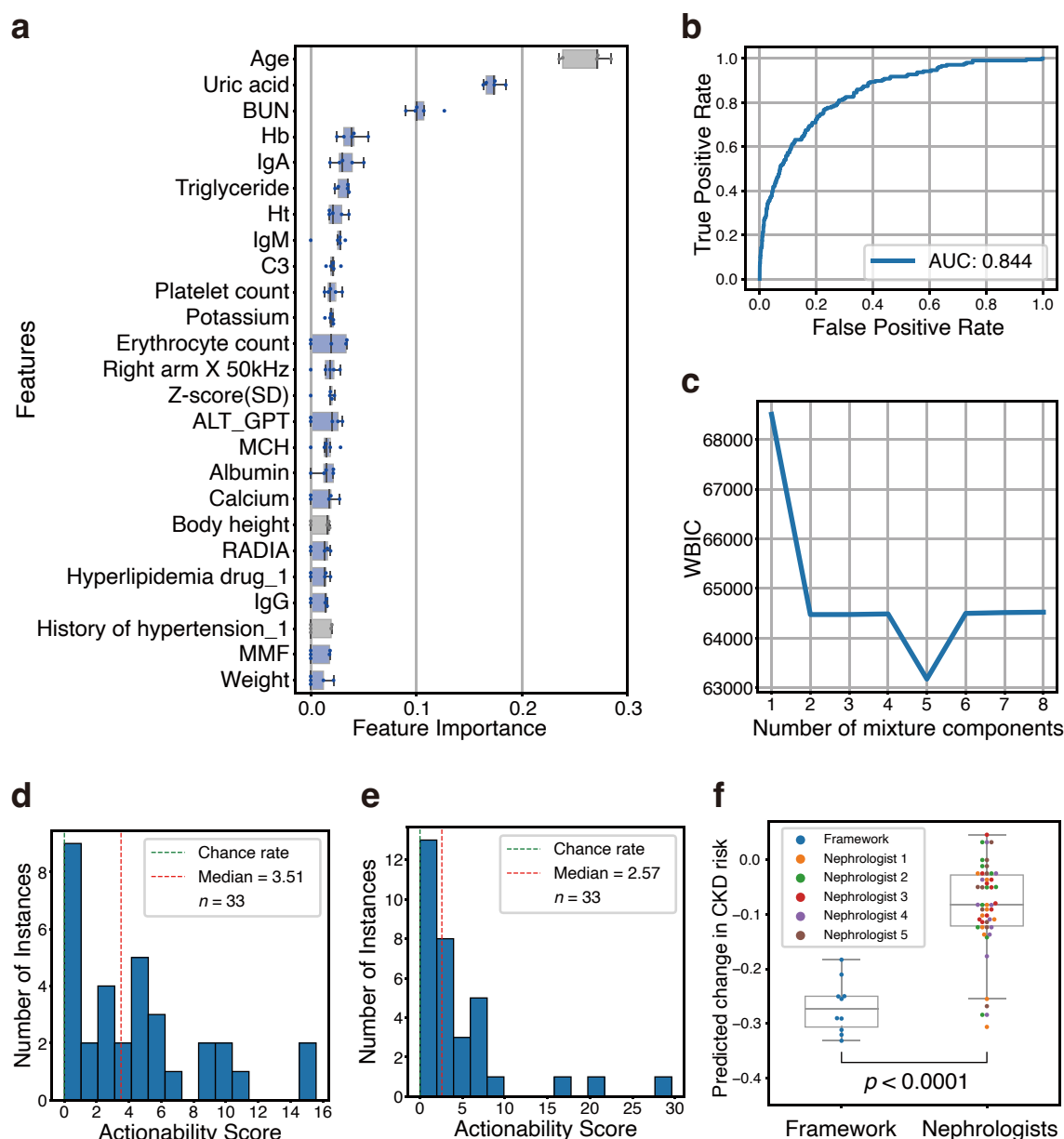
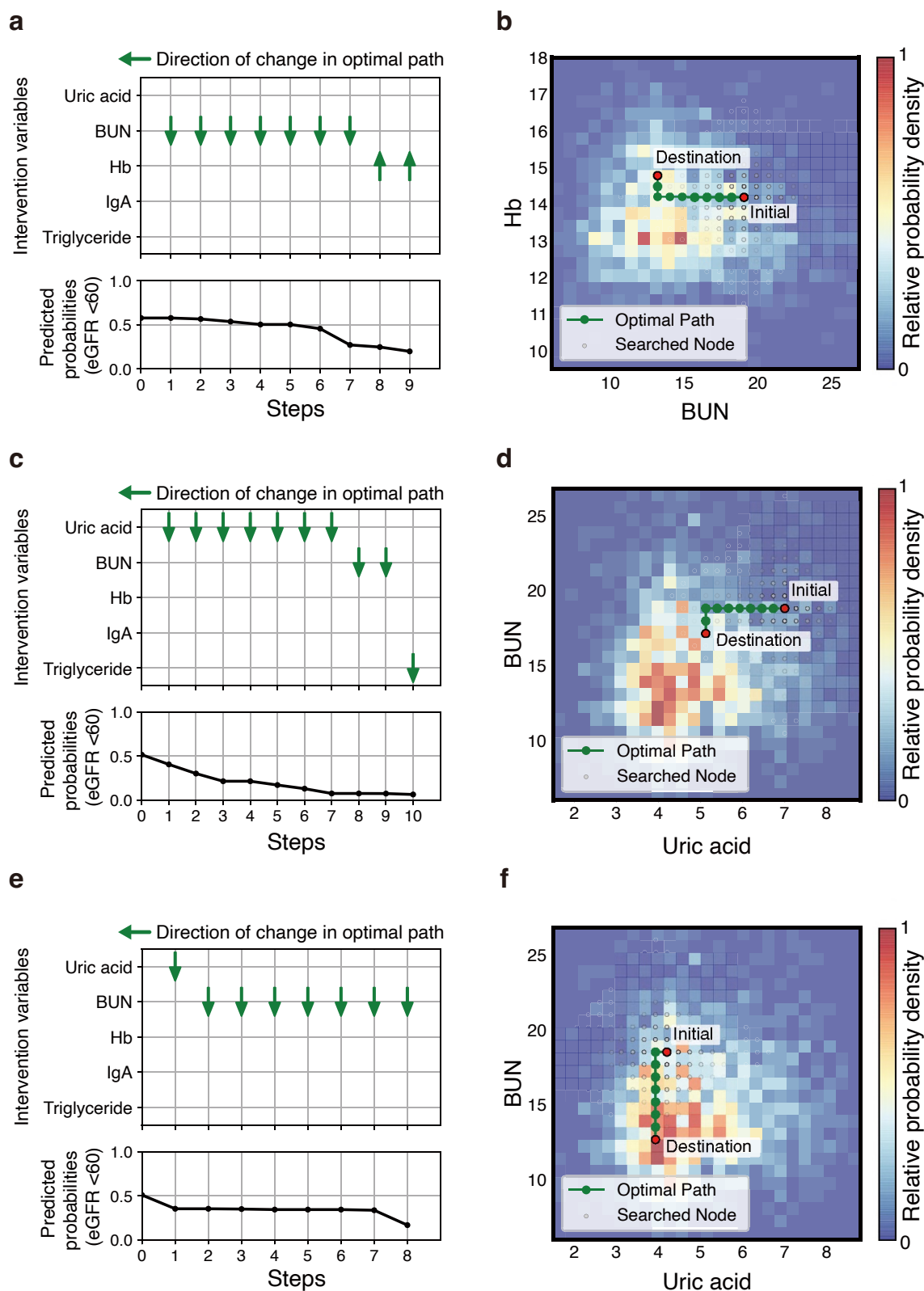


Fig. 7 Application of proposed framework on chronic kidney disease (CKD) risk classification task using the Iwaki health promotion project (IHPP) dataset. **a** Feature importance: these 25 features were selected by recursive feature elimination (RFE) to predict the estimated glomerular filtration rate (eGFR) category. RFE was performed with five-fold cross-validation, and the feature importance when 25 variables remained is shown for each fold ($n = 5$). The plot color represents the following: gray: variables which cannot be intervened, and blue: intervenable variables. Details of features are described in Supplementary Data 1. **b** Classification model score. AUC area under the curve. **c** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **d, e** Histogram of actionability scores with intervention variables based on data-driven selection (**d**) or hypothesis-driven selection (**e**) at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path. **f** Comparison of predicted CKD risk reduction between framework-proposed paths and nephrologist-selected paths. Health improvement paths constructed based on hypothesis-driven intervention variables using our framework were compared with the nephrologist-selected paths among framework-proposed paths and random paths. Each nephrologist evaluated the same randomly selected instances ($n = 10$). Statistical significance was calculated using the Welch's *t*-test (two-sided). In box-plot, center line represents median; box limits, upper and lower quartiles; whiskers, $1.5 \times$ interquartile range.

the prediction probability of eGFR < 60 . In both instances 5 and 6, though the optimal path mainly consisted of improvements in the order of uric acid and BUN, the amount of the total change in each variable was different.

Second, we also planned paths with another intervention variables set based on hypothesis-driven selection. Four variables that would fluctuate with the clinical guideline–recommended treatments⁴⁶ were selected by nephrologists: triglyceride, RADIA,

weight and Hb (Supplementary Table 4). The number of applicable instances was 33. Actionability scores were greater than zero in 32/33 instances, and the median was 2.57 (Fig. 7e). Furthermore, we performed nephrologist assessments of the paths proposed by the framework. First, the reductions in CKD risk, defined as the probability of eGFR < 60 given by the prediction model, were compared between framework-proposed paths and nephrologist interventions. The most suitable path was



selected from a framework-proposed path and nine random paths by the nephrologists. Based on the guideline⁴⁶, the direction of improvement when generating random paths was set as follows: increase for Hb and decrease for triglyceride, RADIA, and weight. The framework-proposed paths exhibited a significant decrease in predictive CKD risk compared to the paths selected by the

nephrologists (Fig. 7f). Subsequently, the utility of the planned paths and random paths were evaluated by the nephrologists. The ratio of paths that nephrologists evaluated as informative in framework-proposed paths was 0.52 ± 0.26 (Supplementary Table 3). Details of the nephrologists' evaluation of suggestive individual instances are provided in the Supplementary Notes.

Fig. 8 Examples of personal actionable paths for treatment with intervention variables based on data-driven selection in chronic kidney disease (CKD) risk classification task. The optimal paths for improving the response variable predicted by the ML model are represented for randomly selected three examples: instance 4 (**a**) and (**b**), instance 5 (**c**) and (**d**), and instance 6 (**e**) and (**f**). **a, c, e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d, f** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: hemoglobin (Hb) and blood urea nitrogen (BUN) (**b**), and BUN and uric acid (**d**) and (**f**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. 3D plots of the path are shown in Supplementary Fig. 8. eGFR estimated glomerular filtration rate, IgA immunoglobulin A.

Discussion

In this study, we proposed a framework for planning paths to improve the prediction values of ML models. We demonstrated that the proposed framework could plan paths through nodes with high probabilities using the synthetic datasets. Furthermore, our proposed framework was capable of planning actionable paths to improve the predicted SBP and CKD risk values in the actual health dataset. The results of applying our framework for the eGFR regression task and the hypertension risk classification task described in the Supplementary Notes also show the successful path planning using our framework. Though conventional surrogate models applied in XAI methods, such as LIME and SHAP, are useful for identifying individual factors that contribute to prediction, they cannot provide the probability of taking variable states. In our framework, the construction of a stochastic surrogate model based on hierarchical Bayesian modeling enabled the estimation of joint probability densities for virtually changed variables and actionable path planning. Our framework suggests realistic concrete treatment processes that are actionable on humans for personal health improvement.

As shown in Figs. 6 and 8, our framework could visually present concrete improvement paths at the individual level. The direction of change in intervention variables was consistent with conventional clinical knowledge for intervention in patients with hypertension^{33–39,42} and CKD^{46,48,50}. For example, high blood glucose levels have been reported to be a risk factor of hypertension, and reduction of triglyceride is an appropriate intervention for patients with impaired renal function. Furthermore, the planned paths consisted of a sequence of changes in intervention variables and could be translated into actual clinical treatments using correspondence tables between the variables and clinical guideline-recommended treatments (Supplementary Tables 2 and 4). In this way, our framework could plan health improvement paths using the intervention variables that would change as a result of clinicians' guidance based on clinical guidelines. Even if the guidelines to be referred to would be changed depending on the geographical region and clinical domain in practical use, the framework-proposed paths could be associated with the clinicians' guidance by using the appropriate guidelines. As a result of clinicians' assessment, the paths planned using our framework exhibited significant improvement in predictive response variables compared to the paths selected by clinicians (Figs. 5f and 7f). Also, the improvement paths presented by our framework were informative for clinicians to some extent (Supplementary Table 3). However, the ratio of paths that clinicians evaluated as practical was not high (Supplementary Table 3). In this study, though we performed an objective path planning without considering clinical constraints, practical adjustments would be needed (detailed in the Supplementary Notes). Tool-assisted goal-settings are expected to help time-constrained clinicians and contribute to better health improvement in patients⁵³. Our framework can provide clinicians with understandable and informative health improvement plans based on patient health data and given intervention variables.

Accordingly, this can be suitable for patient–clinician collaborative decision making on health interventions.

In the remaining part of this section, we provide methodological considerations of the proposed framework. Although the experiments were performed under specific settings, our framework has generality in terms of its methodology owing to three aspects. First, we used XGBoost²⁹ to build the prediction models to be explained in the main manuscript. Because our framework operates in a model-agnostic manner, other high-performance ML models can be used. The Supplementary Notes provide the results of applying the framework to random forest and support vector machine (Supplementary Figs. 9–29 and Supplementary Tables 5–6). In the synthetic datasets, similar paths were planned for the same instances regardless of applied ML algorithms (Fig. 4 and Supplementary Figs. 9–15). However, in the actual dataset, the paths differed in the same instances depending on the ML algorithms (Figs. 6, 8, and Supplementary Figs. 16–29). The actual dataset was more complicated than the synthetic datasets, which resulted in the construction of the prediction models with different properties. Our framework aimed to provide an optimal path to the destination node that would improve the predictions of the ML model. Therefore, the destination node itself was changed when the properties of the ML model were different, which resulted in large differences in the planned paths. Second, although we assumed a mixture distribution of normal or categorical distributions for the explanatory variables in the hierarchical Bayesian modeling (Fig. 2), distributions can be selected according to the data. This is expected to be applied to some extent to medical data, which is often accompanied by a significant amount of noise and missing values. Finally, we selected intervention variables based on the data-driven or hypothesis-driven manner in the path planning with the health checkup data. Also, the unit cell size in planning was set to 0.2 σ of the data distribution. These selection methods and values can be flexibly changed according to the application or the patient's request and environment.

Our objective in path planning was to obtain a path to the best-predicted value with the set conditions of the number of iterations L . In real situations, application-dependent or clinical constraints would exist. A typical case is that the values of explanatory and/or response variables should be less/more than the reference values in all nodes along the path. For example, there were cases where the predicted SBP values were temporarily increased from the original value in our experiment (Fig. 6c). This would be better avoided in clinical situations. By slightly changing the path search condition to exclude undesirable nodes, our framework can be applied to these cases. In addition, there would be use cases where the target value of the response variable or intervention variable should be determined by clinicians (Supplementary Notes). Our framework can be applied by modifying the termination conditions of the search to reach the target value.

From the perspective of expanding the proposed framework, the high-computational cost when many intervention variables exist or when calculating a long-term path for treatment needs to be considered. Subject to our experimental conditions

($L = 20,000$ and five intervention variables), ~ 10 min were required for path planning per instance. Under the condition of more intervention variables, the number of steps of the planned paths decreased, and the calculation time became longer even in the same iteration count, L (Supplementary Figs. 30,31). A more practical path planning can be expected by combining our framework with techniques for finding intervention points, such as counterfactual explanations^{54–58}. Counterfactual explanations usually present intervention goal values of the explanatory variables for changing the response variable without considering the intervention process. Our framework can plan actionable paths to the intervention goal values decided by counterfactual explanations. For this case, a more efficient path planning algorithm, such as the A* search algorithm⁵⁹, can be applied in our framework. Similar to our actionable path planning approach, some studies have been conducted in recent years to obtain intervention points with minimum costs⁵⁸. These approaches enable fast searches by assuming linearity. Our framework is more suitable for use with sophisticated nonlinear ML models, i.e., in cases where linearity is difficult to assume, such as those pertaining to medical checkup data.

Our study has some limitations. First, the health checkup dataset used in this study was obtained from a single area and had a small sample size. This could have contributed to the low-prediction score of the SBP regression model (Fig. 5b). Although the dataset problem does not impair the validity of the proposed framework as a methodology, we performed supplementary experiments that apply the framework on the public datasets to support the framework's validity (Supplementary Notes). Besides, though we selected variables that could be directly or indirectly modifiable as intervention variables because of the nature of health checkup data, it is preferable to build a prediction model with more variables that can be directly intervened. Also, because of the limitation of the observational study data, the stochastic surrogate model only calculated the joint probability density of a set of variables and did not explicitly express the intervention effect considering the causality⁶⁰. By estimating the causal effect of each intervention variable on the response variable through more controlled studies related to individual applications, it may be possible to construct a stochastic surrogate model considering the causality and to present health-improvement paths that are more consistent with the clinician's consideration. Second, in path planning, we set a commonly used and easy-to-interpret grid graph, which resulted in the constraint of changing one variable at a time. Because our framework intervened in explanatory variables stepwise, the fluctuation on the correlated variables was observed as a pattern in which multiple variables alternated. In most cases, multiple variables fluctuated in the optimal paths (Supplementary Fig. 32). In actual application, graphs other than the grid can also be applied according to requirements. Finally, we encountered a problem when we verified the clinical effectiveness of the paths planned by using our framework. Planned paths in the synthetic dataset indicated that our framework would perform correctly. Also, we evaluated that the change directions of intervention variables in some paths were consistent with clinical knowledge in the clinical application. The clinician evaluation results suggested that while the framework would be promising, there are points to be adjusted for its practical application. It is necessary to verify the effectiveness of the paths through a prospective cohort study to suit the real-world applications of our framework.

In conclusion, we proposed a framework to plan actionable health improvement processes at the individual level. Using the synthetic dataset, we proved that our framework could plan actionable paths through the nodes with high probabilities. Furthermore, we successfully demonstrated that health-improving

paths planned for lowering blood pressure and improving CKD risk based on the application of our framework to the actual health checkup dataset were actionable and consistent with clinical knowledge. Our framework can present reasonable and personalized health improvement plans based on ML model predictions in a wide range of situations that is expected to contribute to decision making in the medical field. Further studies should focus on the prospective clinical validation of actionable paths planned by using the framework proposed herein. Our framework may provide clinicians with deeper insights by proposing definite and actionable treatment paths through the use of the ML model.

Methods

Synthetic 3D dataset. We generated a simple 3D dataset to verify whether our framework can plan paths by transiting the nodes with high probabilities in the variable space to improve the predicted response values of the prediction model. The dataset was generated from three, 3D normal distributions to ensure that straight paths were not always actionable (Supplementary Fig. 2). Each distribution generated 200 data points that consisted of x_1 , x_2 , and x_3 . The response variables were set to the sum of x_1 , x_2 , and x_3 with Gaussian noise ($\sigma = 2$). The dataset consisted of 600 data points and randomly split into training data (80%) and test data (20%).

IHPP dataset. To evaluate our framework, we used the IHPP dataset. In this study, we considered the use cases to plan actionable paths to improve the SBP for a regression task and CKD risk for a classification task. CKD was defined by eGFR value, which was calculated using the equation for Japanese population⁶¹. According to the guideline risk classifications, the eGFR cutoff value to 60 mL/min/1.73 m²⁶². The IHPP has annually acquired a wide range of health checkup data that comprise the molecular biology, physiology, biochemistry, lifestyle, and the socio-environmental aspects of residents of the Iwaki district, Hirosaki City, Aomori Prefecture, Japan. In this study, we targeted 12,803 health checkup instances for 13 years (2005–2017). Given the existence of cases where the same person participated in multiple years, the number of unique participants was 3132 (Table 1). The dataset was randomly split into training (80%) and test data (20%) by participants (Supplementary Table 1). This study was approved by the Ethics Committee of Hirosaki University School of Medicine (approval number: 2019-009) and was conducted according to the recommendations of the Declaration of Helsinki. All participants provided written informed consent.

Construction of prediction models. In the experiments, XGBoost²⁹, which is based on a gradient-boosting decision tree algorithm, was used to create the prediction models. In general, XGBoost is a high-performance, nonlinear model. We also examined prediction models based on other nonlinear ML algorithms in supplementary experiments: random forest and support vector machine. The hyperparameters of the model were determined by fivefold cross-validation of the training data. For preprocessing, continuous explanatory variables were standardized by the mean and standard deviation. In addition to this preprocessing, techniques such as other data-dependent preprocessing were applied. These data-dependent preprocessing is described in the results section.

Stochastic surrogate model with hierarchical Bayesian modeling. The graphical model representation of the surrogate model constructed in this study is shown in Fig. 2. In this case, $\mathbf{x}_{\text{cont}}$ represents continuous explanatory variables, such as body composition and blood test data, $\mathbf{x}_{\text{disc}}$ represents discrete explanatory variables, such as sex, and y represents a response variable, such as the blood pressure value. We denote the measured explanatory values as $\mathbf{x}$ and the predictions by the prediction model as $\hat{y}$. z is the parameter of the mixture components, and k represents each mixture component. The data generative process on a regression task is formulated as follows

$$k \sim \text{Categorical}(\pi) \quad (1)$$

$$\mathbf{x}_{\text{cont}} \sim N(\mathbf{m}_k, \Sigma_k) \quad (2)$$

$$\mathbf{x}_{\text{disc}} \sim \text{Categorical}(\phi_{\text{disc},k}) \quad (3)$$

$$y \sim N(\mu_k, \sigma) \quad (4)$$

$$\mu_k = \beta_{1,k} + \beta_{2,k}^T \mathbf{x}_{\text{cont}} + \beta_{3,k}^T \mathbf{x}_{\text{disc}} \quad (5)$$

$$\sigma = \frac{\text{RMSE}_{\text{test}}}{2} \quad (6)$$

where $\text{RMSE}_{\text{test}}$ is an RMSE of the regression model to adjust to the

nonstandardized y scale. The priors for other hyperparameters in Eqs. (1)–(5) are defined as follows

$$\beta_{1,k} \sim N(y_{\text{mean}}, 5y_{\text{std}}) \quad (7)$$

$$\beta_{2,k} \sim \text{DoubleExponential}(0, 1) \quad (8)$$

$$\beta_{3,k} \sim \text{DoubleExponential}(0, 1) \quad (9)$$

$$\pi \sim \text{Dirichlet}(1) \quad (10)$$

$$\mathbf{m}_k \sim N(0, 5\mathbf{I}) \quad (11)$$

$$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 2.5)) \quad (12)$$

$$\phi_{\text{disc},k} \sim \text{Dirichlet}(1) \quad (13)$$

where y_{mean} and y_{std} represent the mean and standard deviation values of the predicted response variable to adjust to the nonstandardized y scale, respectively. Σ_k is a diagonal matrix with elements according to the Cauchy distribution. In a classification task, y is a binary variable and Eqs. (4)–(7) are changed to the following formula accordingly:

$$y \sim \text{Bernoulli}(\theta_k) \quad (14)$$

$$\theta_k = \frac{1}{1 + e^{-\mu_k}} \quad (15)$$

$$\mu_k = \beta_{1,k} + \beta_{2,k}^T \mathbf{x}_{\text{cont}} + \beta_{3,k}^T \mathbf{x}_{\text{disc}} \quad (16)$$

$$\beta_{1,k} \sim N(0, 5) \quad (17)$$

We used PyStan⁶² to estimate the parameters using the Markov chain Monte Carlo algorithm (iteration = 1500, warm-up = 500). For model selection, the WBIC was calculated for each model²⁸. For a stable training, instances with values outside the 3σ range calculated using the training data in the continuous explanatory variables were excluded as outliers.

The priors were selected from noninformative prior distributions or weakly informative prior distributions. The relationships between prior distribution hyperparameters and WBIC are shown in Supplementary Figs. 33–36. Furthermore, we set 1–8 as the range of mixture components from the viewpoint of calculation costs and supplementary experimental results (Supplementary Figs. 37–40). The supplemental results show that the planned paths almost unchanged under a superabundant number of mixture components.

All analyses were carried out using custom software written in Python. Open-source Python packages (pandas, numpy, scipy, xgboost²⁹, scikit-learn, matplotlib, and pystan⁶²) were used in our framework.

Path planning using a stochastic surrogate model. We calculated a path for the treatment for each instance based on a breadth-first search algorithm²⁷. The intervention variable space was regarded as a grid graph, and the grid points (nodes) were connected to plan a path. We defined the probability of the node as the probability of taking the node calculated using the surrogate model. Furthermore, the actionability was defined as the product of nodal probabilities on a specified path. The purpose of this algorithm was used to obtain the most actionable (optimal) path for each node. The output path was the optimal path to the node with the most improved predictive value within the search iteration count, L . From the computational perspective, we used the negative logarithm of actionability as a path cost to minimize, which is mathematically synonymous with maximizing actionability.

The pseudocode of this algorithm is shown in Fig. 3. In lines 2–5, the cost and visited state (the optimal path was searched or not) of nodes were initialized. The search started from the initial node in line 6. We obtained a list of nodes adjacent to the currently selected node in the grid graph in line 8 of the pseudocode. Subsequently, the costs for these nodes were updated in lines 10–12. The following node was selected in line 16. After reaching the predetermined count L , the optimal path to the node with the best ML model prediction value was selected as the planned path in line 18. If multiple nodes with the same predictions existed, the path with the minimum cost was selected.

Actionability score. To evaluate the actionability of the paths, we defined the actionability score expressed by the following equation: $\log(\text{optimal path actionability}) - \log(\text{baseline path actionability})$, where the optimal path actionability is the actionability of the path planned using our framework (Supplementary Fig. 41). The baseline path actionability is the geometric mean of ten actionabilities of paths that connect both ends of the optimal path by the shortest procedure in a random manner. The actionability score indicates the log scale of the ratio of the actionability of the optimal path to the actionability of the baseline path. Hence, the higher the actionability score, the better the optimal path is planned through nodes with higher probabilities than the baseline path. When the actionability score is zero, the optimal path has the same actionability as the baseline path.

Clinician assessment of health-improvement paths. We conducted two steps of assessments. First, we compared the predicted value reductions between the framework-proposed paths and clinicians' interventions. We randomly selected items from the intervention variables and generated random paths in which these items were intervened in a random order on the grid graph in the appropriate direction based on the treatment policy of clinical guidelines. The number of interventions, that is, steps, in the random path was set to be the same as the path planned using the framework in each instance. Nine patterns of random improvement paths were created for each of ten randomly selected instances. For each instance, clinicians selected an improvement path considered to be the closest to their treatment policy among ten blinded improvement paths, which comprised of a path planned by our framework and nine random paths (Supplementary Fig. 6). When the clinicians made a selection, only the order of interventions and the initial values of measurement items of the instance was presented, and the transition of the predictive response variable was not displayed. Five board-certified members of the Japanese Circulation Society assessed the paths in the SBP regression task and five board-certified nephrologists of the Japanese Society of Nephrology in the CKD risk classification task. Regarding the changes in the predictive response variables of the paths proposed by the framework and the paths selected by the clinicians, a two-tailed Welch's t -test was used to determine a significant difference between the two groups.

Second, we assessed the utility of the improvement paths presented by our framework. The clinicians were provided with the initial values of measurement items of the instance, as well as two blinded paths: the framework-proposed path and the random path (Supplementary Fig. 7). The order of interventions along with the accompanied transition of the predictive response variable and the path projected onto 2D heatmaps, which exhibited the actual data distribution, for each pair of intervention variables were presented. The clinicians assessed the practicality and informativeness of paths respectively for ten instances.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The synthetic datasets can be generated from the code in the repository provided in the Code availability section. The datasets used in the Supplementary Information are open available on UCI machine learning repository⁶³ and Trevor Hastie's Software page at <https://web.stanford.edu/~hastie/Papers/LARS/>. The health checkup data used in this study were collected in the Iwaki Health Promotion Project (IHPP) and transferred to a secure data center with restricted access controls in a de-identified format. The de-identified data are available from Hirosaki University School of Medicine (contact via e-mail: coi@hirosaki-u.ac.jp) for academic research purposes only and for researchers who meet the criteria for access to the data. Researchers need to be approved by the research ethics review committees of both the Hirosaki University School of Medicine and their affiliation. Three months are to be expected for the access request to be approved. All other data in this study are included in this article or are available from the corresponding author upon reasonable request.

Code availability

We provide the custom scripts to perform our framework at (https://github.com/clinfo/actionable_path_planning) with an assigned (<https://doi.org/10.5281/zenodo.4706186>).

Received: 13 November 2020; Accepted: 23 April 2021;

Published online: 25 May 2021

References

- Boult, C. et al. Early effects of "guided care" on the quality of health care for multimorbid older persons: a cluster-randomized controlled trial. *J. Gerontol.* **63**, 321–327 (2008).
- Wolff, J. L. et al. Effects of guided care on family caregivers. *Gerontologist* **50**, 459–470 (2010).
- Boyd, C. M. et al. The effects of guided care on the perceived quality of health care for multi-morbid older persons: 18-month outcomes from a cluster-randomized controlled trial. *J. Gen. Intern. Med.* **25**, 235–242 (2010).
- Djulgovic, B. & Guyatt, G. H. Evidence-based practice is not synonymous with delivery of uniform health care. *JAMA* **312**, 1293 (2014).
- Goetz, L. H. & Schork, N. J. Personalized medicine: motivation, challenges, and progress. *Fertil. Steril.* **109**, 952–963 (2018).
- Chan, I. S. & Ginsburg, G. S. Personalized medicine: progress and promise. *Annu. Rev. Genomics Hum. Genet.* **12**, 217–244 (2011).
- Snyderman, R. & Williams, R. S. Prospective medicine: the next health care transformation. *Acad. Med.* **78**, 1079–1084 (2003).
- Djulgovic, B. & Guyatt, G. H. Progress in evidence-based medicine: a quarter century on. *Lancet* **390**, 415–423 (2017).

9. Lenzen, S. A., Daniëls, R., van Bokhoven, M. A., van der Weijden, T. & Beurskens, A. Disentangling self-management goal setting and action planning: a scoping review. *PLoS ONE* **12**, 1–22 (2017).
10. Kim, J., Thrift, A. G., Nelson, M. R., Bladin, C. F. & Cadilhac, D. A. Personalized medicine and stroke prevention: where are we? *Vasc. Health Risk Manag.* **11**, 601–611 (2015).
11. Murphy, M. E., Fahey, T. & Smith, S. M. Computer-based clinical decision support for general practitioners. *Fam. Pract.* **31**, 497–498 (2014).
12. Darcy, A. M., Louie, A. K. & Roberts, L. W. Machine learning and the profession of medicine. *JAMA* **315**, 551–552 (2016).
13. Ahmed, Z., Mohamed, K., Zeeshan, S. & Dong, X. Q. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database* **2020**, 1–35 (2020).
14. Ho, D. S. W., Schierding, W., Wake, M., Saffery, R. & O'Sullivan, J. Machine learning SNP based prediction for precision medicine. *Front. Genet.* **10**, 1–10 (2019).
15. Lenert, M. C., Matheny, M. E. & Walsh, C. G. Prognostic models will be victims of their own success, unless. *J. Am. Med. Inform. Assoc.* **26**, 1645–1650 (2019).
16. Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, oD. B. What do we need to build explainable AI systems for the medical domain? arxiv:1712.09923 (2017).
17. Adadi, A. & Berrada, M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* **6**, 52138–52160 (2018).
18. Caruana, R. et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In *Proc. 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD'15* 1721–1730 (2015).
19. Zhang, A., Teng, L. & Alterovitz, G. OUP accepted manuscript. *J. Am. Med. Inform. Assoc.* **0**, 1–8 (2020).
20. Ribeiro, M. T., Singh, S. & Guestrin, C. “Why should I trust you?” Explaining the predictions of any classifier. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13–17-Aug, 1135–1144 (2016).
21. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, 4766–4775 (2017).
22. Katuwal, G. J. & Chen, R. machine learning model interpretability for precision medicine. arxiv:1610.09045 (2016).
23. Elshaw, R., Sherif, Y., Al-Mallah, M. & Sakr, S. Interpretability in healthcare a comparative study of local machine learning interpretability techniques. In *Proc. IEEE Symposium on Computer-Based Medical Systems* 2019-June, 275–280 (2019).
24. Elshaw, R., Al-Mallah, M. H. & Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med. Inform. Decis. Mak.* **19**, 146 (2019).
25. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).
26. Becquemont, L. et al. Personalized medicine: how to switch from the concept to the integration into the clinical development plan to obtain marketing authorization. *Therapies* **67**, 349–357 (2012).
27. Skiena, S. S. *The Algorithm Design Manual*. Springer Science+Business Media (Springer, London, 2008). <https://doi.org/10.1007/978-1-84800-070-4>.
28. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14**, 867–897 (2013).
29. Chen, T. & Guestrin, C. XGBoost. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vol. 42 785–794 (ACM, 2016).
30. le Thi, H. A., Nguyen, V. V. & Ouchani, S. Gene selection for cancer classification using DCA. *Lect. Notes Comput. Sci.* **5139**, 62–72 (2008).
31. Kannel, W. Fifty years of Framingham study contributions to understanding hypertension. *J. Hum. Hypertens.* **14**, 83–90 (2000).
32. Rockwood, M. R. H. & Howlett, S. E. Blood pressure in relation to age and frailty. *Can. Geriatr. J.* **14**, 2–7 (2011).
33. Foulds, H. J. A., Bredin, S. S. D. & Warburton, D. E. R. The relationship between hypertension and obesity across different ethnicities. *J. Hypertens.* **30**, 359–367 (2012).
34. Ye, S. et al. Associations of body composition with blood pressure and hypertension. *Obesity* **26**, 1644–1650 (2018).
35. Wada, T., Kawasaki, Y. & Inaji, J. Establishing borderline and at-risk regions for estimated skeletal muscle mass of legs determined with a body composition meter. *Ningen Dock Int.* **2**, 14–18 (2014).
36. Tatsumi, Y. et al. Fasting blood glucose predicts incidence of hypertension independent of HbA1c levels and insulin resistance in middle-aged Japanese: the Saku study. *Am. J. Hypertens.* **32**, 1178–1185 (2019).
37. Kuwabara, M. et al. Fasting blood glucose is predictive of hypertension in a general Japanese population. *J. Hypertens.* **37**, 167–174 (2019).
38. Shimizu, Y. et al. Gamma-glutamyl transpeptidase (γ -GTP) has an ambivalent association with hypertension and atherosclerosis among elderly Japanese men: a cross-sectional study. *Environ. Health Prev. Med.* **24**, 1–7 (2019).
39. Yamada, Y. et al. The relationship between serum gamma-glutamyl transpeptidase levels and hypertension: common in drinkers and nondrinkers. *Hypertension Res.* **18**, 295–301 (1995).
40. Adrogué, H. J. & Madias, N. E. Sodium and potassium in the pathogenesis of hypertension. *N. Engl. J. Med.* **356**, 1966–1978 (2007).
41. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
42. Umemura, S. et al. The Japanese society of hypertension guidelines for the management of hypertension (JSH 2019). *Hypertens. Res.* **42**, 1235–1481 (2019).
43. Lippi, G., Targher, G. & Guidi, G. C. Relationship between γ -glutamyltransferase, fasting plasma glucose, and triglycerides in the general population. *Clin. Chem.* **53**, 1866–1867 (2007).
44. Liang, J. et al. Serum gamma-glutamyltransferase is associated with impaired fasting glucose in Chinese adults: the cardiometabolic risk in Chinese (CRC) study. *Cell Biochem. Biophys.* **70**, 1823–1828 (2014).
45. Jo, A. & Mainous, A. G. Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: a nationally representative cross-sectional study. *BMJ Open* **8**, 1–6 (2018).
46. Japanese Society of Nephrology. Evidence-based clinical practice guideline for CKD 2013. *Clin. Exp. Nephrol.* **18**, 346–423 (2014).
47. Weinstein, J. R. & Anderson, S. The aging kidney: physiological changes. *Adv. Chronic Kidney Dis.* **17**, 302–307 (2010).
48. Obermayr, R. P. et al. Elevated uric acid increases the risk for kidney disease. *J. Am. Soc. Nephrol.* **19**, 2407–2413 (2008).
49. Klein, L. et al. Admission or changes in renal function during hospitalization for worsening heart failure predict postdischarge survival: results from the outcomes of a prospective trial of intravenous milrinone for exacerbations of chronic heart failure (OPTIME-CHF). *Circul. Heart Fail.* **1**, 25–33 (2008).
50. Lanktree, M. B., Thériault, S., Walsh, M. & Paré, G. HDL cholesterol, LDL cholesterol, and triglycerides as risk factors for CKD: a mendelian randomization study. *Am. J. Kidney Dis.* **71**, 166–172 (2018).
51. Rossert, J. et al. Effect of early correction of anemia on the progression of CKD. *Am. J. Kidney Dis.* **47**, 738–750 (2006).
52. Hall, J. et al. Obesity, hypertension, and chronic kidney disease. *Int. J. Nephrol. Renovasc. Dis.* **7**, 75–88 (2014).
53. Mann, D. M., Palmisano, J. & Lin, J. J. A pilot randomized trial of technology-assisted goal setting to improve physical activity among primary care patients with prediabetes. *Prev. Med. Rep.* **4**, 107–112 (2016).
54. Wachter, S., Mittelstadt, B. & Russell, C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harv. J. Law Technol.* **31**, 841–887 (2017).
55. van Looveren, A. & Klaise, J. Interpretable counterfactual explanations guided by prototypes. arxiv:1907.02584 (2019).
56. Tolomei, G., Silvestri, F., Haines, A. & Lalmas, M. Interpretable predictions of tree-based ensembles via actionable feature tweaking. In *Proc. ACM SIGKDD 716 International Conference on Knowledge Discovery and Data Mining Part F1296*. 465–474 (2017).
57. Laugel, T., Lesot, M.-J., Marsala, C., Renard, X. & Detyniecki, M. Inverse classification for comparison-based interpretability in machine learning. *Commun. Comput. Inf. Sci.* **853**, 100–111 (2017).
58. Ustun, B., Spangher, A. & Liu, Y. Actionable recourse in linear classification. In *Proc. 2019 Conference on Fairness, Accountability, and Transparency FAT** 2019. 10–19 (2019).
59. Goldberg, A. V. & Harrelson, C. Computing the shortest path: A* search meets graph theory. In *Proc. Annual ACM-SIAM Symposium on Discrete Algorithms* 156–165 (2005).
60. Robins, M. & James, M. A. H. *Causal Inference—what if*. (Chapman & Hall/CRC: Boca Raton, 2020).
61. Matsuo, S. et al. Revised equations for estimated GFR from serum creatinine in Japan. *Am. J. Kidney Dis.* **53**, 982–992 (2009).
62. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
63. Dua, D. & Graff, C. *UCI Machine Learning Repository*. <http://archive.ics.uci.edu/ml> (2019).

Acknowledgements

This work was supported by JST, Center of Innovation Program (JPMJCE1302), and Kyowa Hakko Bio Co., Ltd. The author, Kazuki Nakamura, thanks Kazushi Shoji, Miho Komatsu, Takashi Ishida, and Yuko Shimanami, employees of Kyowa Hakko Bio Co., Ltd., for their generous support. The authors thank Minoru Sakuragi, M.D. for helpful comments on our paper. The authors would like to thank cardiologists Kazuki Matsushita, M.D., Masahiro Kimura, M.D., Ph.D., Osamu Baba, M.D., Ph.D., Sawa Miyagawa, M.D., and Shuhei Tsuji M.D. for their cooperation in clinical assessment. The authors appreciate nephrologists Daisuke Takada, M.D., Naoya Torii, M.D., Noriaki

Sato, M.D., Shusuke Hiragi, M.D., MBA, Ph.D., and Takuya Ishimura, M.D. for their cooperation in clinical assessment.

Competing interests

Kazuki Nakamura is an employee of Kyowa Hakko Bio Co., Ltd. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23319-1>.

Correspondence and requests for materials should be addressed to Y.O.

Peer review information *Nature Communications* thanks Liangyuan Hu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Supplementary Information

Health improvement framework for actionable treatment planning using a surrogate Bayesian model

Kazuki Nakamura^{1,2}, Ryosuke Kojima², Eiichiro Uchino², Koh Ono³, Motoko Yanagita^{4,5}, Koichi Murashita⁶, Ken Itoh⁷, Shigeyuki Nakaji⁸ and Yasushi Okuno^{2*}

¹ Research & Business Development Department, Kyowa Hakko Bio Co., Ltd., Tokyo, 100-0004, Japan.

² Department of Biomedical Data Intelligence, Graduate School of Medicine, Kyoto University, Kyoto, 606-8507, Japan.

³ Department of Cardiovascular Medicine, Graduate School of Medicine, Kyoto University, Kyoto, 606-8507, Japan.

⁴ Department of Nephrology, Graduate School of Medicine, Kyoto University, Kyoto, 606-8507, Japan.

⁵ Institute for the Advanced Study of Human Biology, Kyoto University, Kyoto, 606-8507, Japan.

⁶ Center of Innovation Research Initiatives Organization, Hirosaki University, Zaifu-cho, Hirosaki, Aomori, 036-8562, Japan.

⁷ Department of Stress Response Science, Hirosaki University Graduate School of Medicine, Zaifu-cho, Hirosaki, Aomori, 036-8562, Japan.

⁸ Department of Social Health, Hirosaki University Graduate School of Medicine, Zaifu-cho, Hirosaki, Aomori, 036-8562, Japan.

* Address correspondence to Yasushi Okuno, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto, 606-8507, Japan, TEL: +81-75-751-4881, FAX: +81-75-751-4881, E-mail: okuno.yasushi.4c@kyoto-u.ac.jp

Supplementary Notes

Four supplementary experiments are described below. First, we conducted a more extensive evaluation of our framework with a five-dimensional (5D) synthetic dataset considering the real-life applications than the three-dimensional (3D) synthetic dataset. Second, details of the clinicians' evaluations of the improvement paths planned using our framework are described. Third, applications on the hypertension risk classification task and estimated glomerular filtration rate (eGFR) regression task using Iwaki Health Promotion Project (IHPP) dataset are described. Fourth, we applied our framework on a non-health dataset and another simple health dataset because the prediction model based on the health dataset used in the main text had low-prediction scores.

Validation of framework on 5D synthetic dataset

We generated a 5D dataset to verify whether our framework can plan paths by transiting the nodes with high probabilities in the variable space. This synthetic dataset was generated from five, 5D normal distributions (Supplementary Fig. 42). Assuming the correlation between explanatory variables, covariance was randomly set. Each distribution generated 200 data points that consisted of x_i ($i \in \{1, \dots, 5\}$). The response variables were set to the sum of x_i with Gaussian noise ($\sigma = 2$). The dataset consisted of 1,000 data points and randomly split into training data (80%) and test data (20%).

We built regression models based on XGBoost, random forest (RF), and support vector machine (SVM) (Supplementary Fig. 12a–c). Subsequently, stochastic surrogate models using hierarchical Bayesian modeling were constructed (Supplementary Fig. 12d–f). Using the stochastic surrogate models with lowest widely applicable Bayesian information criterion (WBIC), we planned paths to decrease the value of the response variable. All explanatory variables were selected for intervention variables, and the unit cell size of the grid was set to 0.5σ in the training data for each explanatory variable. The path search algorithm was executed with $L = 20,000$ for each instance, and an optimal path to the node with the lowest predictive value was acquired. Planned paths were more actionable than the baseline paths (Supplementary Fig. 12g–i). The planned paths of two randomly selected instances for each prediction model are shown in Supplementary Figs. 13–15. We demonstrated that our framework applied to the dataset with covariance and could be used to plan paths with five intervention variables.

Details of utility assessments of health improvement paths planned by framework

The clinician assessments were performed on the utility of the framework related to the systolic blood pressure (SBP) regression task and chronic kidney disease (CKD) risk classification task as described in the main text. This section provides a discussion of suggestive instances in clinicians' evaluation.

Regarding instance 11 in the SBP regression task, the clinician evaluated that while lowering blood glucose might be reasonable in lowering blood pressure, the destination value of blood glucose was too low, which could pose other risks associated with hypoglycemia (Supplementary Fig. 43a–c). In this study, the search range, i.e., upper and lower limit, for the intervention variables were not defined to perform the objective path planning in the path search. While this can lead to more effective blood pressure-improving effects with interventions that clinicians do not usually take, it can also have side effects on the patient's health. Therefore, for example, regarding the blood glucose level, it would be better to avoid undesirable intervention by setting the lower limit based on the guideline or clinician's knowledge in practical application. Regarding instance 12, the clinicians disagreed with the framework-proposed path in which blood glucose and gamma glutamyl transferase (γ -GTP) were intervened instead of clearly high body mass index (BMI) (Supplementary Fig. 43d–f). As is shown in the heatmap, there were few instances where the BMI exceeds 30 in the dataset, which might cause overfitting on the original nonlinear prediction model construction. For example, if there are accidentally a few instances with obesity but low blood pressure, the prediction model may present predictive blood pressure lower than the actual regarding obesity instances. Since our framework presents the optimal path to the node with the most improved response variable value in the search iteration count based on the predictive model, the prediction model has a major impact on the destination node selection. Although we used health checkup data for a specific region, it would be desired to build a robust predictive model using a larger sample size dataset.

Regarding instance 13 in the CKD risk classification task, the framework-proposed path was composed of the alternating interventions in high triglyceride and right ankle diastolic blood pressure (RADIA), which is blood pressure-related indicators (Supplementary Fig. 44a–c). Also, the path was consistent with the distribution of actual data for these variables. From these results, the clinicians

evaluated that this health-improvement path was plausible and could lead to behavioral changes in the patient. In this study, we used a grid graph for the path search algorithm, and only one variable is intervened at a time. Paths would be planned by repeating interventions, and simultaneous fluctuations of multiple variables could be represented by alternating fluctuations. Regarding instance 14, the clinicians evaluated that intervention on overweight and blood pressure would be necessary though the framework-proposed path consisted of intervention in triglyceride and hemoglobin (Hb) (Supplementary Fig. 44d–f). While the path planned to improve the response variable without specifying target values of intervention variables might present new possibilities, it could be better in some cases that the clinician set target values of intervention variables. By changing the search end condition in the path search algorithm, our framework can present the optimal path to the node of the targeted-intervention variable values determined by the clinicians.

Application of framework on hypertension risk classification task using IHPP dataset

We applied our framework to the hypertension risk classification task. According to the guideline for hypertension, the SBP cutoff value for the classification task was set to 140 mmHg¹. As in the main text, we performed explanatory variable reduction using recursive feature elimination (RFE). Important features selected by XGBoost-based RFE comprised items related to hypertension, such as age, body composition (leg score and BMI), blood glucose, and γ -GTP^{2–10} (Supplementary Fig. 22a). Following our framework, classification models were built after the replacement of missing values with the use of multiple imputations. Area under the curve (AUC) of the models are summarized in Supplementary Table 6. Next, we constructed stochastic surrogate models based on multiple imputed RFE-selected features and predicted values (Supplementary Fig. 22b–d). Though we initially set the range of mixture components to 1–8, the lowest WBIC values were obtained when the numbers of mixture components were eight in XGBoost and RF. Therefore, we expanded the range to 16 in these cases. Subsequently, we performed path planning using the stochastic surrogate models selected based on WBIC. The top-five variables that could be intervened were selected as intervention variables based on feature importance in a data-driven manner: γ -GTP, blood glucose, aspartate transaminase (AST [GOT]), leg score, and immunoglobulin G (IgG). Assuming a scenario wherein the task is to improve participants with higher SBP, relevant instances were selected according to the following

criteria: XGBoost-predictive class of SBP ≥ 140 mmHg¹, and no missing values in the intervention variables. The number of applicable instances was 162. We executed the path-search algorithm with $L = 20,000$ for each instance and acquired a path to the node with the lowest predictive SBP value in count L . The histograms of the actionability score are shown in Supplementary Fig. 22e–g. The paths of the two randomly selected instances are shown in Supplementary Fig. 23–25.

Application of framework on eGFR regression task using IHPP dataset

We applied our framework to the eGFR regression task. As in the main text, we performed explanatory variable reduction using RFE. Important features selected by RFE comprised items related to renal function, such as age, uric acid, blood urea nitrogen (BUN), anemia-related factors (Hb and erythrocyte count), and body composition-related factors (right arm X 50 kHz and basal metabolic rate [BMR] score)^{11–15} (Supplementary Fig. 26a). Following our framework, regression models were built after the replacement of missing values with the use of multiple imputations. The scores of the prediction models are shown in Supplementary Fig. 26b–d. Next, we constructed stochastic surrogate models based on hierarchical Bayesian modeling based on multiple imputed RFE-selected features and predicted values (Supplementary Fig. 26e–g). Subsequently, we performed path planning using the stochastic surrogate models selected based on WBIC. The top-five variables that could be intervened were selected as intervention variables based on feature importance in a data-driven manner: uric acid, BUN, right arm X 50 kHz, BMR score, and erythrocyte count. Assuming a scenario wherein the task is to intervene the participants with decreased renal functions, relevant instances were selected according to the following criteria: XGBoost-predictive eGFR < 60 mL/min/1.73 m²¹⁶, and no missing values in the intervention variables. The number of applicable instances was 14. We executed the path-search algorithm with $L = 20,000$ for each instance and acquired a path to the node with the highest predictive eGFR value in count L . The histograms of the actionability score are shown in Supplementary Fig. 26h–j. The paths of the two randomly selected instances are shown in Supplementary Figs. 27–29.

Evaluation on real estate dataset

We obtained a dataset from the UCI Machine Learning Repository^{17,18}. This dataset was used to conduct a regression analysis of the house price of a unit area based on six continuous explanatory variables of real estate data. There were no missing values in the dataset. The dataset was randomly split into training data (80%) and test data (20%).

The trained XGBoost regression model yielded a root-mean-square error (RMSE) of 6.78 and an R-squared value of 0.735 in the test data (Supplementary Fig. 45a). A surrogate model was constructed by hierarchical Bayesian modeling using the original data and the predicted values of the regression model. The lowest WBIC value was obtained when the number of mixture components was two (Supplementary Fig. 45b).

Subsequently, path planning was performed using the surrogate model. Among the six explanatory variables, X_1 (the transaction date) was a variable that was difficult to intervene. Therefore, X_1 was fixed, and the remaining five variables were selected as intervention variables. The unit cell size of the grid was set to 0.2σ in the training data for each explanatory variable. We executed the path search algorithm with $L = 20,000$ for each instance, and the path with the highest predicted value was acquired. The histogram of the actionability score for each instance is shown in Supplementary Fig. 45c. The actionability scores were greater than zero in 75/82 instances, and the median was 5.25.

From these results, we have demonstrated that our framework can be applicable to datasets with higher regression model scores.

Evaluation on public dataset for disease progression

To evaluate the feasibility of our framework on another small health dataset, we used a public dataset on diabetes progression^{19,20}. This dataset was used to conduct a regression of the quantitative measure of diabetes progression one year after the baseline from nine continuous and one discrete explanatory variable (Supplementary Table 7). The dataset is openly available on Trevor Hastie's Software page at <https://web.stanford.edu/~hastie/Papers/LARS/>. This dataset contains no missing values. The dataset was randomly split into training data (80%) and test data (20%).

The feature importance of the trained model is shown in Supplementary Fig. 46a. The XGBoost regression model yielded an RMSE value of 62.19 and an R-squared value of 0.246 in the test data

(Supplementary Fig. 46b). A surrogate model was constructed by hierarchical Bayesian modeling using the original data and the predicted values of the regression model. The lowest WBIC value was obtained when the number of mixture components was two (Supplementary Fig. 46c).

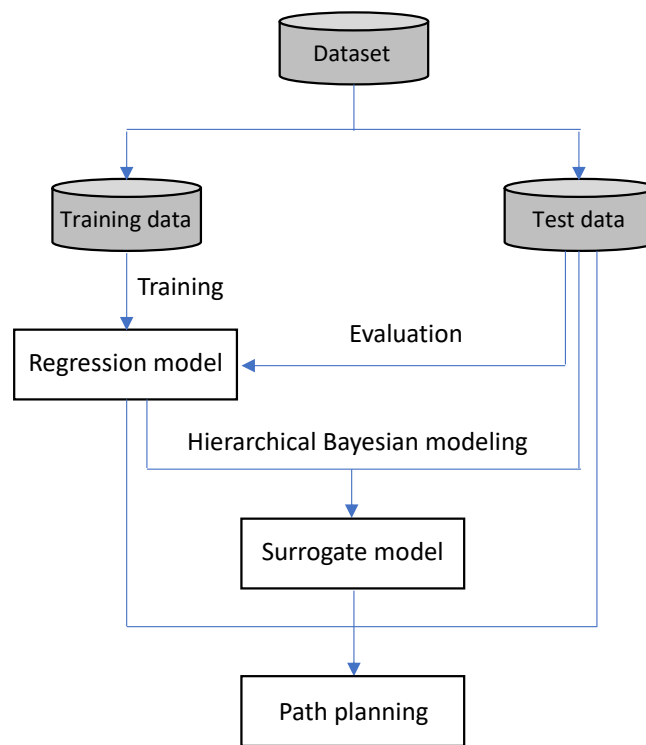
Subsequently, path planning was performed using the surrogate model. Regarding the intervention variables, five variables were selected from the top of the feature importance of the regression model: body mass index (BMI), blood pressure, T-cells, high-density lipoproteins, and lamotrigine (Supplementary Fig. 46a). The remaining variables were fixed. The unit cell size of the grid was set to 0.2σ in the training data for each explanatory variable. We executed the path search algorithm with $L = 20,000$ in each instance, and the path with the lowest predicted value was acquired. The histogram of the actionability score for each instance is shown in Supplementary Fig. 46d. The actionability scores were greater than zero in 83/87 instances, and the median was 2.06. Examples of the paths planned by using the proposed framework are shown in Supplementary Fig. 47. This experiment indicates the feasibility of using the proposed framework in planning actionable paths to improve the predictions of the regression model.

Supplementary References

1. Umemura, S. *et al.* The Japanese Society of Hypertension Guidelines for the Management of Hypertension (JSH 2019). *Hypertension Research* **42**, 1235–1481 (2019).
2. Kannel, W. Fifty years of Framingham Study contributions to understanding hypertension. *Journal of Human Hypertension* **14**, 83–90 (2000).
3. Rockwood, M. R. H. & Howlett, S. E. Blood pressure in relation to age and frailty. *Canadian Geriatrics Journal* **14**, 2–7 (2011).
4. Foulds, H. J. A., Bredin, S. S. D. & Warburton, D. E. R. The relationship between hypertension and obesity across different ethnicities. *Journal of Hypertension* **30**, 359–367 (2012).
5. Ye, S. *et al.* Associations of Body Composition with Blood Pressure and Hypertension. *Obesity* **26**, 1644–1650 (2018).
6. Wada, T., Kawasaki, Y. & Inaji, J. Establishing Borderline and At-risk Regions for Estimated Skeletal Muscle Mass of Legs Determined with a Body Composition Meter. *Ningen Dock International* **2**, 14–18 (2014).
7. Tatsumi, Y. *et al.* Fasting Blood Glucose Predicts Incidence of Hypertension Independent of HbA1c Levels and Insulin Resistance in Middle-Aged Japanese: The Saku Study. *American Journal of Hypertension* **32**, 1178–1185 (2019).
8. Kuwabara, M. *et al.* Fasting blood glucose is predictive of hypertension in a general Japanese population. *Journal of Hypertension* **37**, 167–174 (2019).
9. Shimizu, Y. *et al.* Gamma-glutamyl transpeptidase (γ -GTP) has an ambivalent association with hypertension and atherosclerosis among elderly Japanese men: A cross-sectional study. *Environmental Health and Preventive Medicine* **24**, 1–7 (2019).
10. Yamada, Y. *et al.* The Relationship between Serum Gamma-Glutamyl Transpeptidase Levels and Hypertension: Common in Drinkers and Nondrinkers. *Hypertension Research - Clinical and Experimental* **18**, 295–301 (1995).
11. Wan, C. S. *et al.* Bioelectrical impedance analysis to estimate body composition, and change in adiposity, in overweight and obese adolescents: Comparison with dual-energy x-ray absorptiometry. *BMC Pediatrics* **14**, 1–10 (2014).
12. Weinstein, J. R. & Anderson, S. The Aging Kidney: Physiological Changes. *Advances in Chronic Kidney Disease* **17**, 302–307 (2010).
13. Obermayr, R. P. *et al.* Elevated uric acid increases the risk for kidney disease. *Journal of the American Society of Nephrology* **19**, 2407–2413 (2008).
14. Rossert, J. *et al.* Effect of Early Correction of Anemia on the Progression of CKD. *American Journal of Kidney Diseases* **47**, 738–750 (2006).
15. Hall, J. *et al.* Obesity, hypertension, and chronic kidney disease. *International Journal of Nephrology and Renovascular Disease* **75** (2014).
16. Japanese Society of Nephrology. Evidence-based Clinical Practice Guideline for CKD 2013. *Clinical and Experimental Nephrology* **18**, 346–423 (2014).

17. Dua, D. & Graff, C. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml> (2019).
18. Yeh, I. C. & Hsu, T. K. Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing Journal* **65**, 260–271 (2018).
19. Tibshirani, R., Johnstone, I., Hastie, T. & Efron, B. Least angle regression. *The Annals of Statistics* **32**, 407–499 (2004).
20. Hastie, T. & Efron, B. lars: Least Angle Regression, Lasso and Forward Stagewise. *R package version 1.2* (2013).
21. Matsuo, S. *et al.* Revised Equations for Estimated GFR From Serum Creatinine in Japan. *American Journal of Kidney Diseases* **53**, 982–992 (2009).
22. Leib, E. S., Lewiecki, E. M., Binkley, N. & Hamdy, R. C. Official Positions of the International Society for Clinical Densitometry. *Journal of Clinical Densitometry* **7**, 1–5 (2004).
23. Hye, S. C., Cheon, W. C., Myung, J. P., Hong, M. K. & Yoo, J. H. Clinical value of a desktop spirometer (HI-801) for spirometry screening. *Tuberculosis and Respiratory Diseases* **62**, 276–283 (2007).
24. Kubota, M. *et al.* Reference values for spirometry, including vital capacity, in Japanese adults calculated with the LMS method and compared with previous values. *Respiratory Investigation* **52**, 242–250 (2014).
25. Momma, H. *et al.* Physical fitness tests and type 2 diabetes among Japanese: A longitudinal study from the niigata wellness study. *Journal of Epidemiology* **29**, 139–146 (2019).

Supplementary Figures



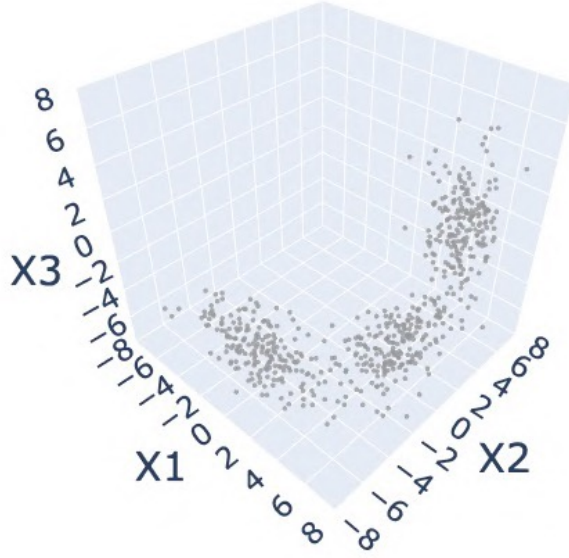
Supplementary Figure 1. Detailed workflow of proposed framework.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad k \in \{1, 2, 3\}$$

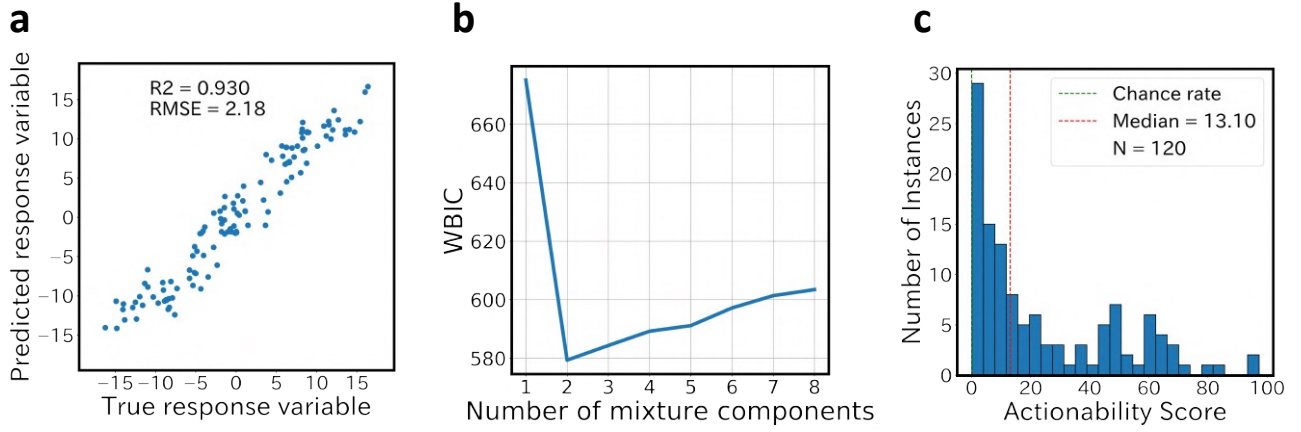
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ -5 \\ -5 \end{bmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 5 \\ 0 \\ -5 \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

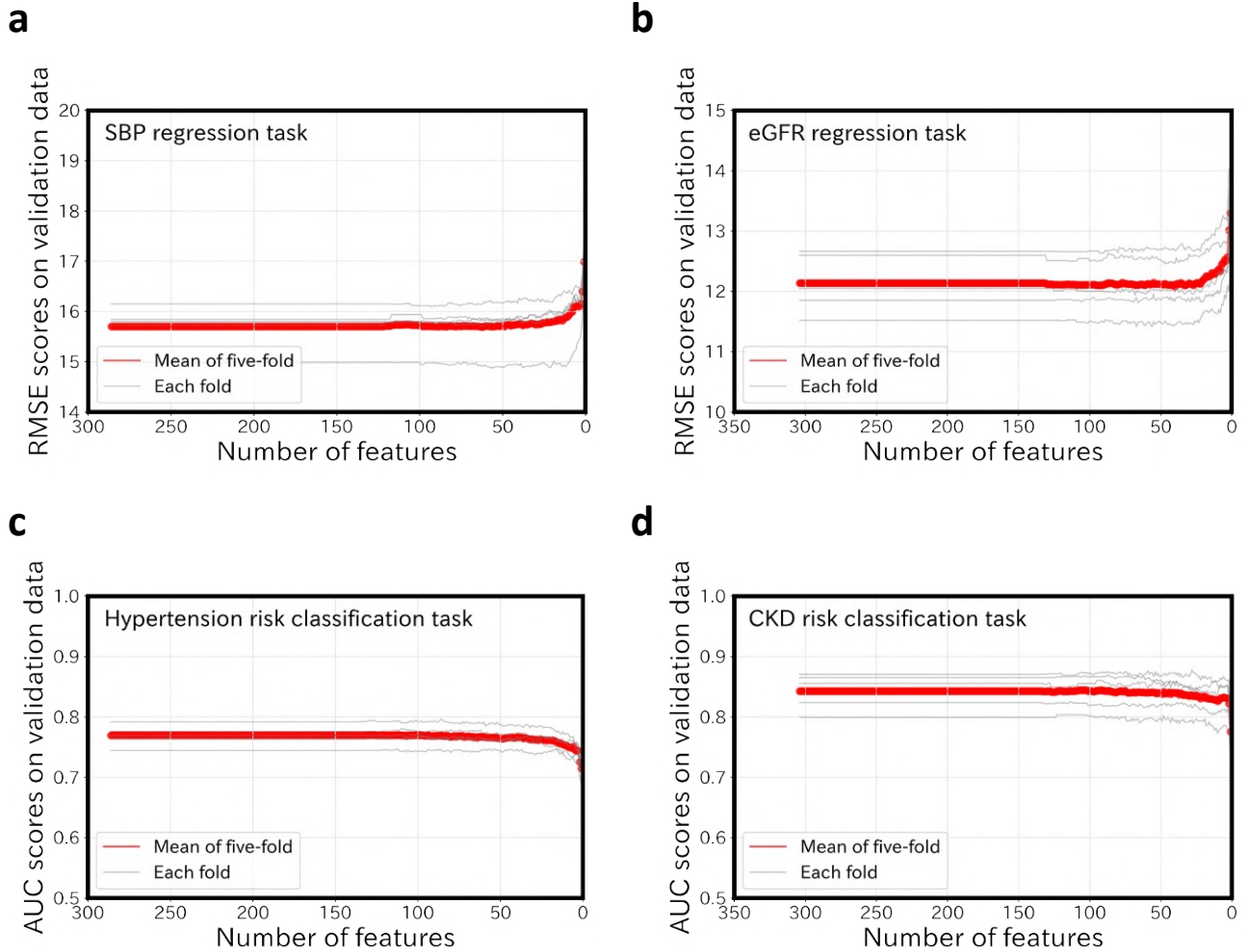
$$\boldsymbol{\mu}_3 = \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$



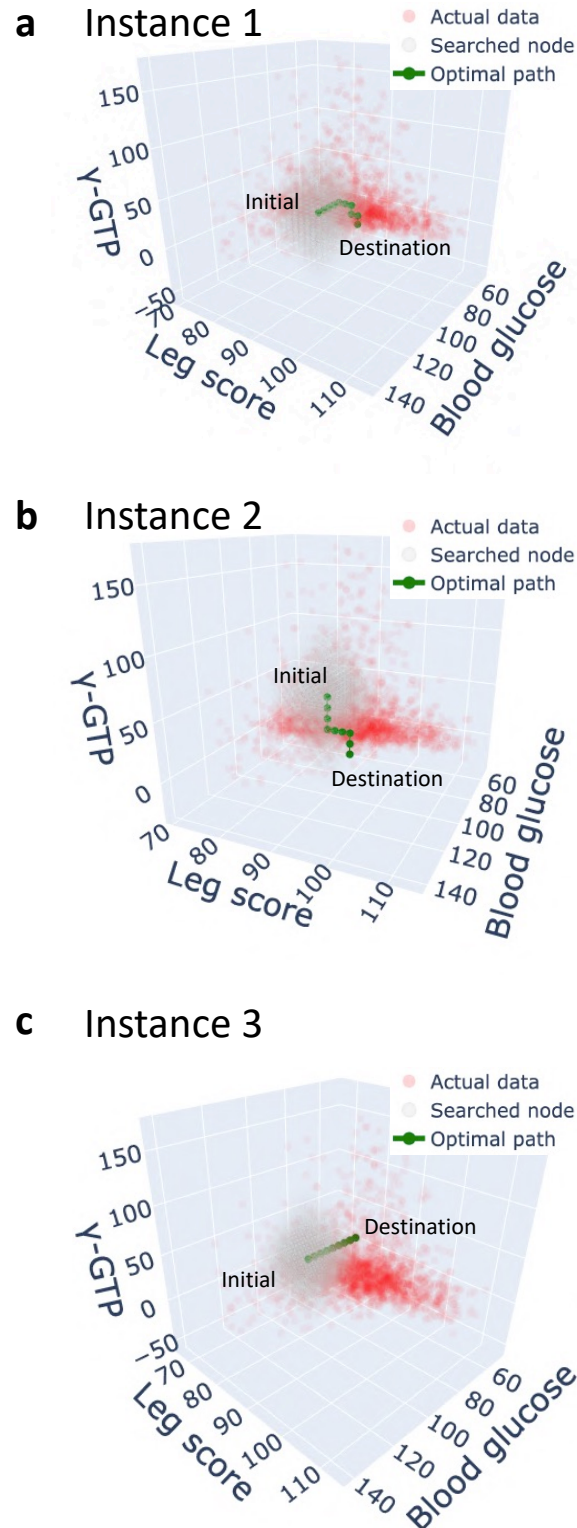
Supplementary Figure 2. Generation of three-dimensional (3D) synthetic dataset. The three 3D normal distributions generated 200 data points that consisted of x_1 , x_2 , and x_3 . Subsequently, a response variable was set to the sum of x_1 , x_2 , and x_3 with Gaussian noise ($\sigma = 2$). The synthetic dataset consisted of a total of 600 data points with explanatory variables (X1, X2, and X3) and a response variable.



Supplementary Figure 3. Results of the proposed framework on three-dimensional (3D) synthetic dataset. a Plot for prediction vs. true response variable. **b** Widely applicable Bayesian information criterion (WBIC) values of the stochastic surrogate models with 1–8 mixture components. **c** Histogram of actionability scores at different instances. The unit cell size of the grid was set to 0.5σ in the training data for each explanatory variable. The path search algorithm was executed with $L = 20,000$ for each instance and acquired a path with the lowest predictive value. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



Supplementary Figure 4. XGBoost model scores during recursive feature elimination (RFE). Features are gradually reduced in RFE, and the scores on the validation data at each stage are shown. **a** Root-mean-squared error (RMSE) scores on systolic blood pressure (SBP) regression. **b** RMSE scores on estimated glomerular filtration rate (eGFR) regression. **c** Area under the curve (AUC) scores on hypertension risk classification. **d** AUC scores on CKD risk classification.

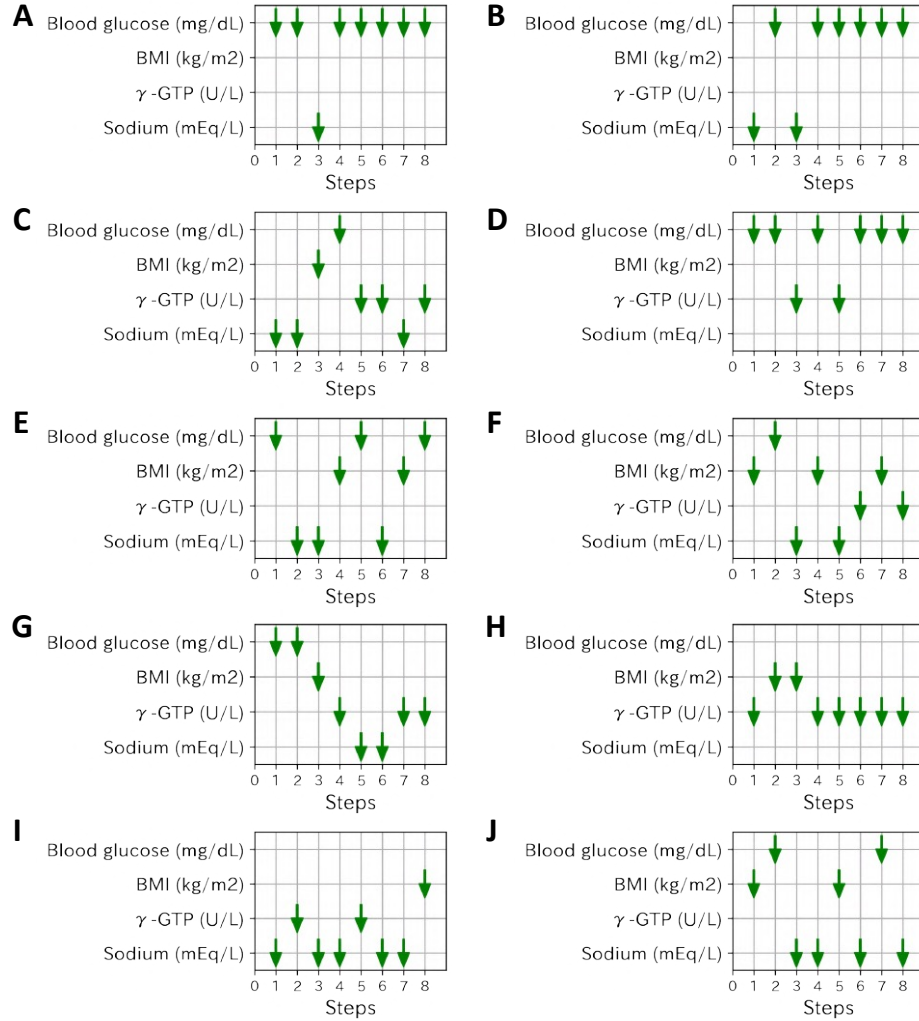


Supplementary Figure 5. 3D plots of optimal paths with intervention variables based on data-driven selection in systolic blood pressure (SBP) regression task. The instances respectively correspond to the instances in Fig. 6 of the main text.

ID: X

Values	Values	Unit of change	Values	Values
Age	Blood glucose	3.60	Erythrocyte count	IgM
Sex	BMI	0.67	MCH	Leg score
Height	γ -GTP	9.47	Lympho%	Inner fat level
Weight	Sodium	0.36	AST_GOT	Left arm R 5kHz
BMI			ALT_GPT	Left arm R 50kHz
Waist			Total protein	Left half X 50kHz
Periumbilical			Potassium	PEF_TIME
SBP			Inorganic Phosphorus	
DBP				
History of hypertension				

Initial predictive SBP: 141 mmHg



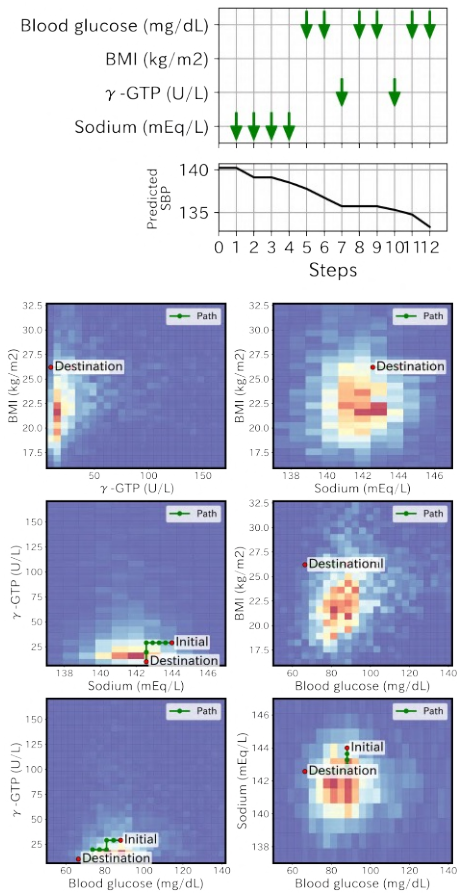
Supplementary Figure 6. Example question sheet for comparison with clinician intervention. An example of an evaluation sheet for the SBP regression task is provided above. The order of interventions and the initial values of measurement items of the instance (masked in this paper to protect personal health information) were presented to clinicians. Clinicians were asked to select an improvement path considered to be the closest to their treatment policy among ten blinded improvement paths (A–J), which comprised of a path planned by our framework and nine random paths. In the transition step, the up or down arrow represents an increase or decrease in the unit change of the intervention variable, respectively. Each clinician evaluated the same ten instances.

ID: Y

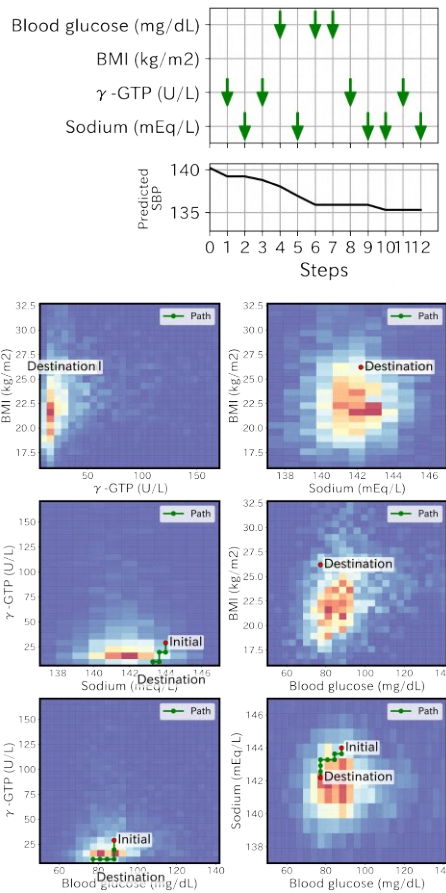
Values	Values	Unit of change	Values	Values
Age	Blood glucose	3.60	Erythrocyte count	IgM
Sex	BMI	0.67	MCH	Leg score
Height	γ -GTP	9.47	Lympho%	Inner fat level
Weight	Sodium	0.36	AST_GOT	Left arm R 5kHz
BMI			ALT_GPT	Left arm R 50kHz
Waist			Total protein	Left half X 50kHz
Periumbilical			Potassium	PEF_TIME
SBP			Inorganic Phosphorus	
DBP				
History of hypertension				

Initial predictive SBP: 140 mmHg

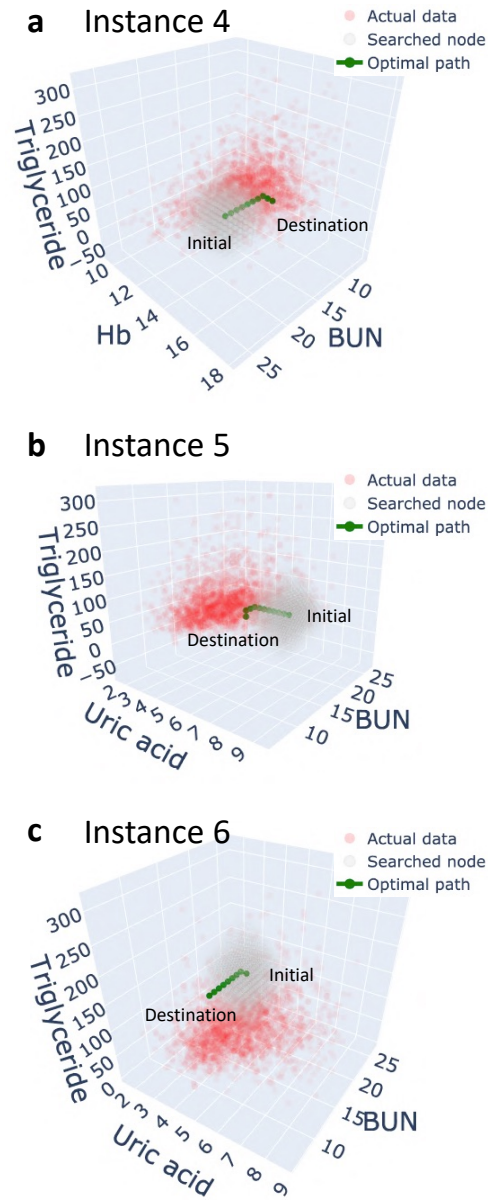
A



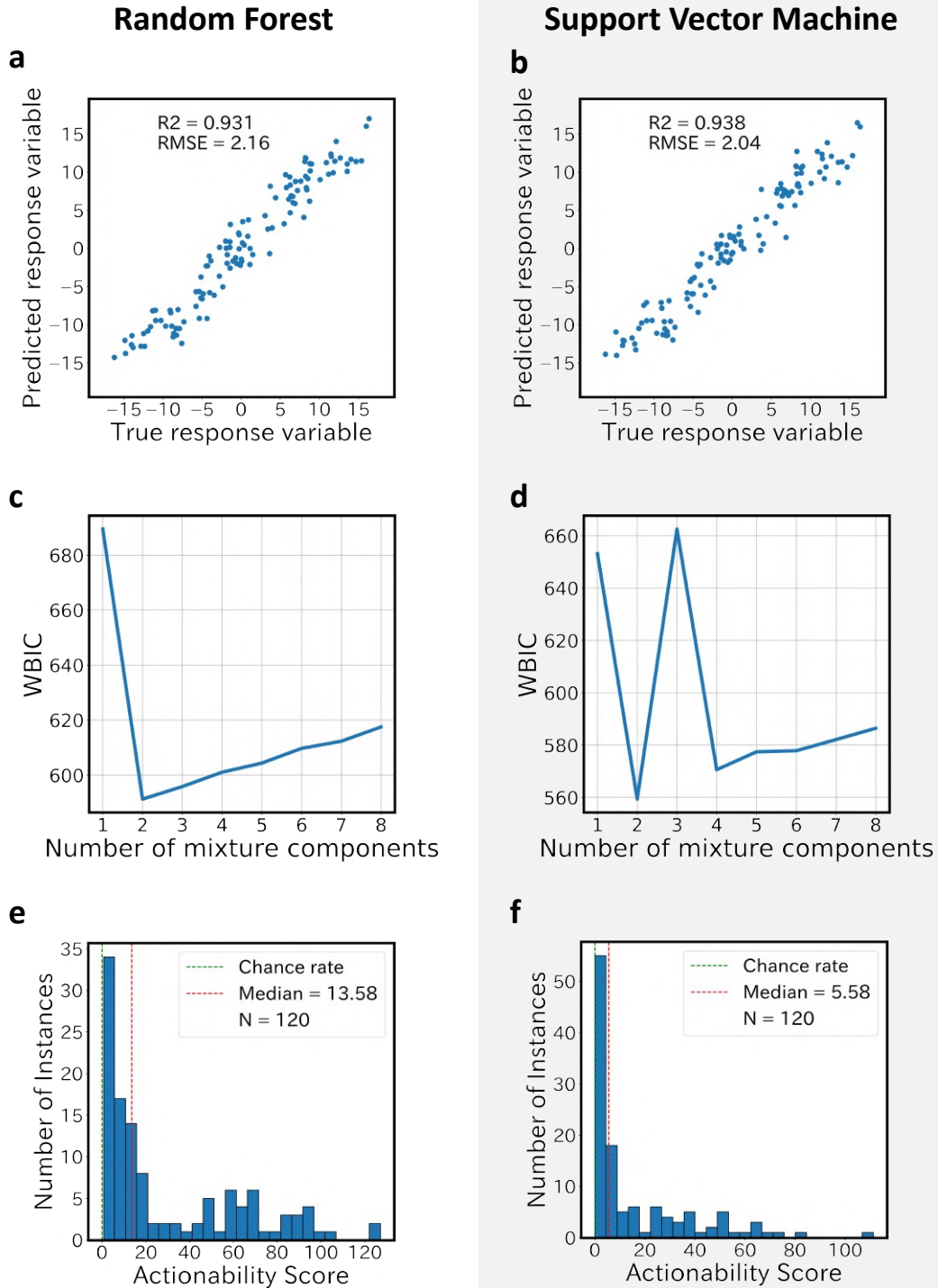
B



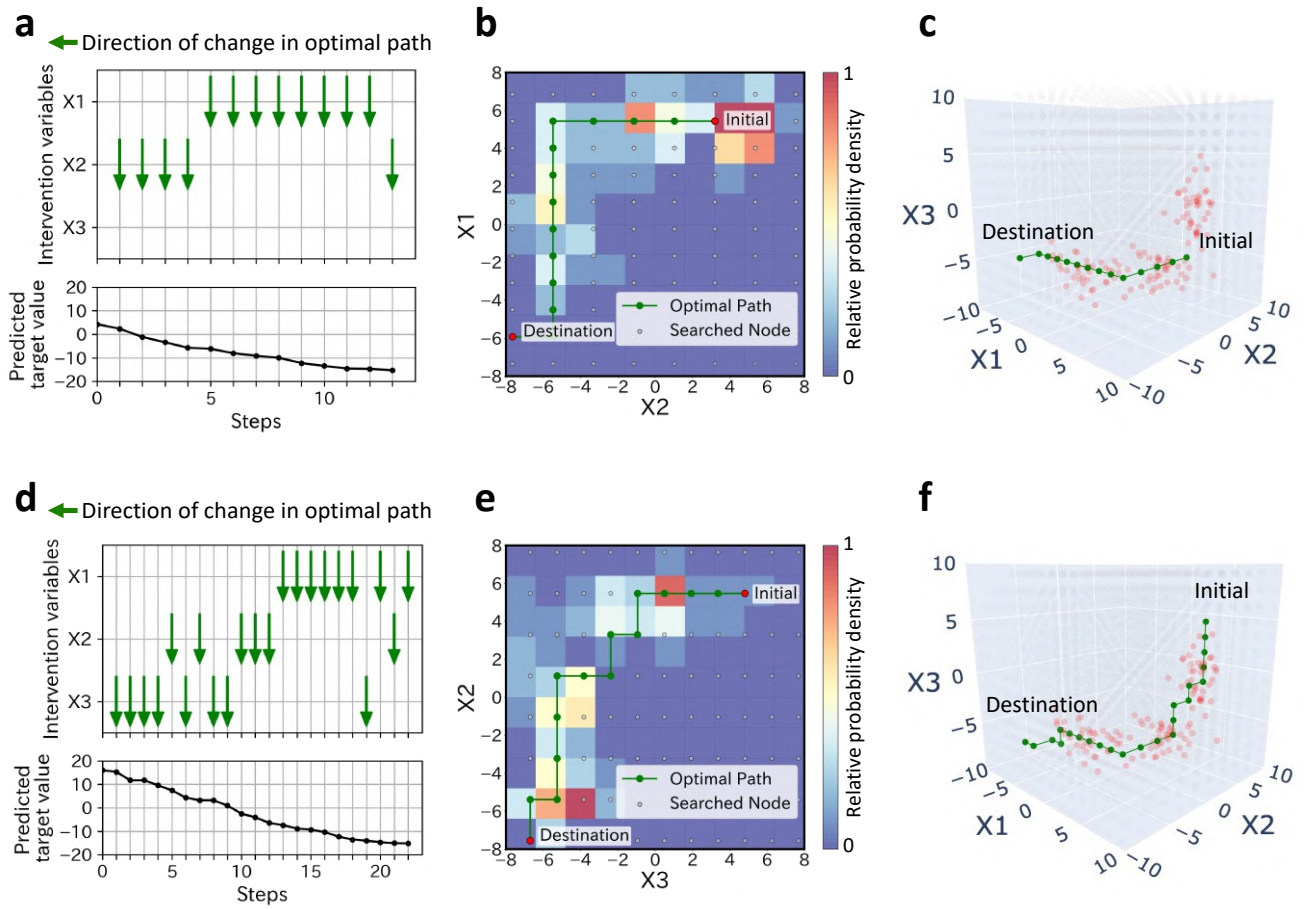
Supplementary Figure 7. Example question sheet for utility assessment. An example of an evaluation sheet for the SBP regression task is provided above. The clinicians were provided with the initial values of measurement items of the instance (masked in this paper to protect personal health information), as well as two blinded paths (A and B). The blinded paths are a path planned by our framework and a random path. For each path, the order of interventions along with the accompanied transition of the predictive response variable and the projection on 2D heatmaps for each pair of intervention variables were presented. In the transition step, the up or down arrow represents an increase or decrease in the unit change of the intervention variable, respectively. Clinicians evaluated the practicality and informativeness of each path with yes/no. Clinicians also commented on the reasons for choosing each evaluation. Each clinician evaluated the same ten instances.



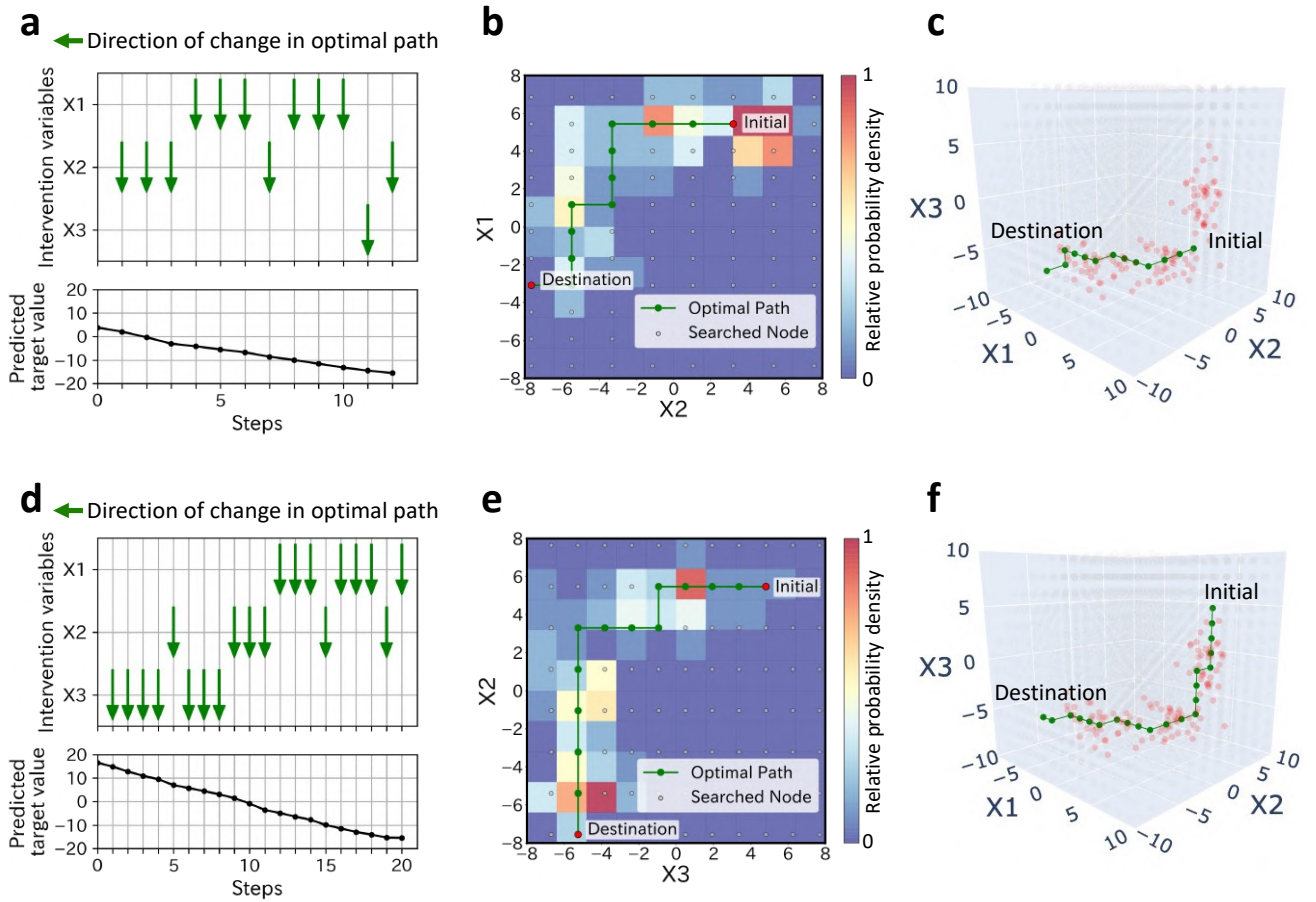
Supplementary Figure 8. 3D plots of optimal paths with intervention variables based on data-driven selection in chronic kidney disease (CKD) risk classification task. The instances respectively correspond to the instances in Fig. 8 of the main text.



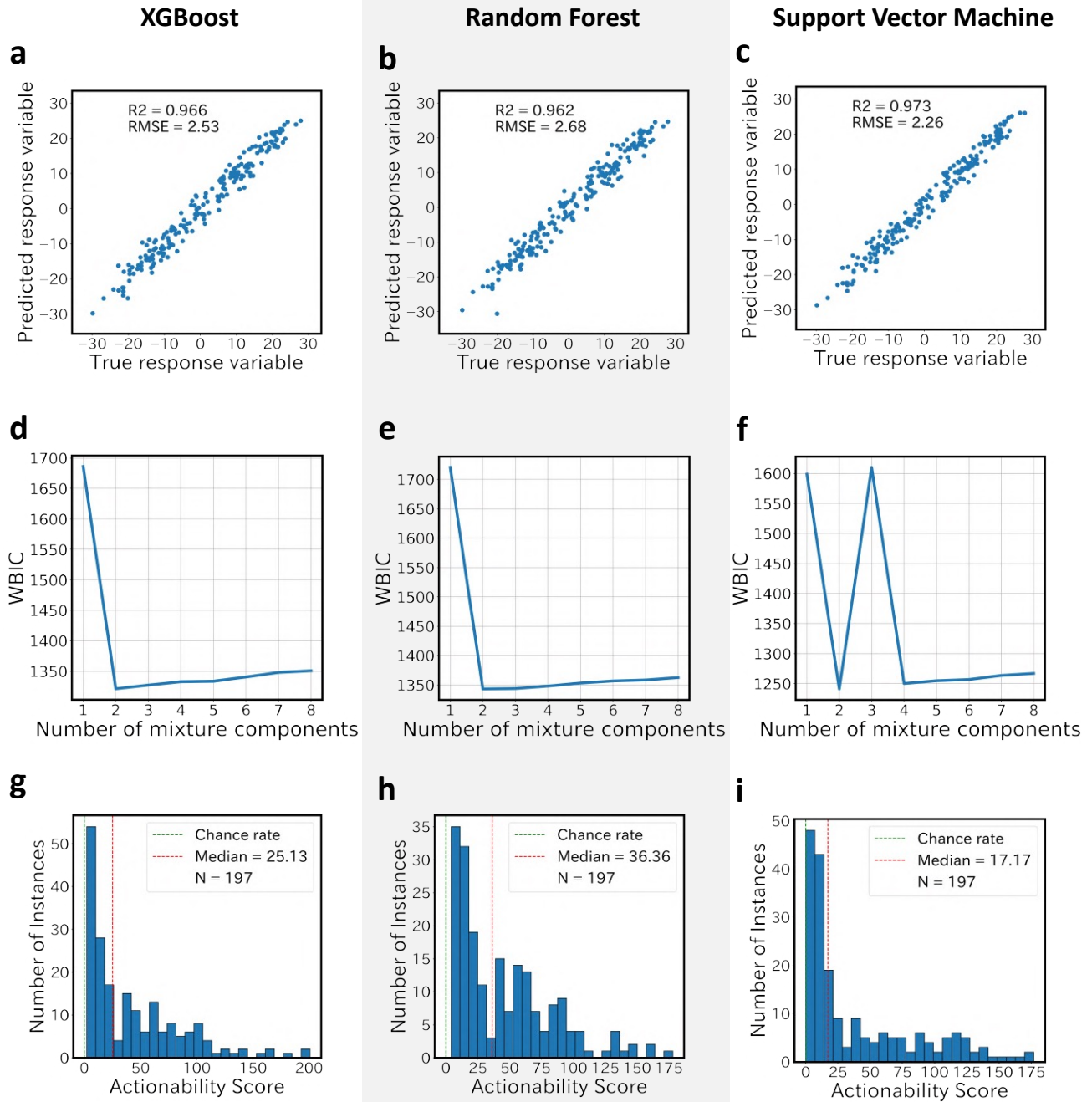
Supplementary Figure 9. Results of the proposed framework on three-dimensional (3D) synthetic dataset using different machine learning (ML) algorithms. The results of applying our framework to prediction models based on nonlinear ML algorithms other than XGBoost are shown: random forest (**a**, **c**, **e**) and support vector machine (**b**, **d**, **f**). **a**, **b** Plots for prediction vs. true response variable. **c**, **d** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **e**, **f** Histogram of actionability scores with intervention variables based on data-driven selection at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



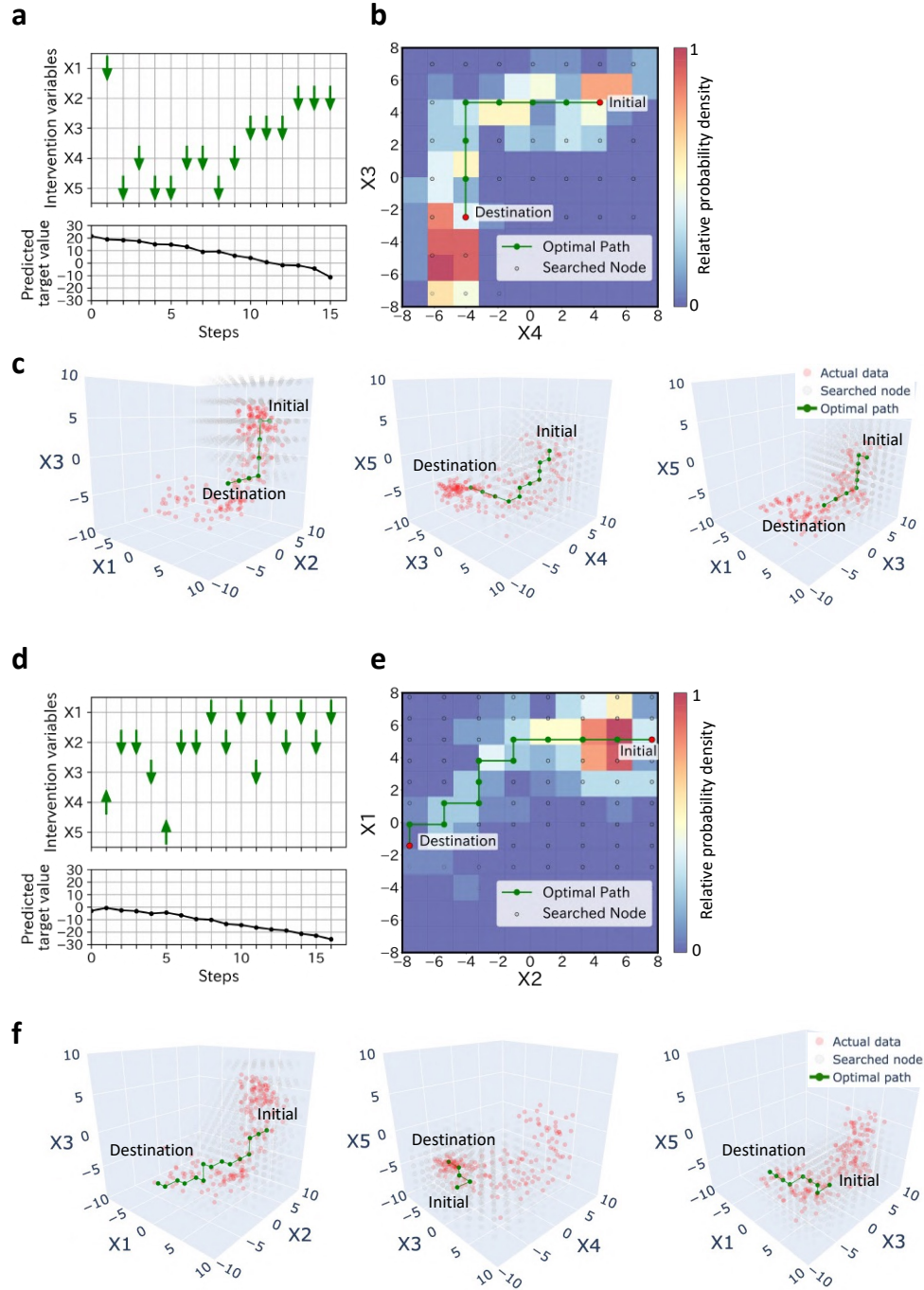
Supplementary Figure 10. Examples of actionable paths planned using random forest (RF) model on three-dimensional (3D) synthetic dataset. The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance A (**a–c**) and instance B (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, e** 2D plots of the path. The 2D plots are shown regarding the selected two variables: X1 and X2 (**b**), and X2 and X3 (**e**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** 3D plots of the path.



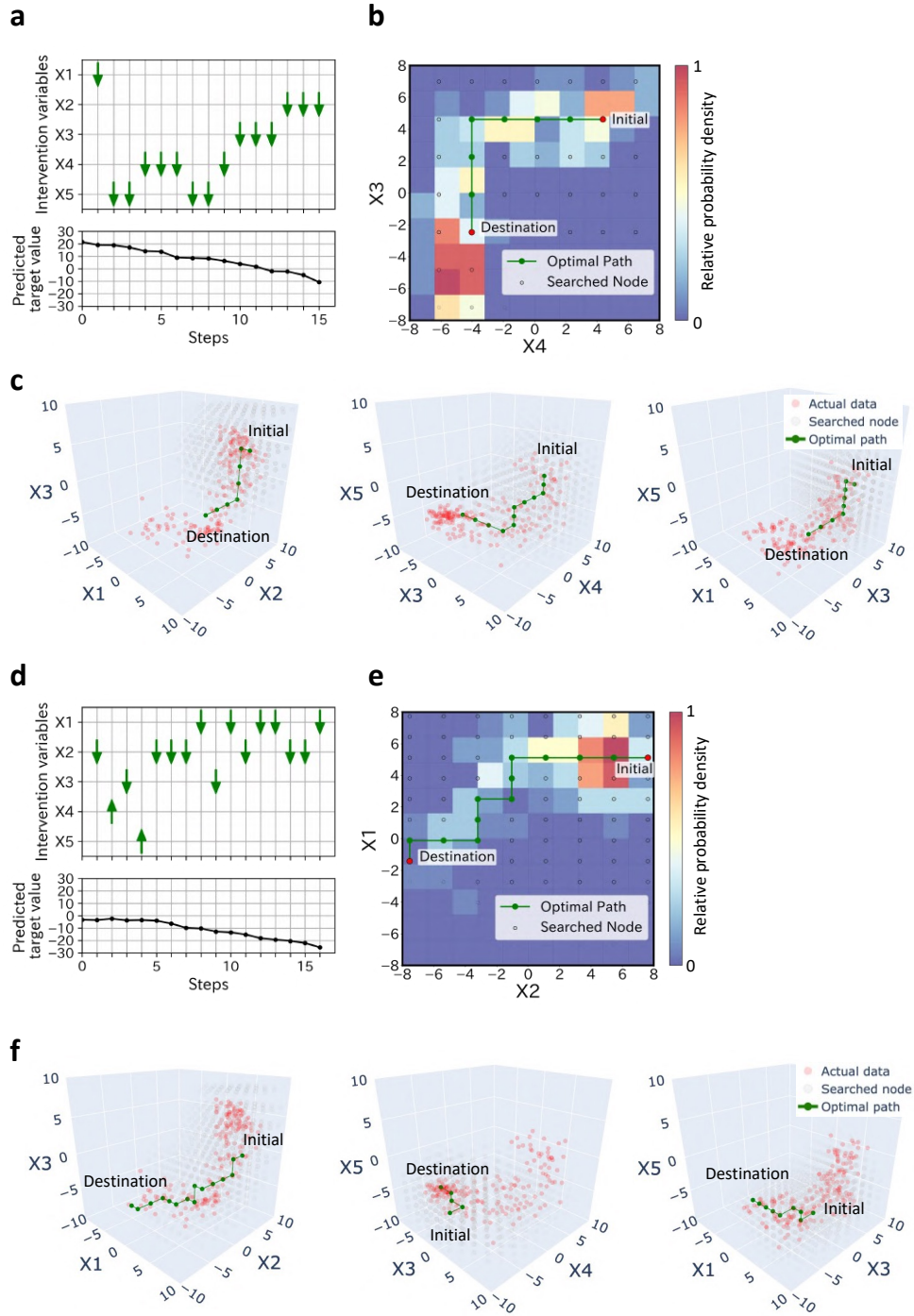
Supplementary Figure 11. Examples of actionable paths planned using support vector machine (SVM) model on three-dimensional (3D) synthetic dataset. The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance A (**a–c**) and instance B (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, e** 2D plots of the path. The 2D plots are shown regarding the selected two variables: X1 and X2 (**b**), and X2 and X3 (**e**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** 3D plots of the path.



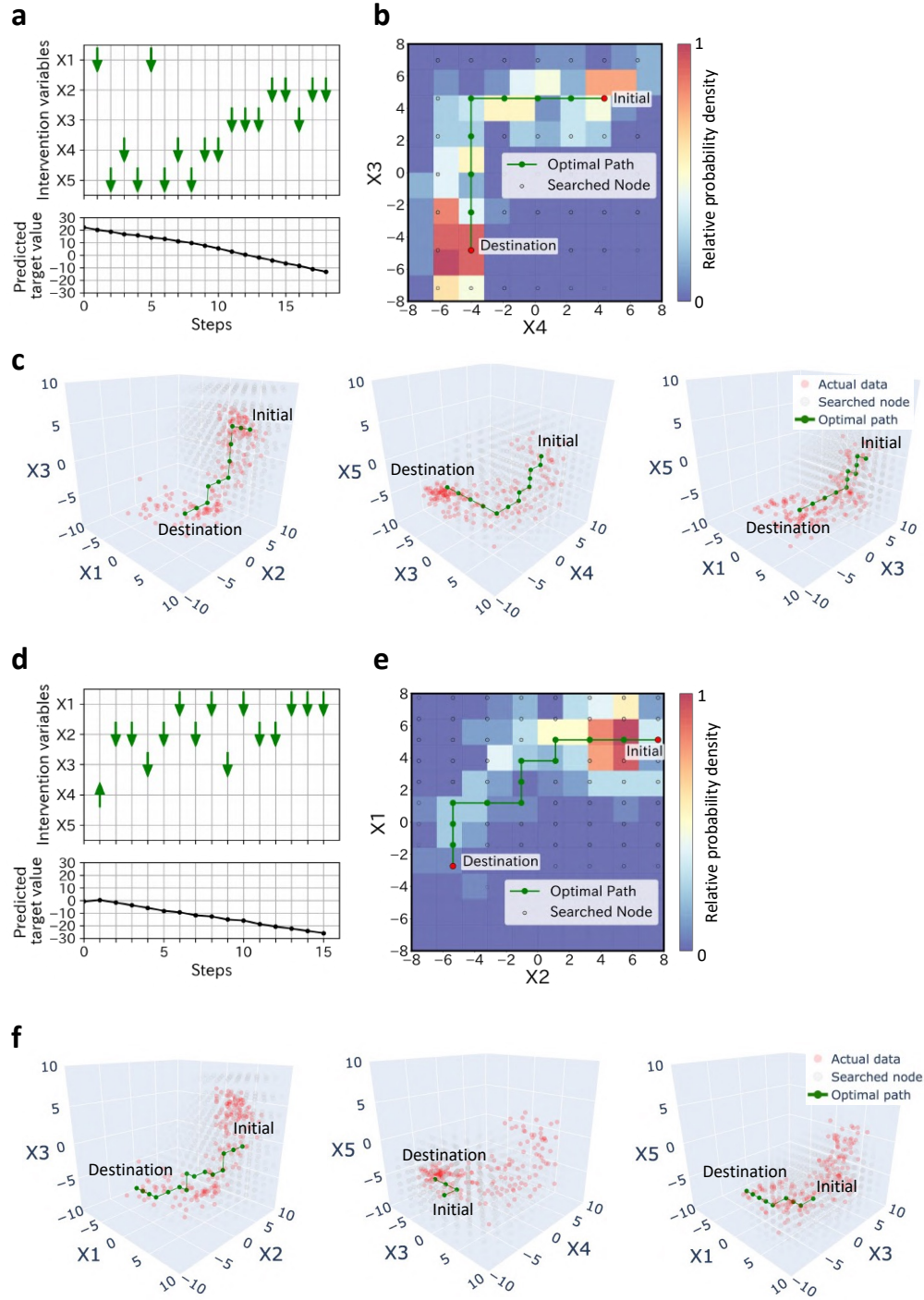
Supplementary Figure 12. Results of the proposed framework on five-dimensional (5D) synthetic dataset. The results of applying our framework to prediction models based on nonlinear machine learning (ML) algorithms are shown: XGBoost (**a, d, g**), random forest (**b, e, h**), and support vector machine (**c, f, i**). **a–c** Plots for prediction vs. true response variable. **d–f** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **g–i** Histogram of actionability scores with intervention variables based on data-driven selection at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



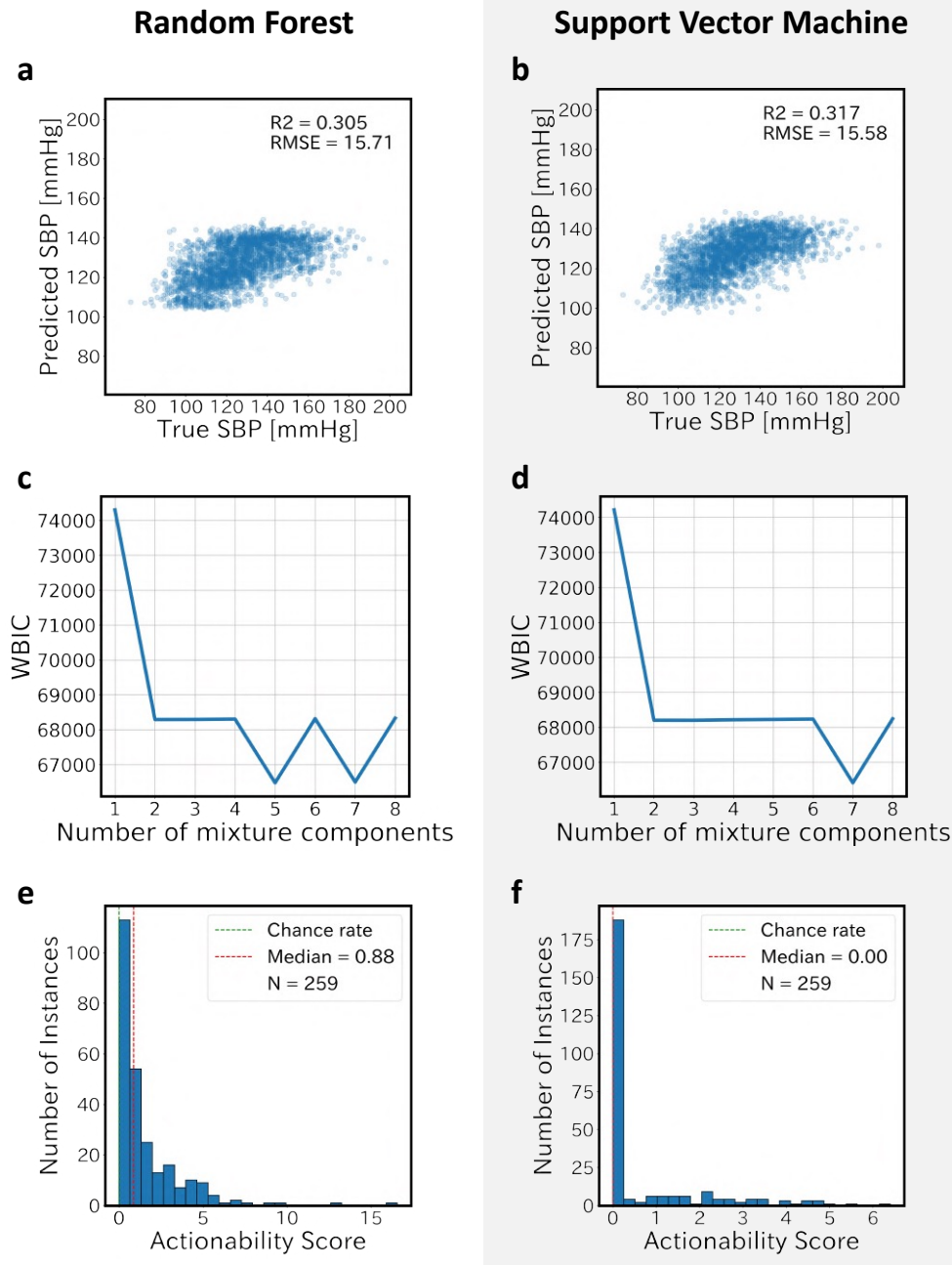
Supplementary Figure 13. Examples of actionable paths planned using XGBoost model on five-dimensional (5D) synthetic dataset. The optimal paths for improving the response variable predicted by the XGBoost model are represented for randomly selected two examples: instance C (a–c) and instance D (d–f). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, e** 2D plots of the path. The 2D plots are shown regarding the selected two variables: X3 and X4 (b), and X1 and X2 (e). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** 3D plots of the path.



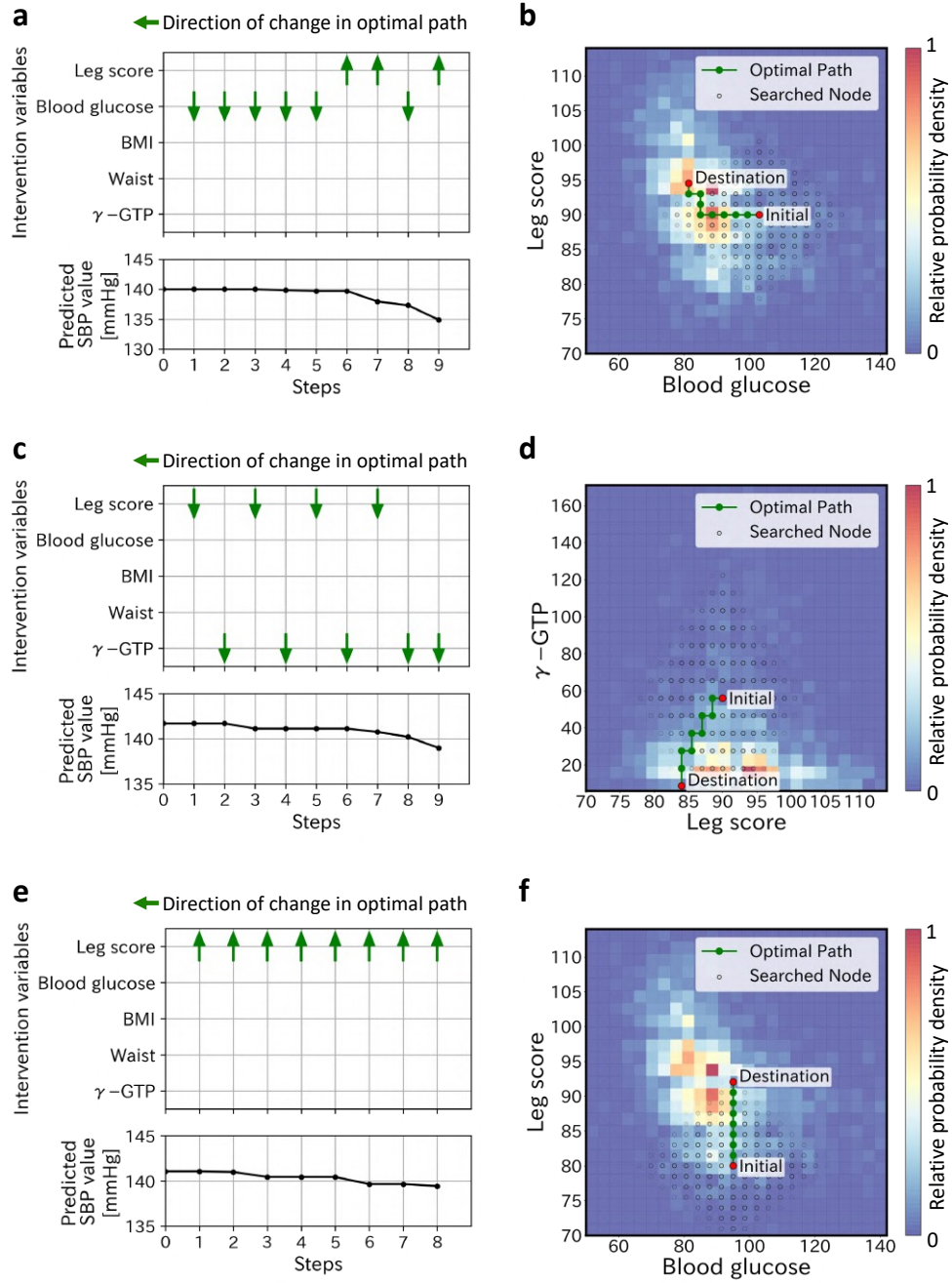
Supplementary Figure 14. Examples of actionable paths planned using random forest (RF) model on five-dimensional (5D) synthetic dataset. The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance C (**a–c**) and instance D (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, e** 2D plots of the path. The 2D plots are shown regarding the selected two variables: X3 and X4 (**b**), and X1 and X2 (**e**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** 3D plots of the path.



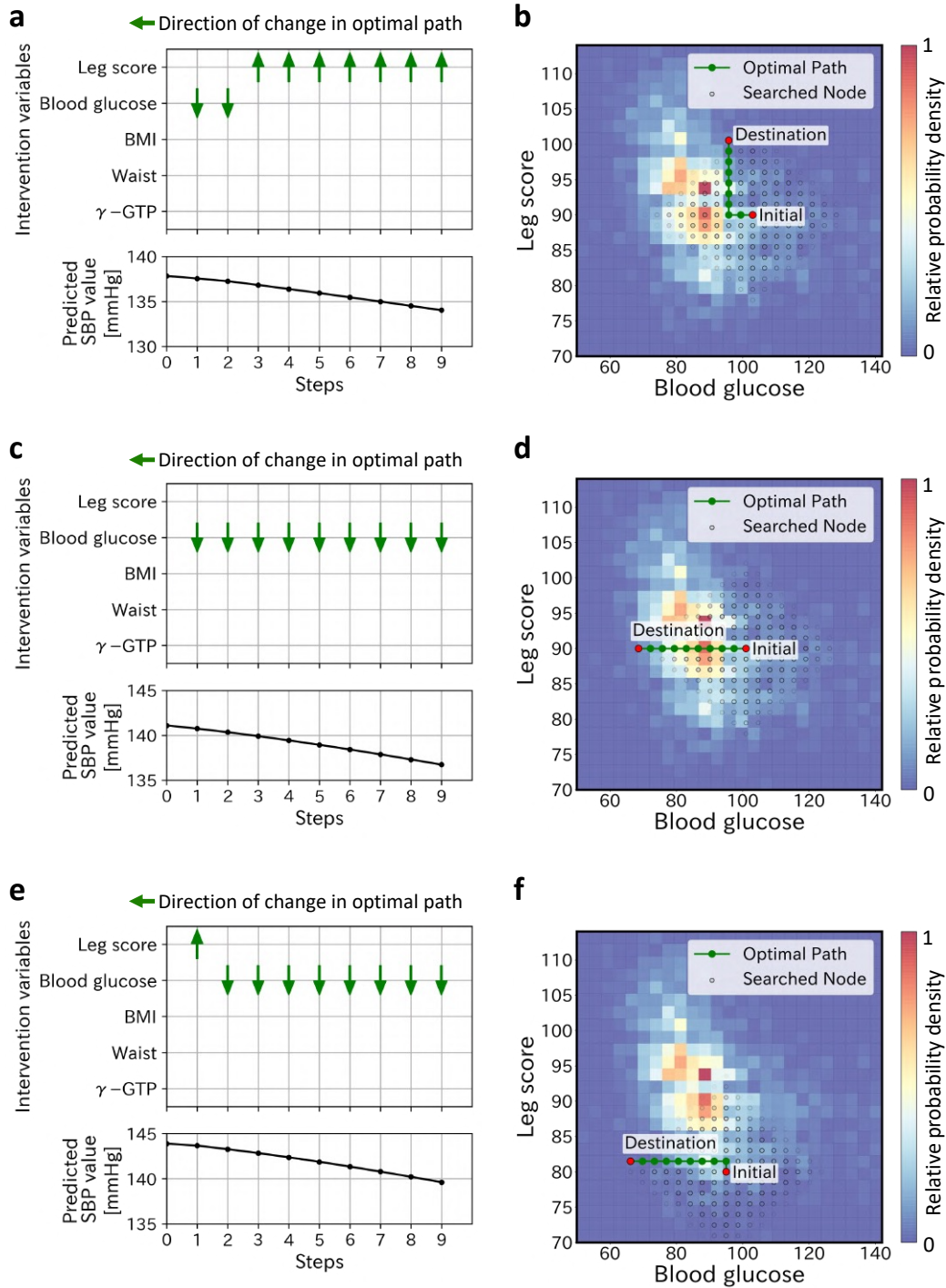
Supplementary Figure 15. Examples of actionable paths planned using support vector machine (SVM) model on five-dimensional (5D) synthetic dataset. The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance C (**a–c**) and instance D (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **B, e** 2D plots of the path. The 2D plots are shown regarding the selected two variables: X3 and X4 (**b**), and X1 and X2 (**e**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed. **c, f** 3D plots of the path.



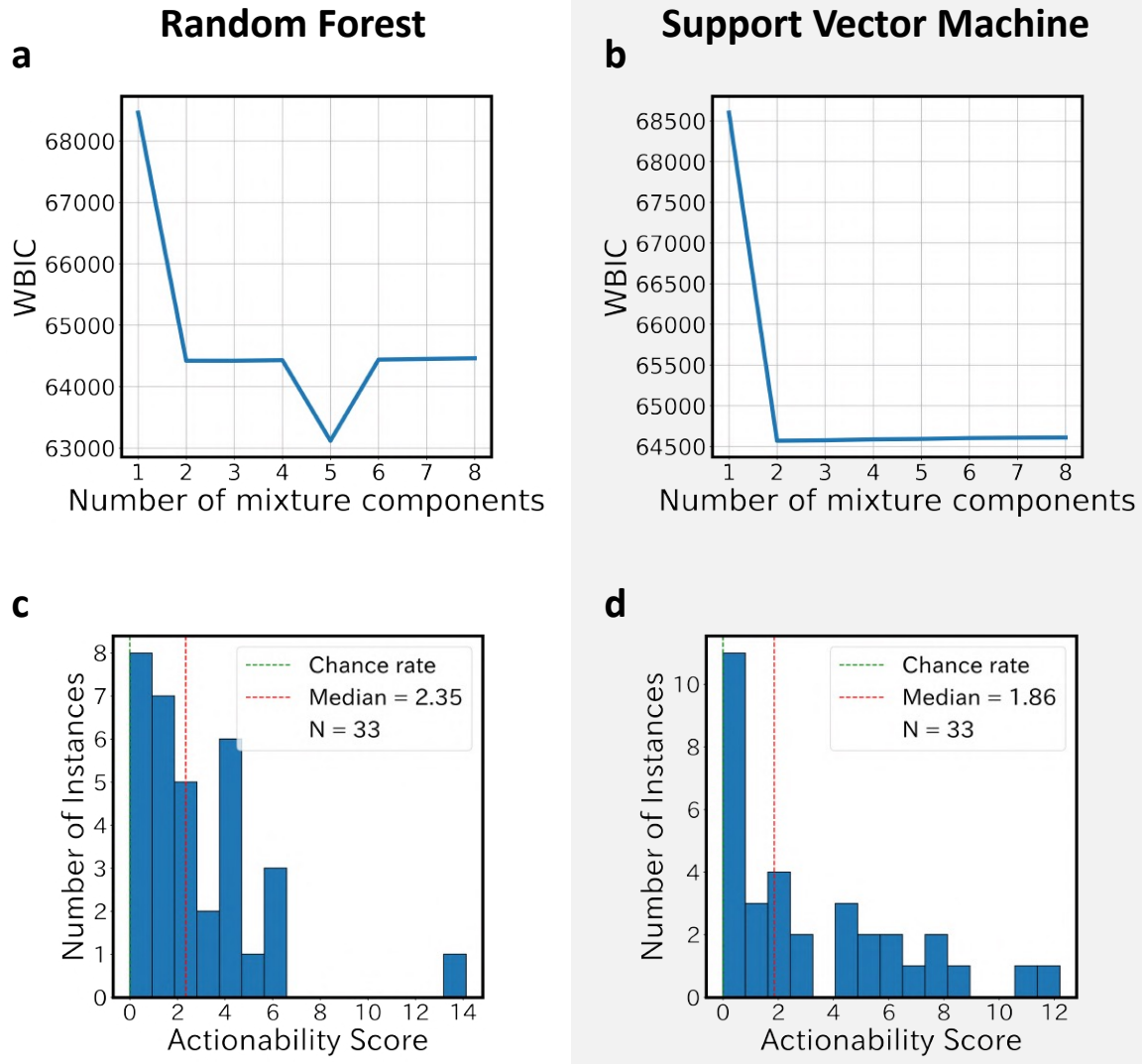
Supplementary Figure 16. Application of proposed framework on systolic blood pressure (SBP) regression task using different machine learning (ML) algorithms. The results of applying our framework to prediction models based on nonlinear ML algorithms other than XGBoost are shown: random forest (**a**, **c**, **e**) and support vector machine (**b**, **d**, **f**). **a**, **b** Plots for prediction vs. true response variable. **c**, **d** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **e**, **f** Histogram of actionability scores with intervention variables based on data-driven selection at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



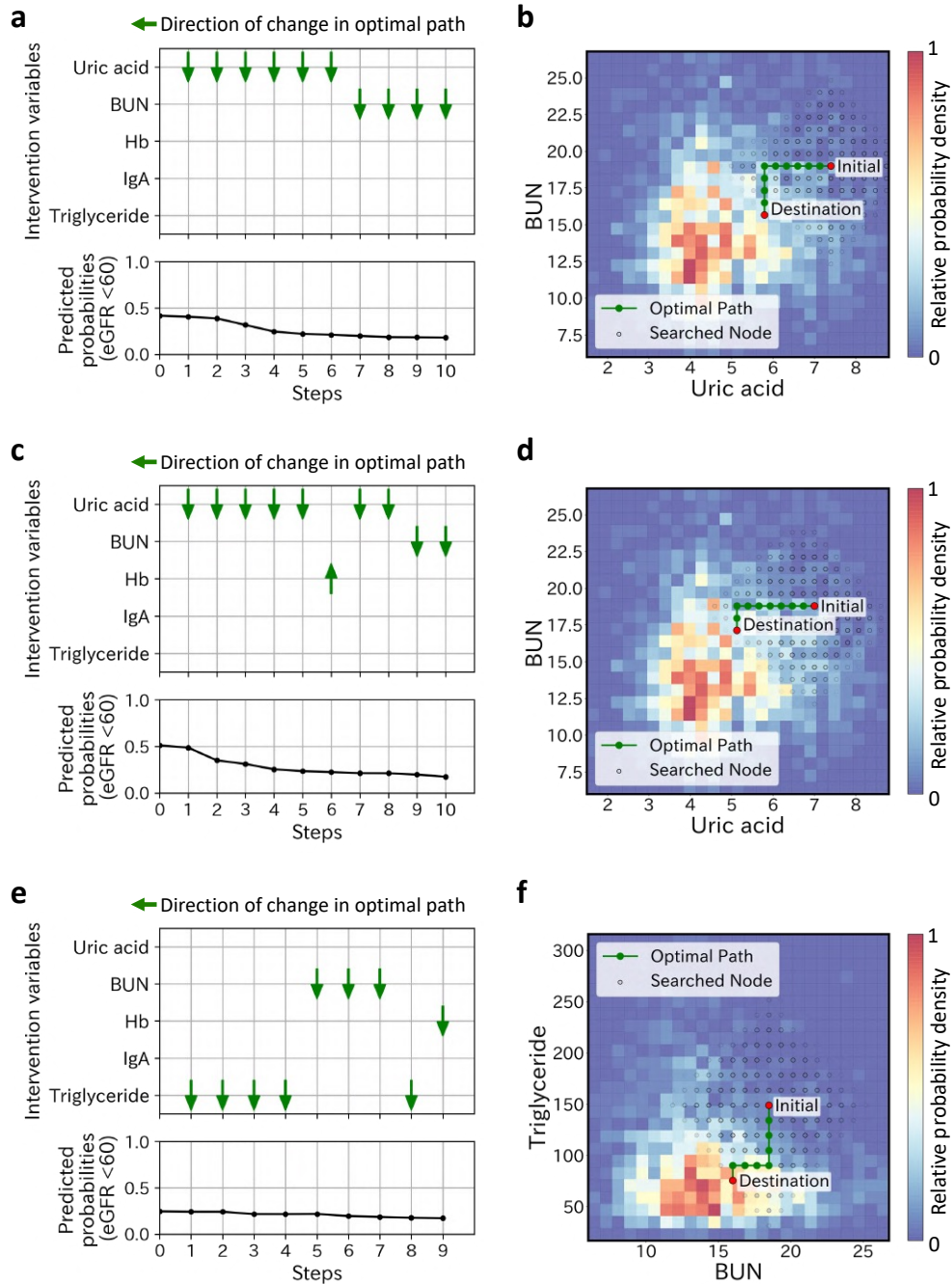
Supplementary Figure 17. Examples of personal actionable paths for treatment in systolic blood pressure (SBP) regression task using random forest (RF). The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance 1 (**a, b**), instance 2 (**c, d**), and instance 3 (**e, f**). **a, c, e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d, f** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: blood glucose and leg score (**b**), leg score and γ -GTP (**d**), and blood glucose and leg score (**f**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



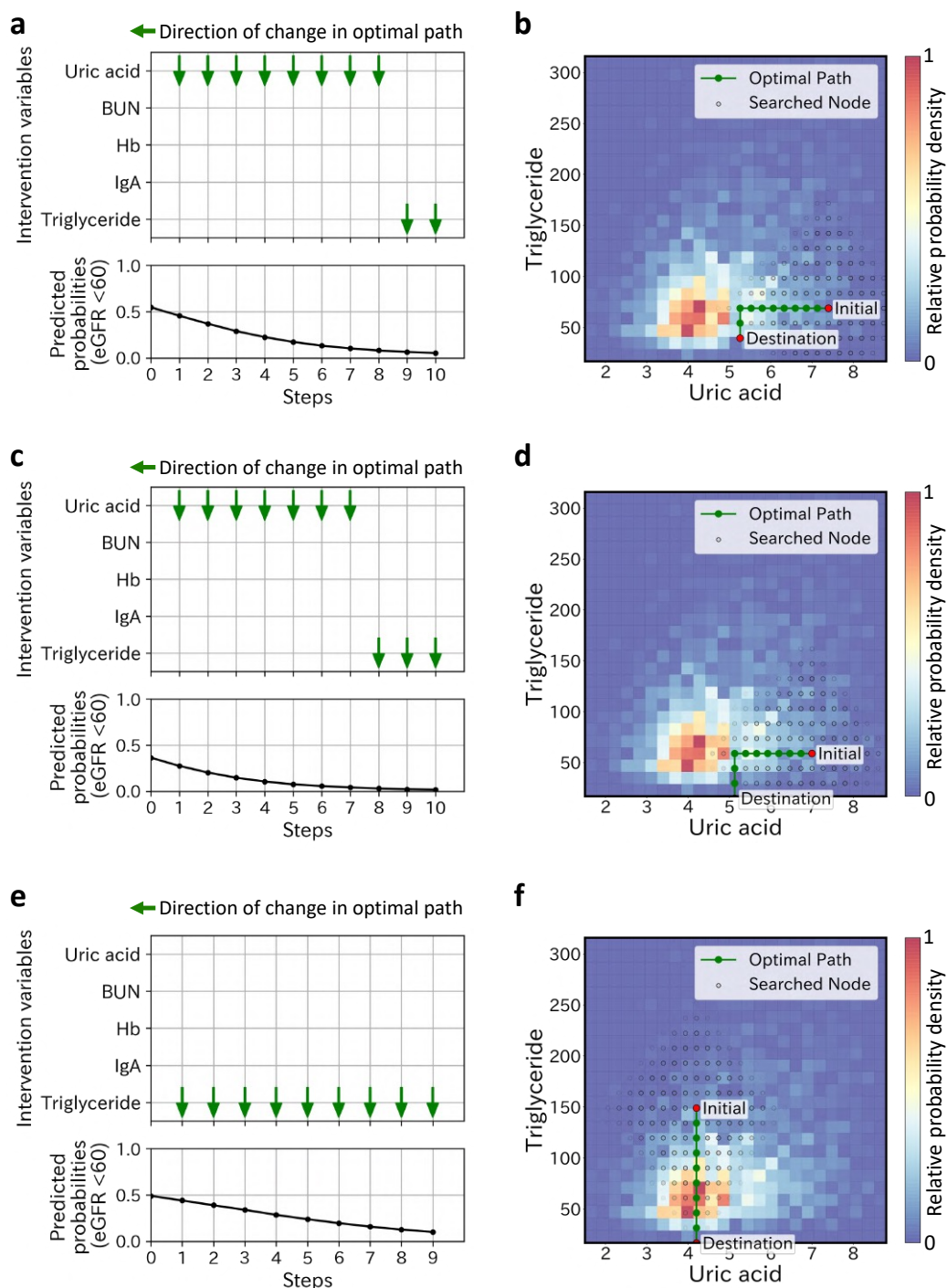
Supplementary Figure 18. Examples of personal actionable paths for treatment in systolic blood pressure (SBP) regression task using support vector machine (SVM). The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance 1 (**a, b**), instance 2 (**c, d**), and instance 3 (**e, f**). **a, c, e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d, f** 2D plots of the path regarding blood glucose and leg score. In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



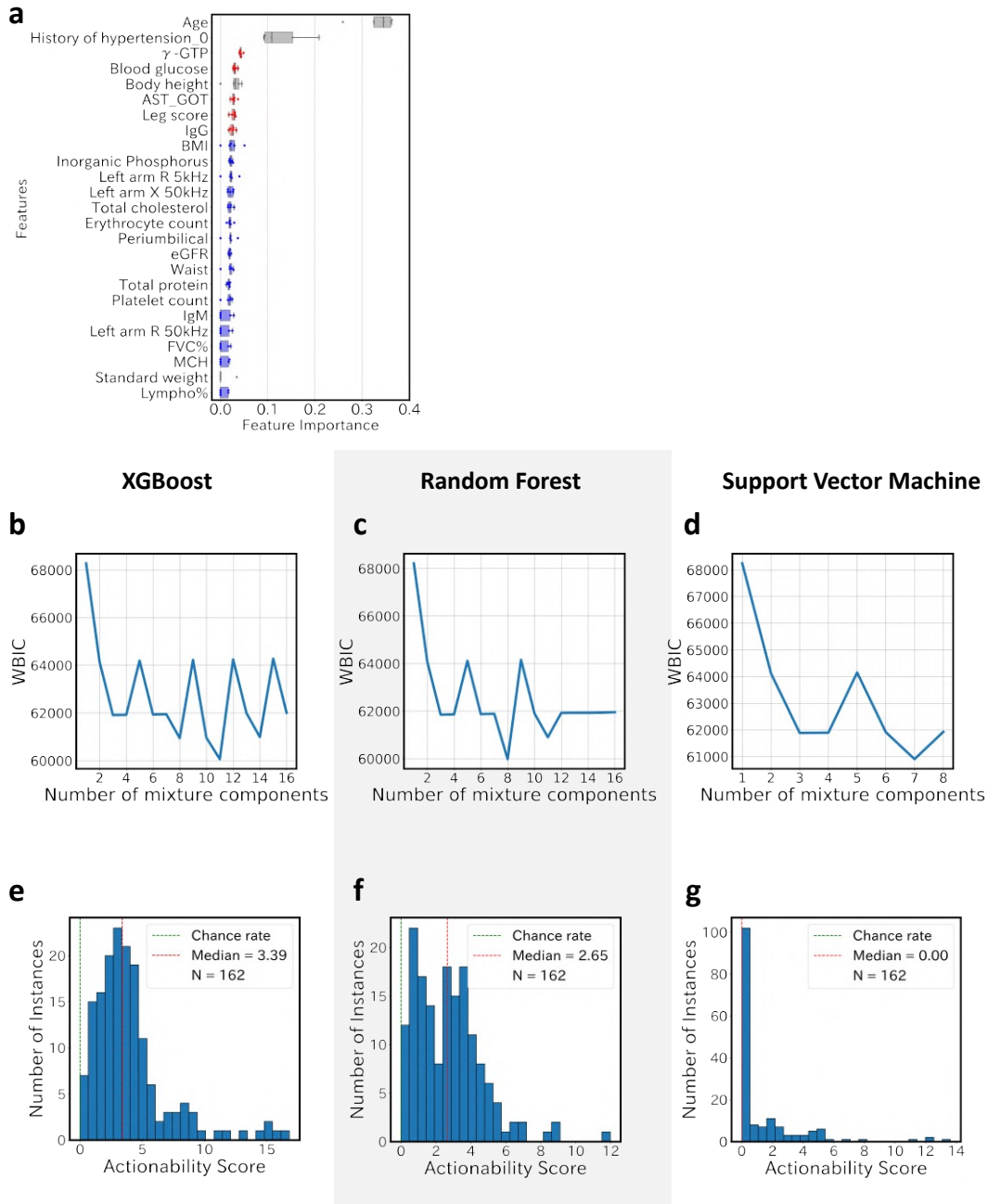
Supplementary Figure 19. Application of proposed framework on chronic kidney disease (CKD) risk classification task using different machine learning (ML) algorithms. The results of applying our framework to prediction models based on nonlinear ML algorithms other than XGBoost are shown: random forest (**a, c**) and support vector machine (**b, d**). **a, b** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **c, d** Histogram of actionability scores with intervention variables based on data-driven selection at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



Supplementary Figure 20. Examples of personal actionable paths for treatment in chronic kidney disease (CKD) risk classification task using random forest (RF). The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance 4 (**a, b**), instance 5 (**c, d**), and instance 6 (**e, f**). **a, c, e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d, f** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: uric acid and blood urea nitrogen (BUN) (**b, d**), and triglyceride and BUN (**f**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.

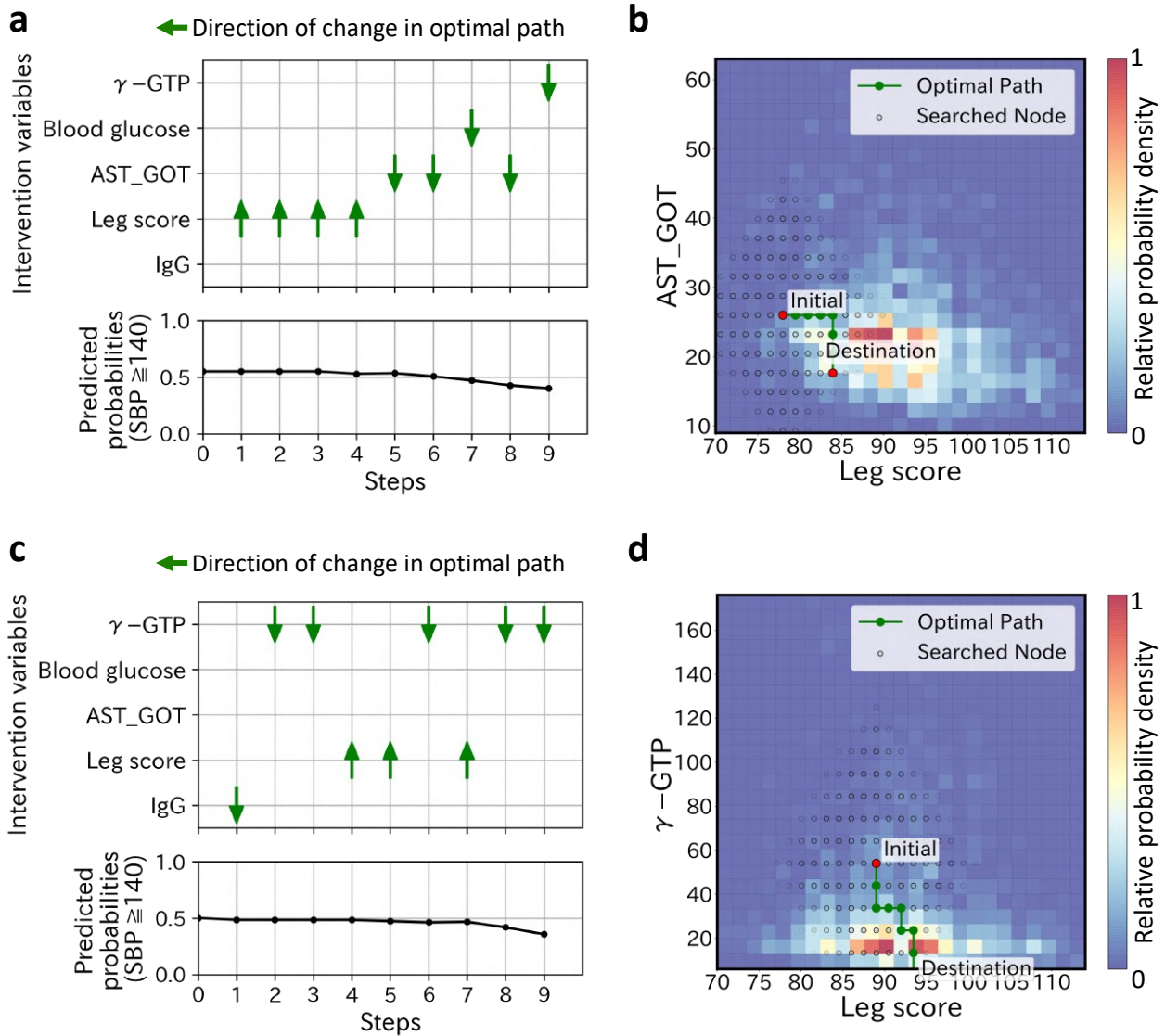


Supplementary Figure 21. Examples of personal actionable paths for treatment in chronic kidney disease (CKD) risk classification task using support vector machine (SVM). The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance 4 (**a**, **b**), instance 5 (**c**, **d**), and instance 6 (**e**, **f**). **a**, **c**, **e** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b**, **d**, **f** 2D plots of the path. The 2D plots are shown regarding triglyceride and uric acid. In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.

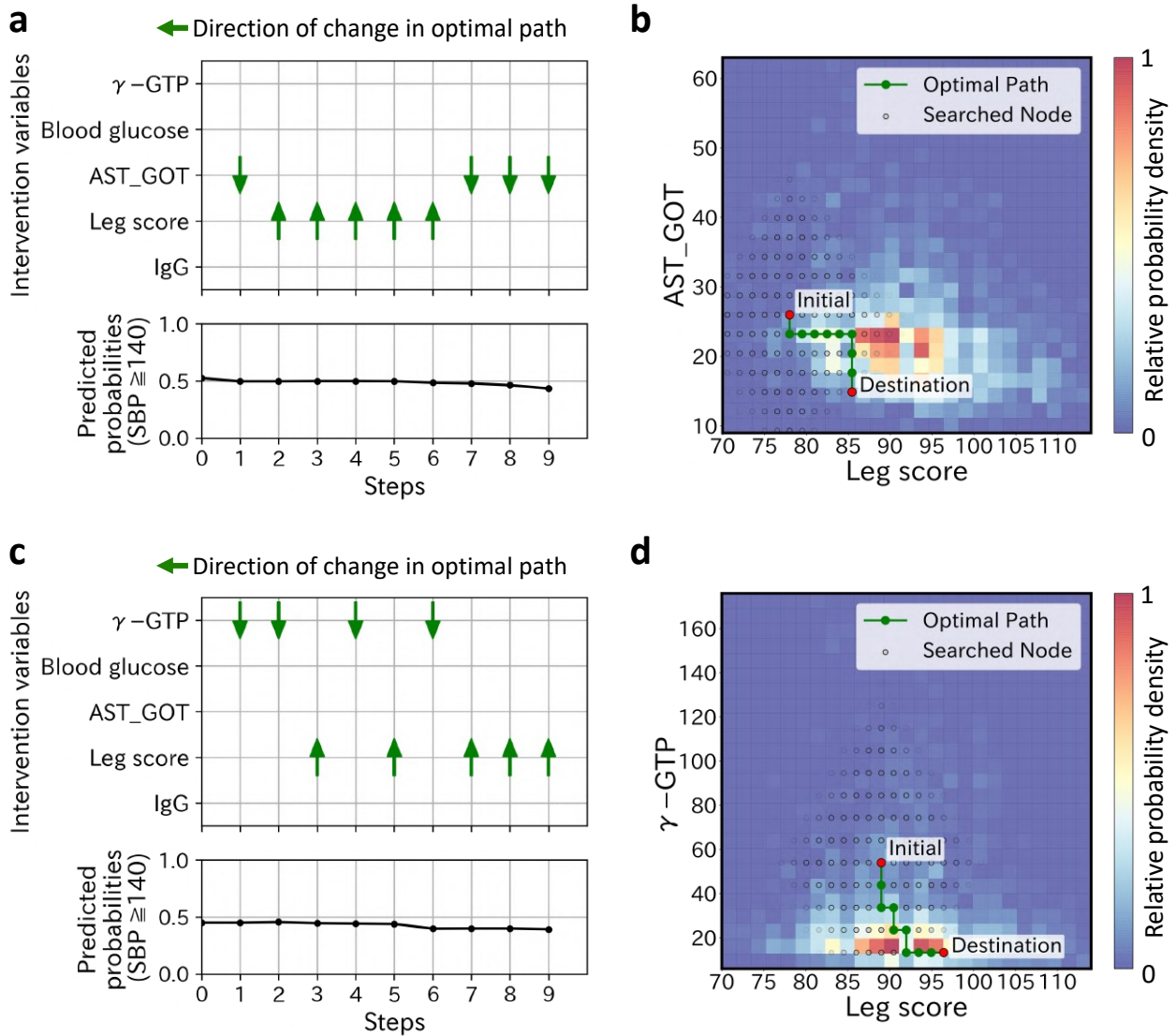


Supplementary Figure 22. Results of the proposed framework on hypertension risk classification task. a

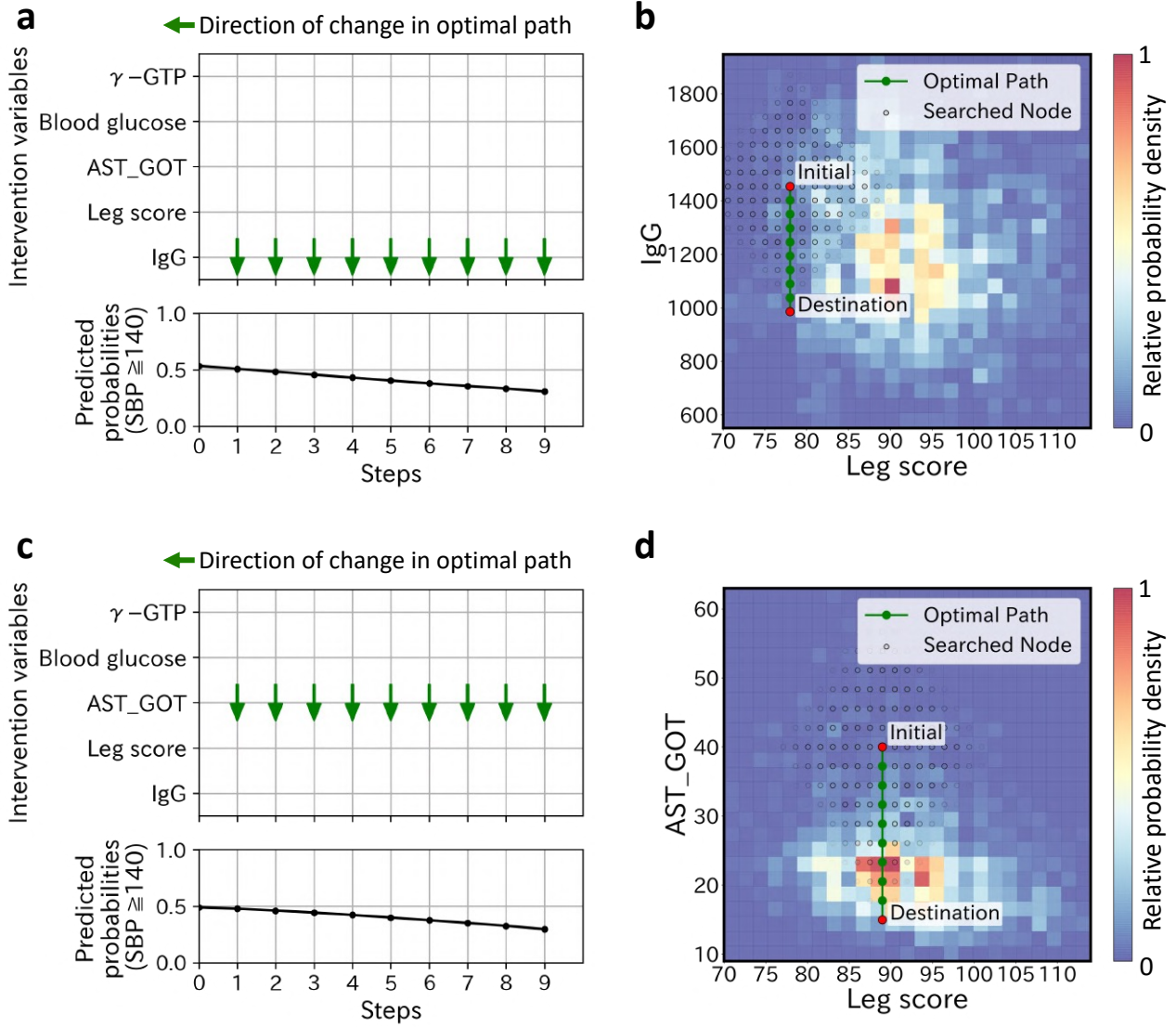
Feature importance. These 25 features were selected by recursive feature elimination (RFE). RFE was performed with five-fold cross-validation, and the feature importance when 25 variables remained is shown for each fold ($n = 5$). The plot color represents the following: red: intervention variables, gray: variables which cannot be intervened, and blue: other variables. In box-plot, center line represents median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Details of features are described in Supplementary Data 1. **b–g** The results of applying our framework to prediction models based on nonlinear machine learning (ML) algorithms: XGBoost (**b**, **e**), random forest (**c**, **f**), and support vector machine (**d**, **g**). **b–d** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models. **e–g** Histogram of actionability scores. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



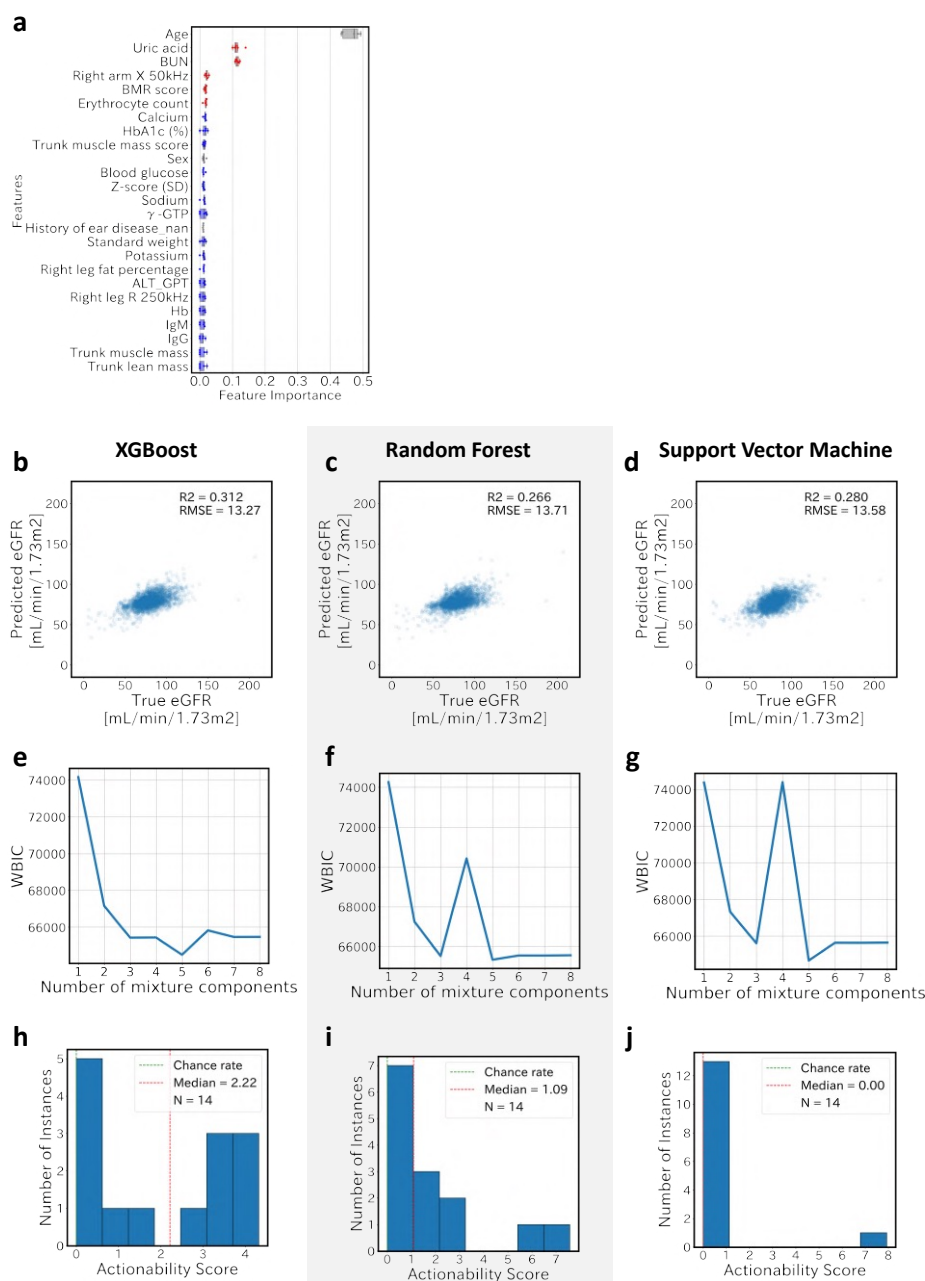
Supplementary Figure 23. Examples of personal actionable paths for treatment in hypertension risk classification task using XGBoost. The optimal paths for improving the response variable predicted by the XGBoost model are represented for randomly selected two examples: instance 7 (**a**, **b**) and instance 8 (**c**, **d**). **a**, **c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b**, **d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: aspartate transaminase (AST [GOT]) and leg score (**b**), and gamma glutamyl transferase (g-GTP) and leg score (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



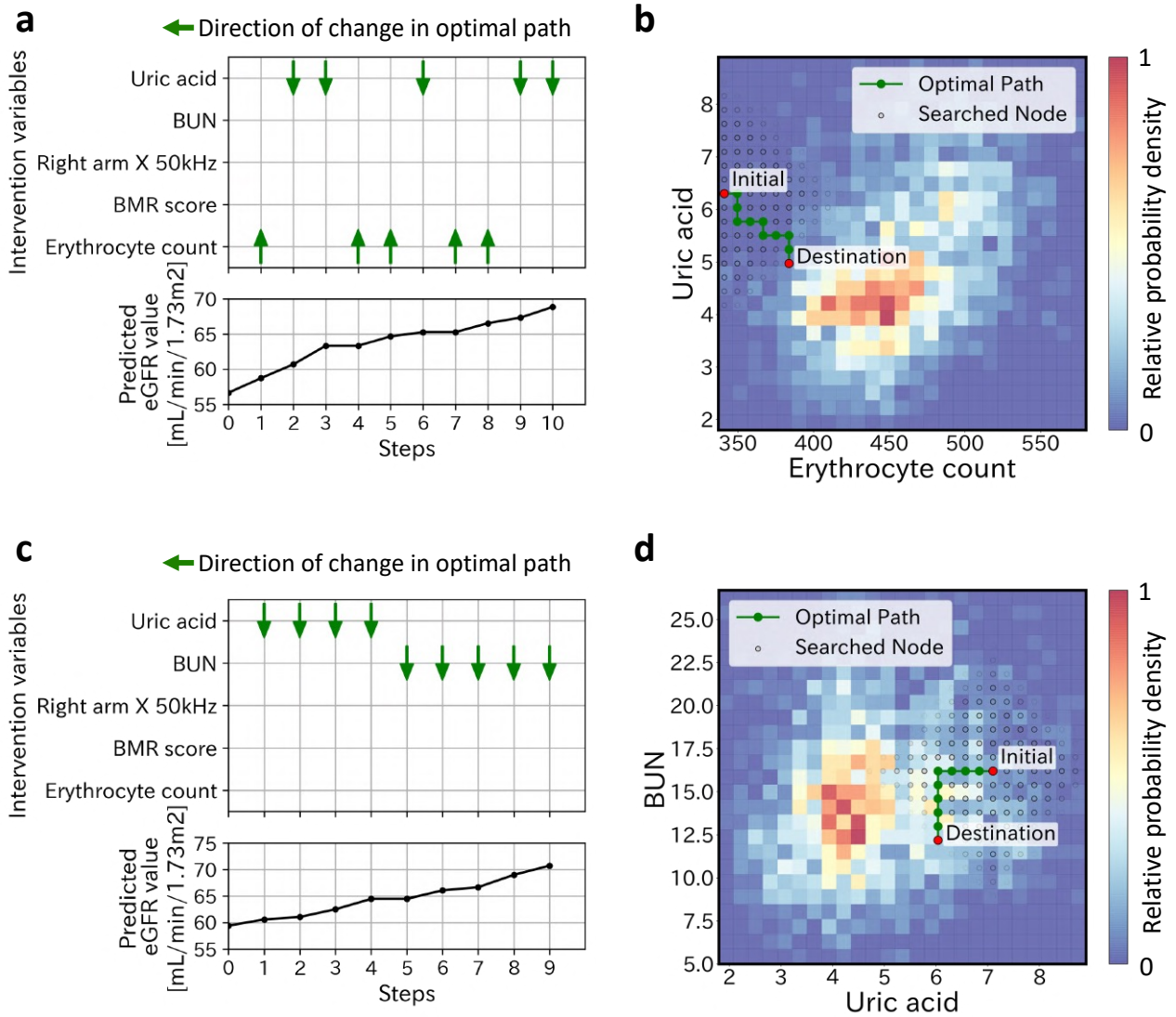
Supplementary Figure 24. Examples of personal actionable paths for treatment in hypertension risk classification task using random forest (RF). The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance 7 (**a, b**) and instance 8 (**c, d**). **a, c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: aspartate transaminase (AST [GOT]) and leg score (**b**), and gamma glutamyl transferase (g-GTP) and leg score (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



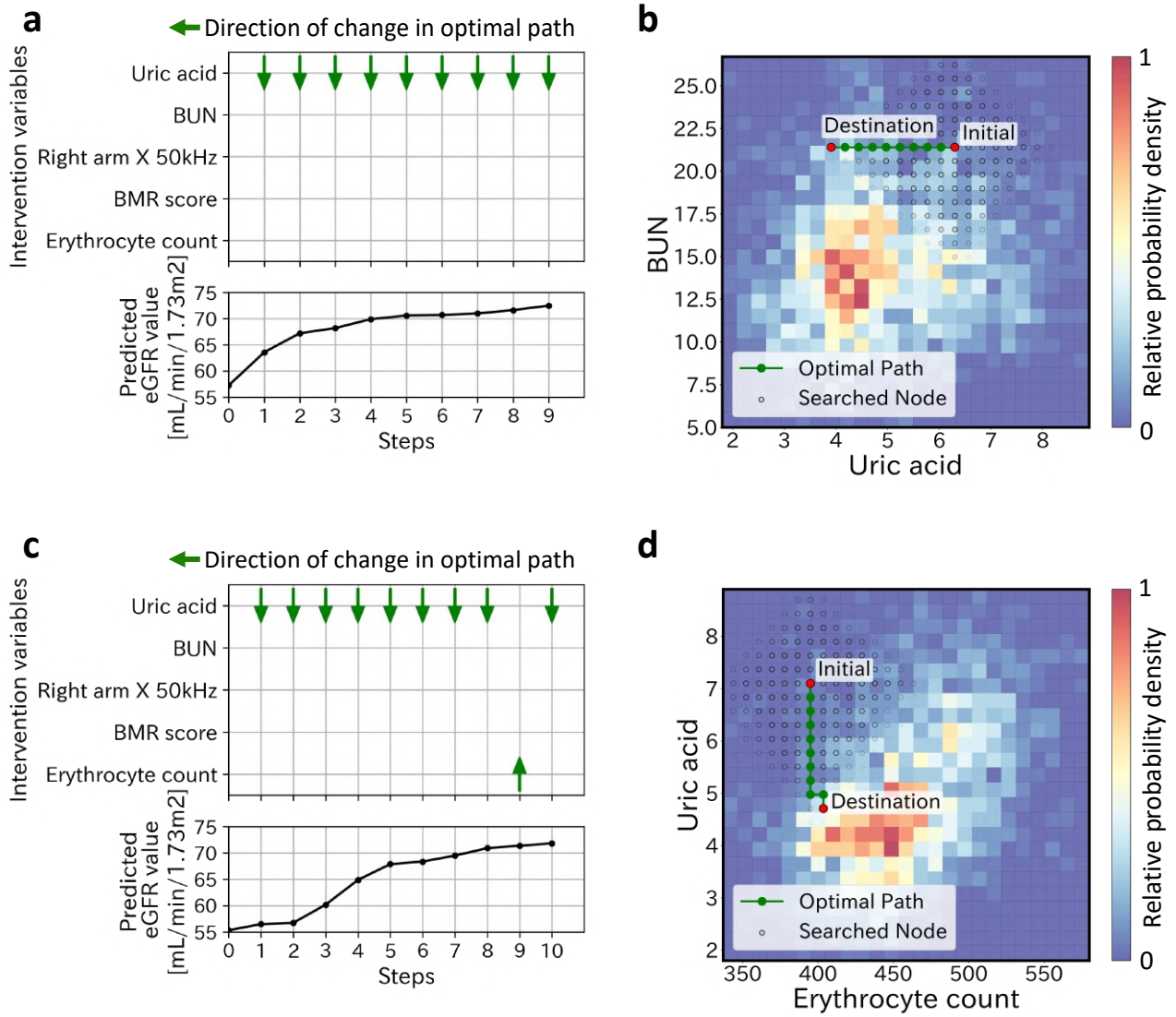
Supplementary Figure 25. Examples of personal actionable paths for treatment in hypertension risk classification task using support vector machine (SVM). The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance 7 (**a, b**) and instance 8 (**c, d**). **a, c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: immunoglobulin G (IgG) and leg score (**b**), and aspartate transaminase (AST [GOT]) and leg score (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



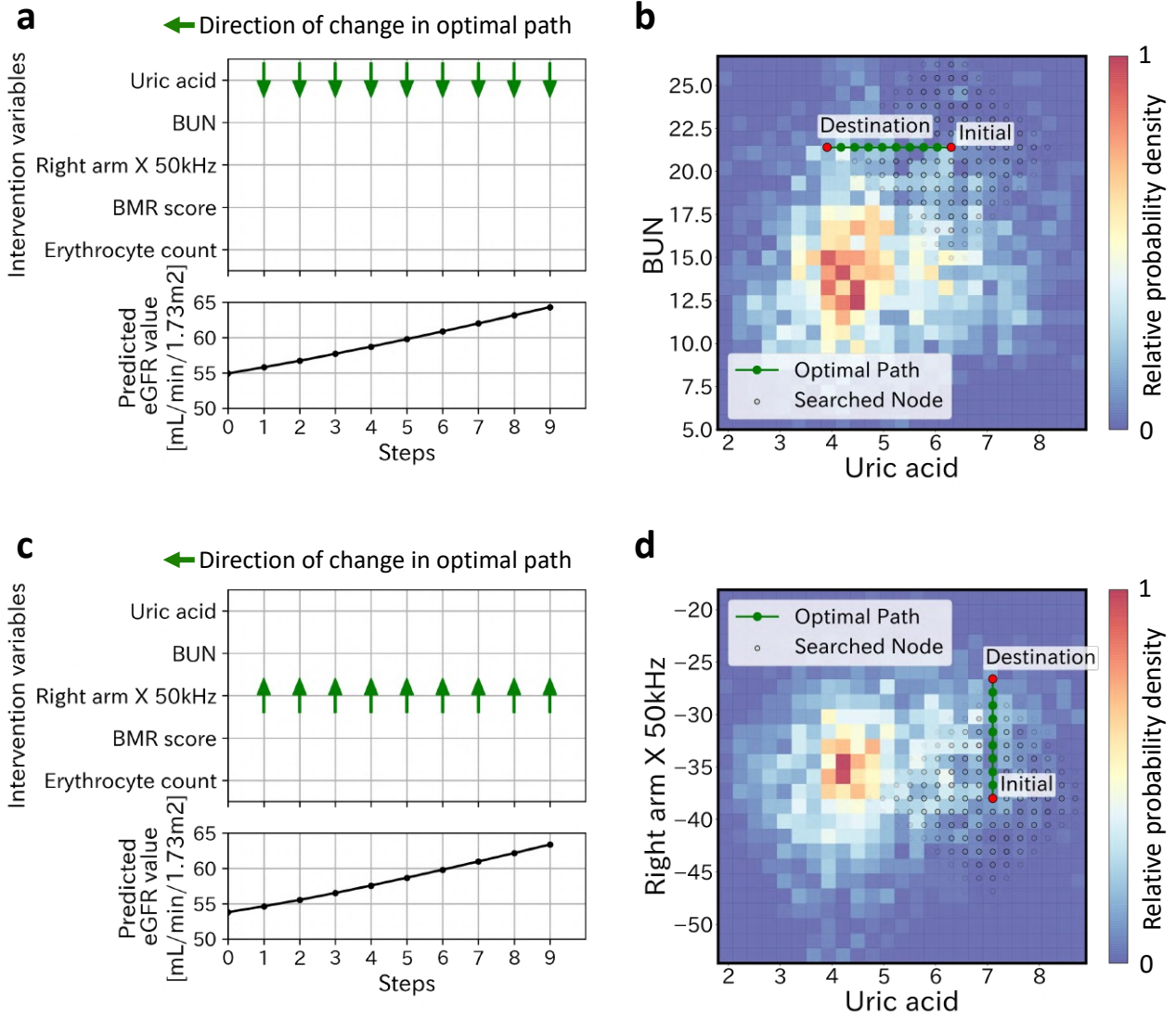
Supplementary Figure 26. Results of the proposed framework estimated glomerular filtration rate (eGFR) regression task. **a** Feature importance. These 25 features were selected by recursive feature elimination (RFE). RFE was performed with five-fold cross-validation, and the feature importance when 25 variables remained is shown for each fold ($n = 5$). The plot color represents the following: red: intervention variables, gray: variables which cannot be intervened, and blue: other variables. In box-plot, center line represents median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Details of features are described in Supplementary Data 1. **b–j** The results of applying our framework to prediction models based on nonlinear machine learning (ML) algorithms: XGBoost (**b**, **e**, **h**), random forest (**c**, **f**, **i**), and support vector machine (**d**, **g**, **j**). **b–d** Plots for prediction vs. true response variable. **e–g** Widely applicable Bayesian information criterion (WBIC) values of stochastic surrogate models with 1–8 mixture components. **h–j** Histogram of actionability scores. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



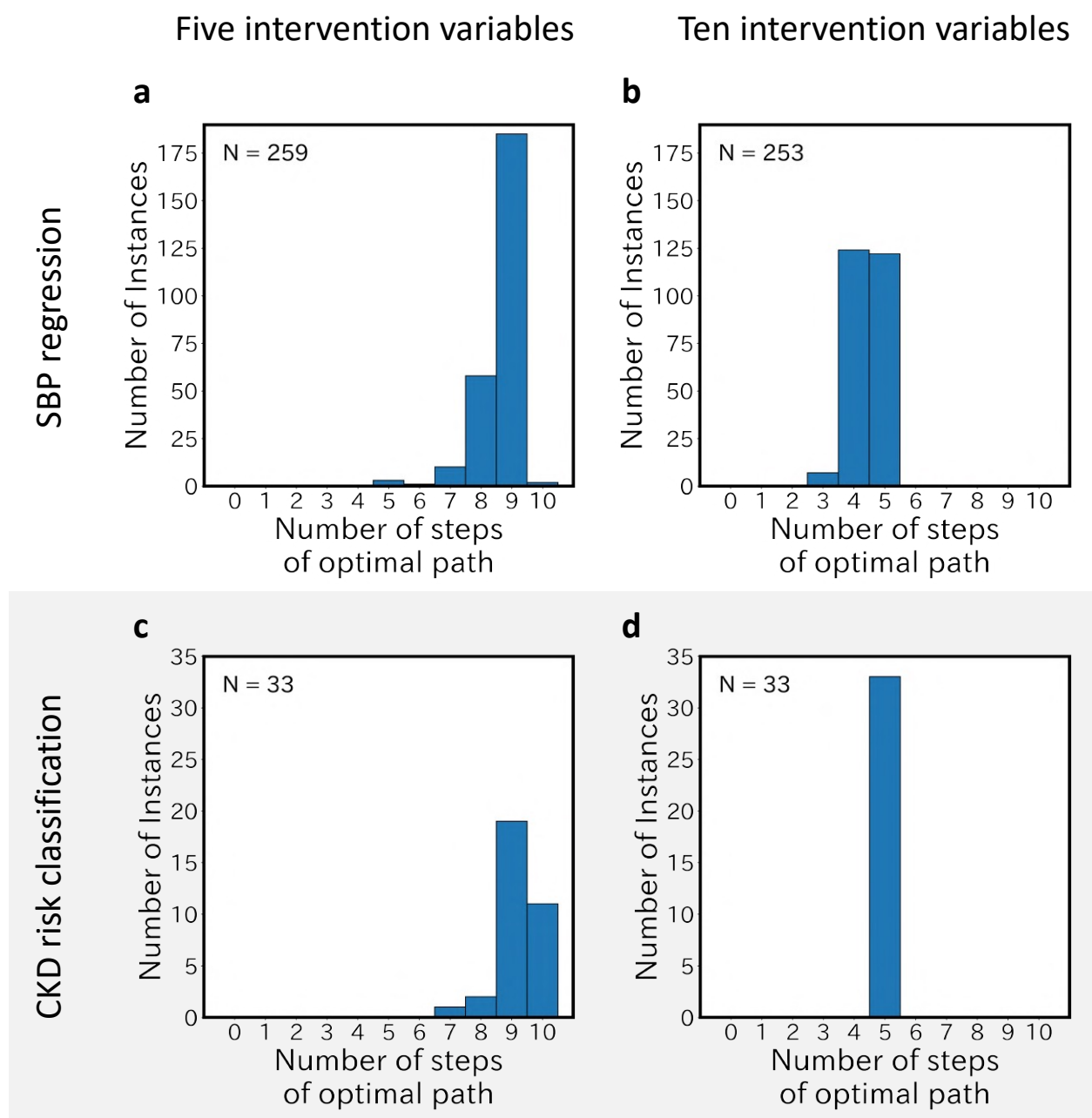
Supplementary Figure 27. Examples of personal actionable paths for treatment in estimated glomerular filtration rate (eGFR) regression task using XGBoost. The optimal paths for improving the response variable predicted by the XGBoost model are represented for randomly selected two examples: instance 9 (**a**, **b**) and instance 10 (**c**, **d**). **a**, **c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b**, **d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: uric acid and erythrocyte count (**b**), and uric acid and blood urea nitrogen (BUN) (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



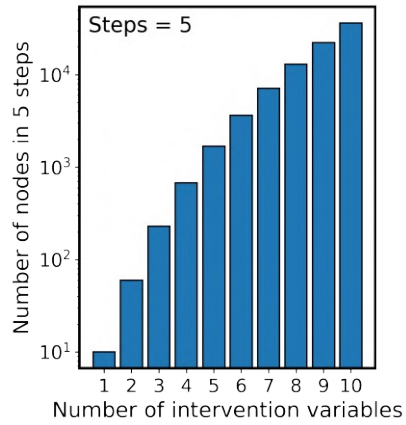
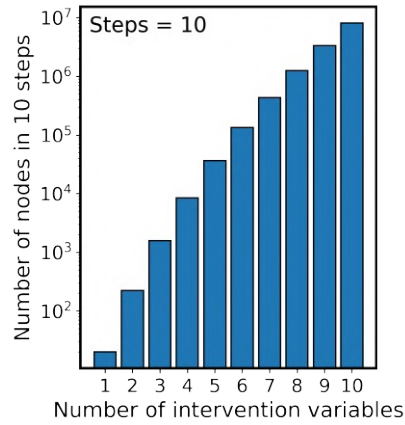
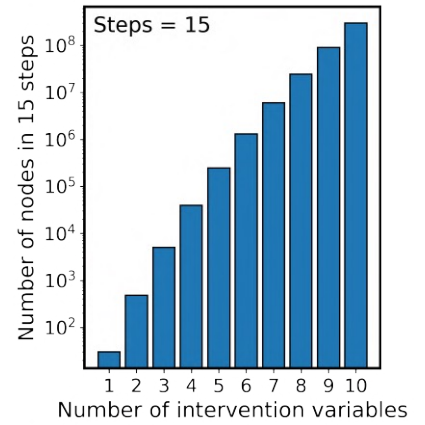
Supplementary Figure 28. Examples of personal actionable paths for treatment in estimated glomerular filtration rate (eGFR) regression task using random forest (RF). The optimal paths for improving the response variable predicted by the RF model are represented for the same instances selected on the application to XGBoost: instance 9 (**a, b**) and instance 10 (**c, d**). **a, c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: uric acid and blood urea nitrogen (BUN) (**b**), and uric acid and erythrocyte count (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



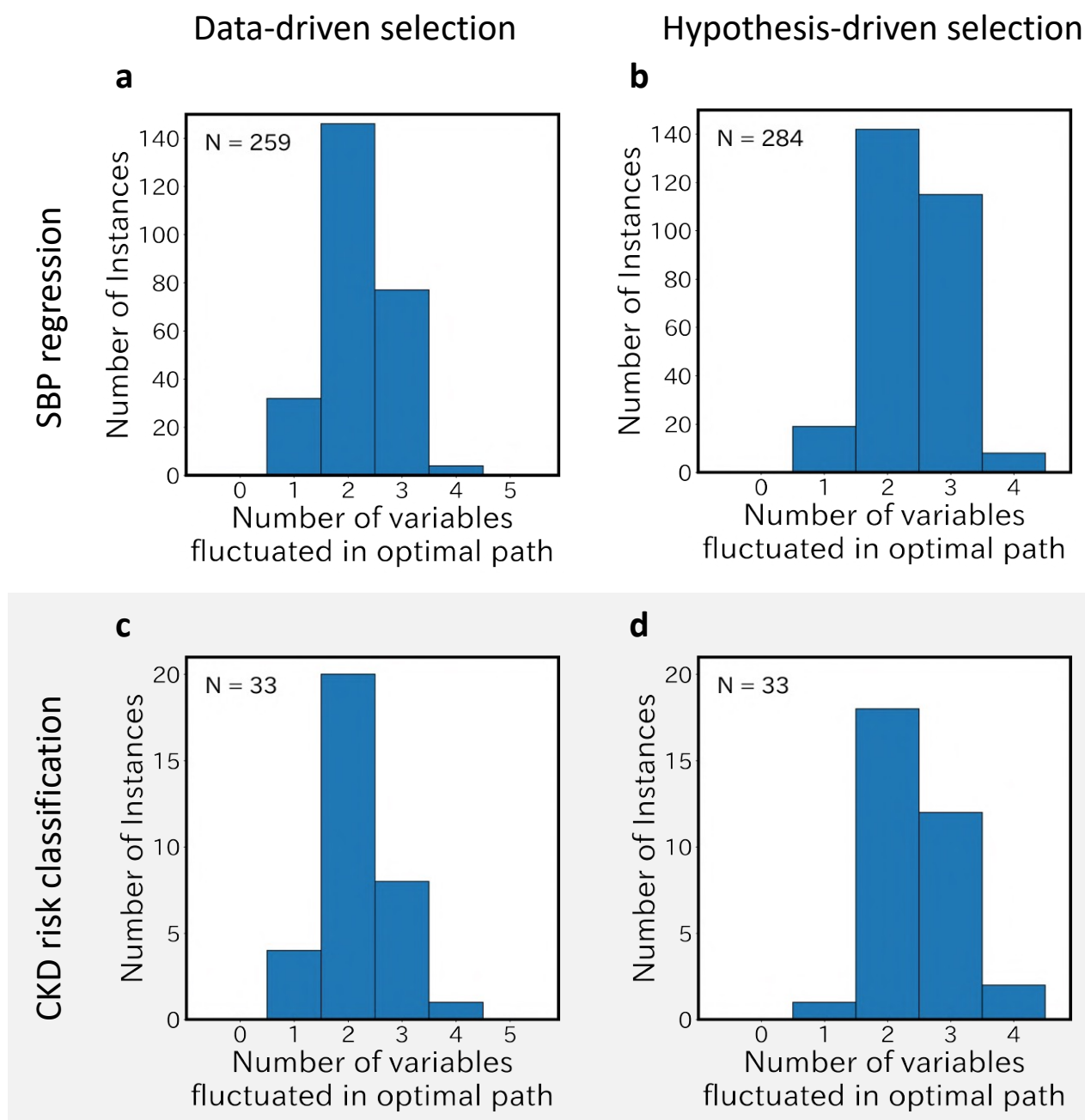
Supplementary Figure 29. Examples of personal actionable paths for treatment in estimated glomerular filtration rate (eGFR) regression task using support vector machine (SVM). The optimal paths for improving the response variable predicted by the SVM model are represented for the same instances selected on the application to XGBoost: instance 9 (**a, b**) and instance 10 (**c, d**). **a, c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d** 2D plots of the path. The 2D plots are shown regarding the two influential variables in the optimal path: uric acid and blood urea nitrogen (BUN) (**b**), and uric acid and right arm X 50 kHz (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



Supplementary Figure 30. Number of steps in optimal paths. The number of steps in the optimal paths for each instance is shown. **a, b** Top-five (**a**) or top-ten (**b**) important features were selected as intervention variables on systolic blood pressure (SBP) regression task. **c, d** Top-five (**c**) or top-ten (**d**) important features were selected as intervention variables on chronic kidney disease (CKD) risk classification task.

a**b****c**

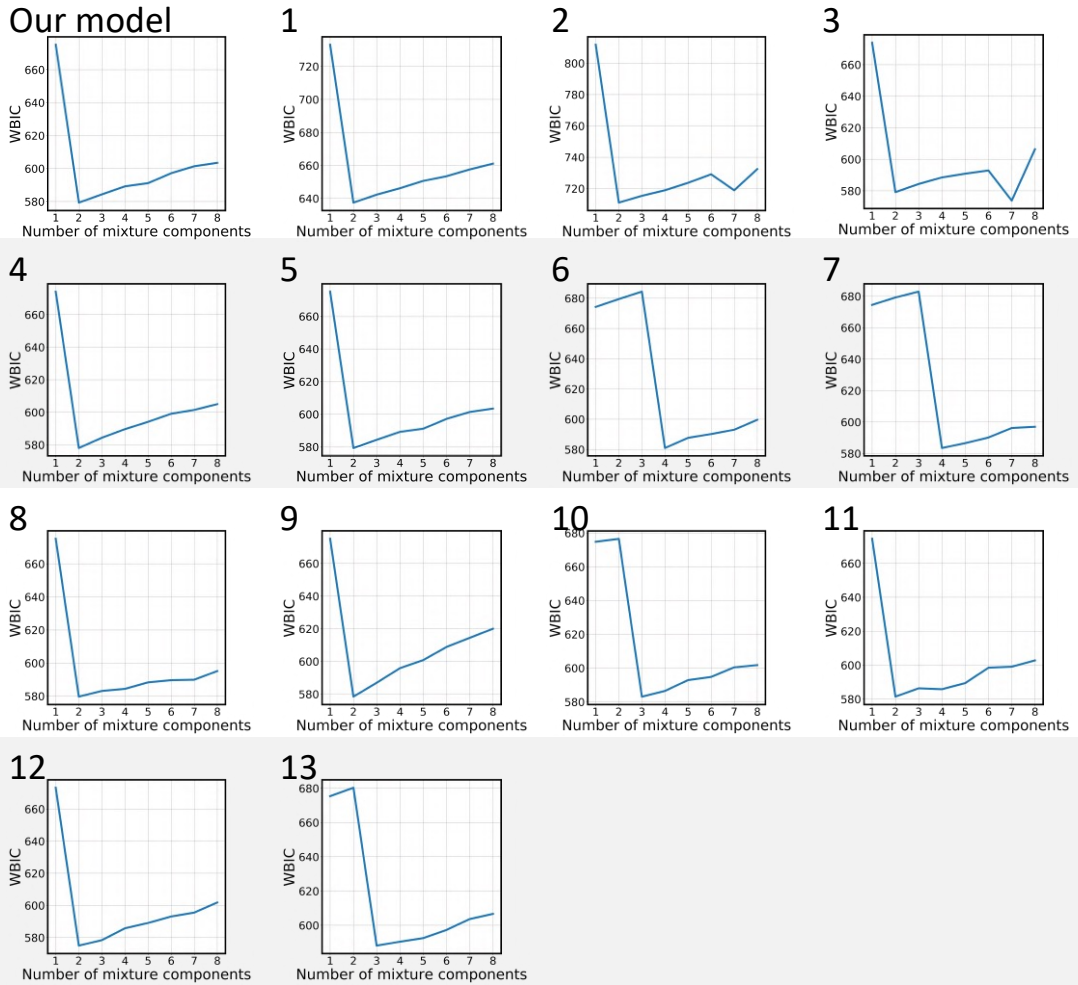
Supplementary Figure 31. Number of nodes within certain number of steps without detours. The number of nodes within 5, 10, or 15 steps are shown. The vertical axis is represented on a logarithmic scale. As the intervention variable increases, the number of nodes increases exponentially, which causes multistep path planning difficult.



Supplementary Figure 32. Number of intervention variables fluctuated in optimal paths. The number of intervention variables that fluctuated in the optimal path for each instance is shown. **a, b** Data-driven (**a**) or hypothesis-driven (**b**) selected intervention variables on systolic blood pressure (SBP) regression task. **c, d** Data-driven (**c**) or hypothesis-driven (**d**) selected intervention variables on chronic kidney disease (CKD) risk classification task.

a

Changes from our model		Changes from our model	
Our model	-	7	$\beta_{2,k} \sim \text{DoubleExponential}(0, 10)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 10)$
1	$\sigma = \text{RMSE}_{\text{test}}$	8	$\pi \sim \text{Dirichlet}(0.5)$
2	$\sigma = 2\text{RMSE}_{\text{test}}$	9	$\pi \sim \text{Dirichlet}(2)$
3	$\beta_{1,k} \sim N(0, 1)$	10	$\mathbf{m}_k \sim N(0, \mathbf{I})$
4	$\beta_{1,k} \sim N(0, 3)$	11	$\mathbf{m}_k \sim N(0, 3\mathbf{I})$
5	$\beta_{2,k} \sim \text{DoubleExponential}(0, 0.1)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 0.1)$	12	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 1))$
6	$\beta_{2,k} \sim \text{DoubleExponential}(0, 5)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 5)$	13	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 5))$

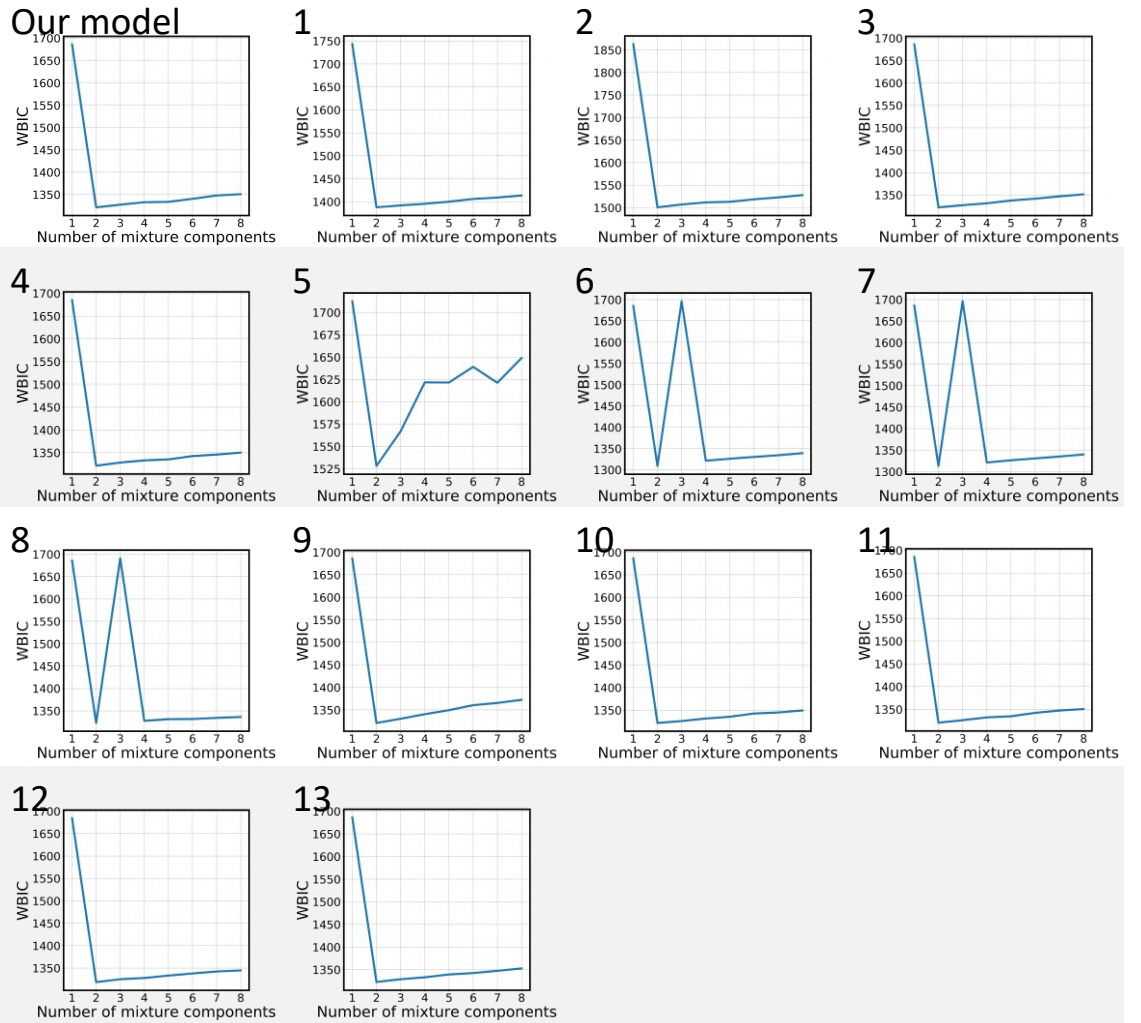
b

Supplementary Figure 33. Sensitivity analysis of hyperparameters in priors of hierarchical Bayesian model using three-dimensional (3D) synthetic dataset. a List of hyperparameters examined. The hyperparameters are changed from our model setting in the main manuscript. **b** Widely applicable Bayesian information criterion (WBIC) of stochastic surrogate models with 1–8 mixture components.

a

Changes from our model		Changes from our model	
Our model	-	7	$\beta_{2,k} \sim \text{DoubleExponential}(0, 10)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 10)$
1	$\sigma = \text{RMSE}_{\text{test}}$	8	$\pi \sim \text{Dirichlet}(\mathbf{0.5})$
2	$\sigma = 2\text{RMSE}_{\text{test}}$	9	$\pi \sim \text{Dirichlet}(\mathbf{2})$
3	$\beta_{1,k} \sim N(0, 1)$	10	$\mathbf{m}_k \sim N(0, \mathbf{I})$
4	$\beta_{1,k} \sim N(0, 3)$	11	$\mathbf{m}_k \sim N(0, 3\mathbf{I})$
5	$\beta_{2,k} \sim \text{DoubleExponential}(0, 0.1)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 0.1)$	12	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 1))$
6	$\beta_{2,k} \sim \text{DoubleExponential}(0, 5)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 5)$	13	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 5))$

b



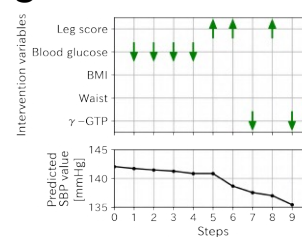
Supplementary Figure 34. Sensitivity analysis of hyperparameters in priors of hierarchical Bayesian model using five-dimensional (5D) synthetic dataset. a List of hyperparameters examined. The hyperparameters are changed from our model setting in the main manuscript. **b** Widely applicable Bayesian information criterion (WBIC) of stochastic surrogate models with 1–8 mixture components.

Our model

1

2

3

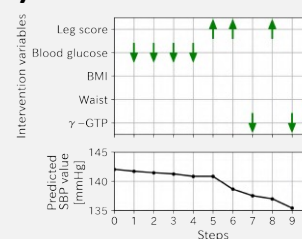


4

5

6

7

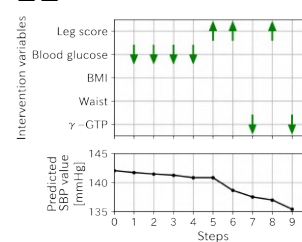


8

9

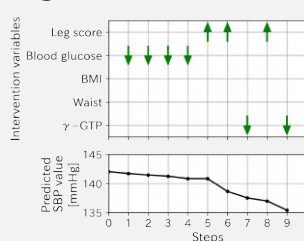
10

11



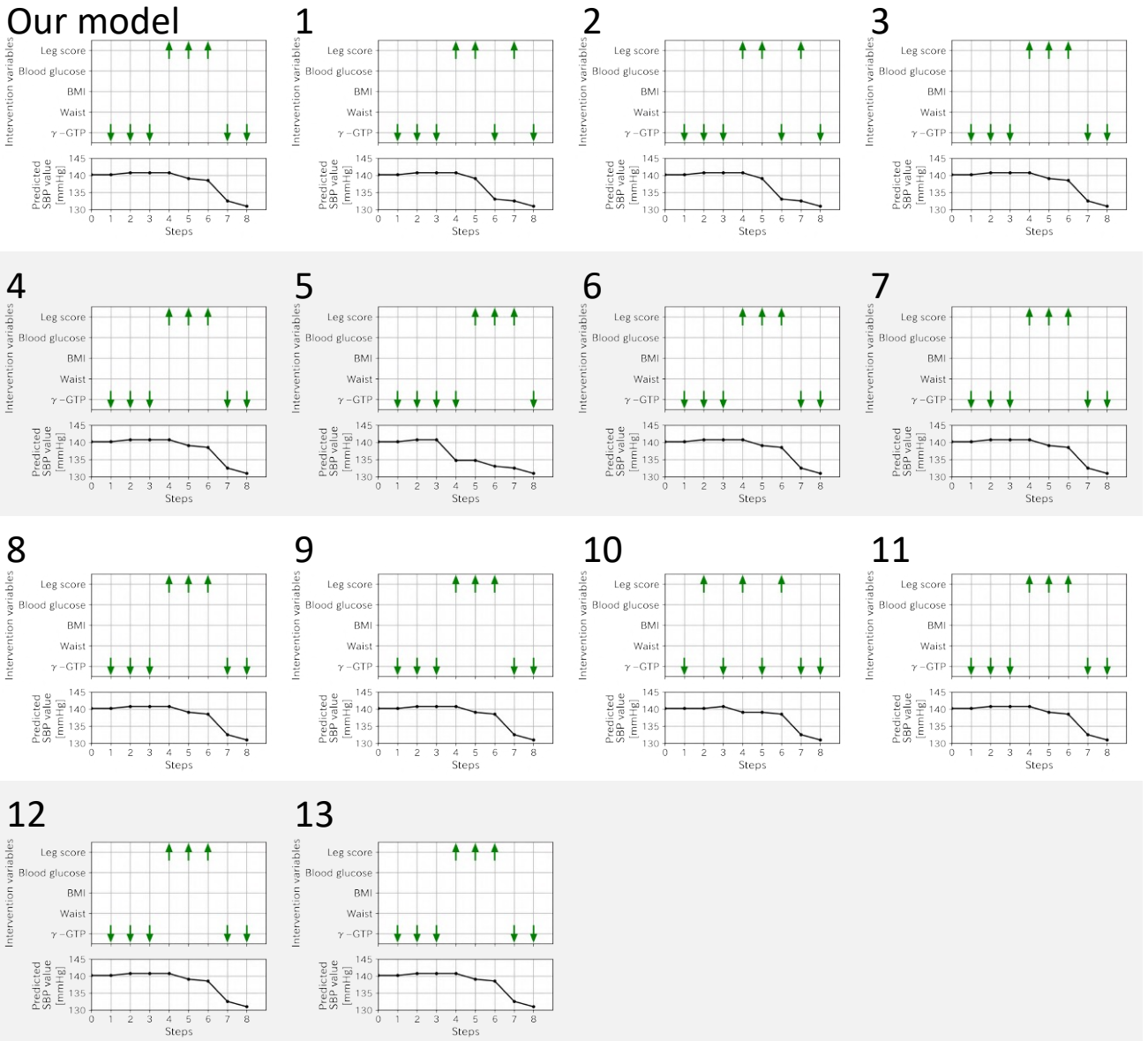
12

13



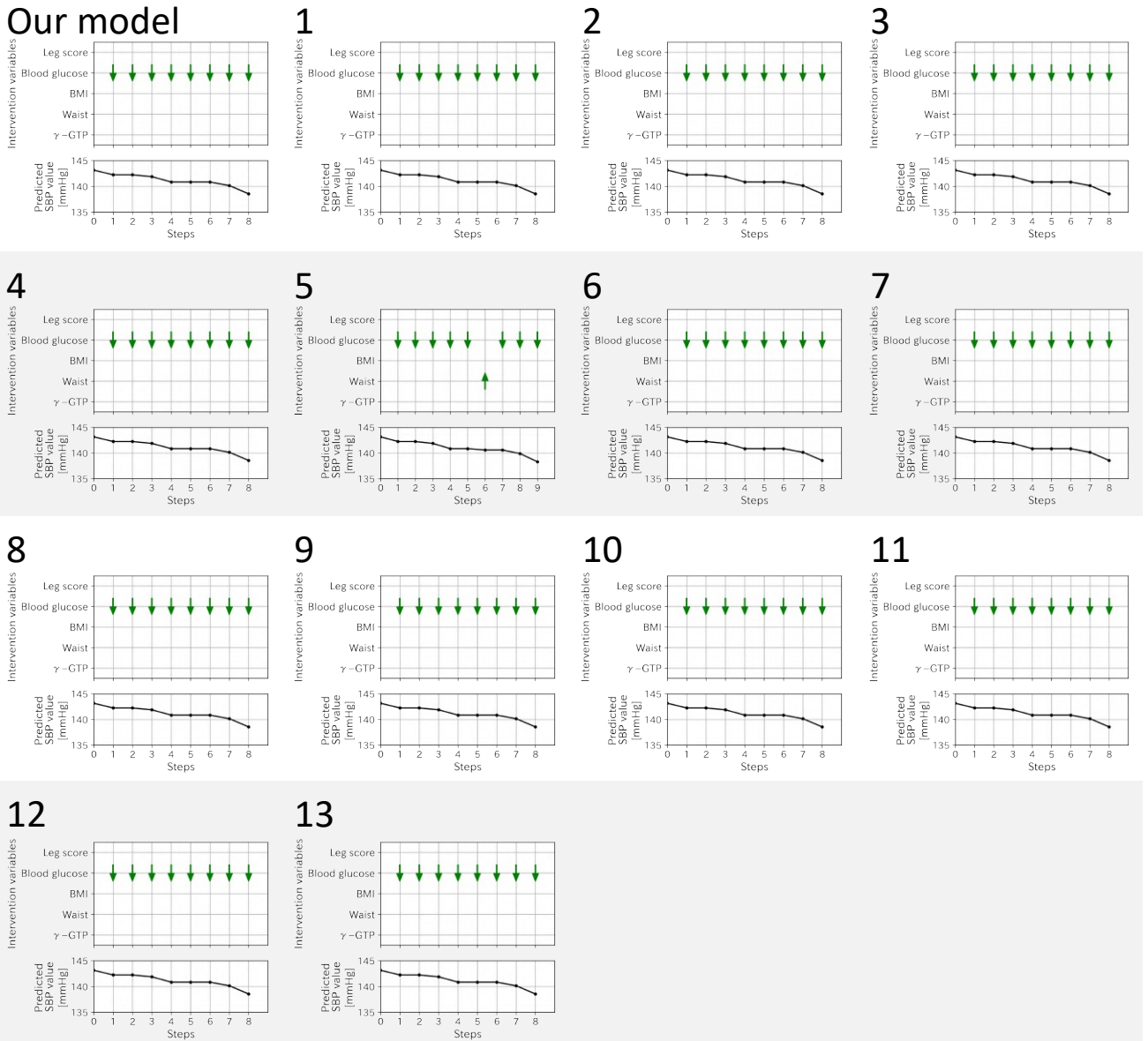
d Instance 2

Our model



e Instance 3

Our model



Supplementary Figure 35. Sensitivity analysis of hyperparameters in priors of hierarchical Bayesian model on systolic blood pressure (SBP) regression task. a List of hyperparameters examined. The hyperparameters are changed from our model setting in the main manuscript. **b** Widely applicable Bayesian information criterion (WBIC) of stochastic surrogate models with 1–8 mixture components. **c–e** Paths planned using stochastic surrogate models for each hyperparameter setting. Each subfigure corresponds to the instances described in Fig. 6 in the main text: instance 1 (**c**), instance 2 (**d**), and instance 3 (**e**).

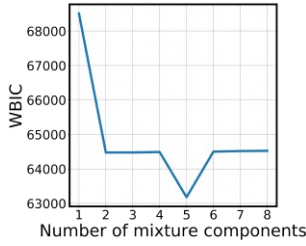
a

Changes from our model	
Our model	-
1	$\beta_{1,k} \sim N(0, 1)$
2	$\beta_{1,k} \sim N(0, 3)$
3	$\beta_{2,k} \sim \text{DoubleExponential}(0, 0.1)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 0.1)$
4	$\beta_{2,k} \sim \text{DoubleExponential}(0, 5)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 5)$
5	$\beta_{2,k} \sim \text{DoubleExponential}(0, 10)$ $\beta_{3,k} \sim \text{DoubleExponential}(0, 10)$

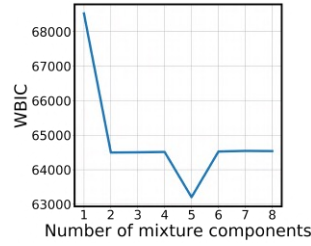
Changes from our model	
6	$\pi \sim \text{Dirichlet}(0.5)$
7	$\pi \sim \text{Dirichlet}(2)$
8	$\mathbf{m}_k \sim N(0, \mathbf{I})$
9	$\mathbf{m}_k \sim N(0, 3\mathbf{I})$
10	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 1))$
11	$\Sigma_k \sim \text{diag}(\text{Cauchy}(0, 5))$

b

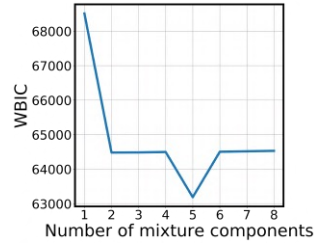
Our model



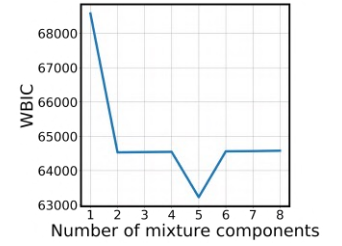
1



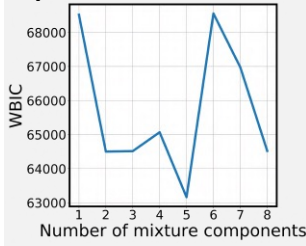
2



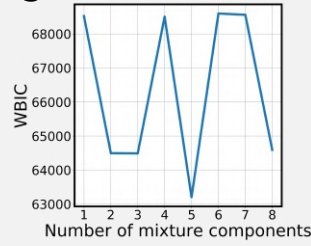
3



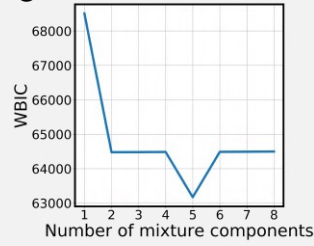
4



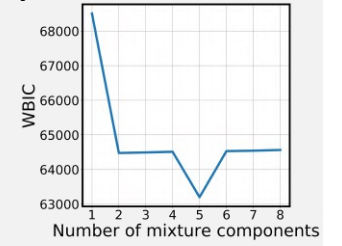
5



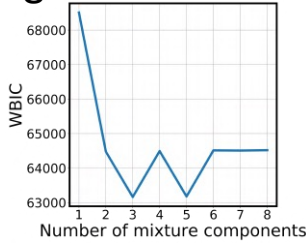
6



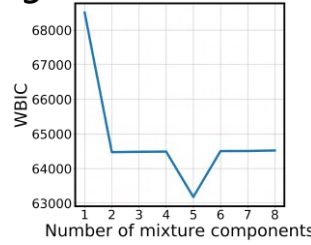
7



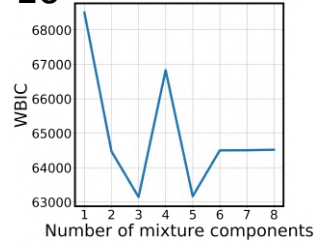
8



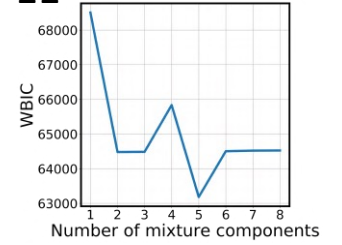
9



10

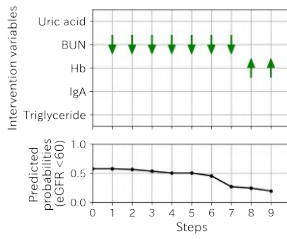


11

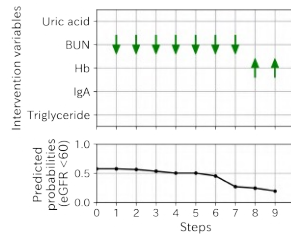


c Instance 4

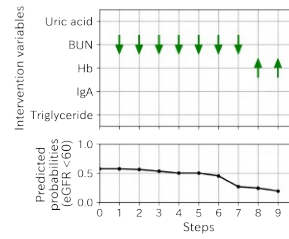
Our model



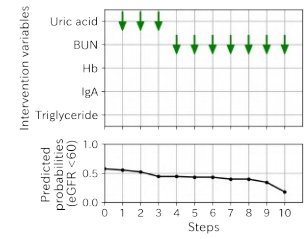
1



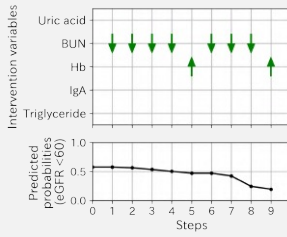
2



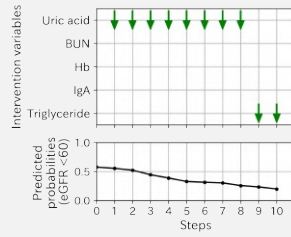
3



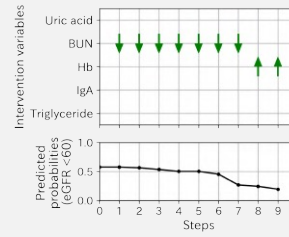
4



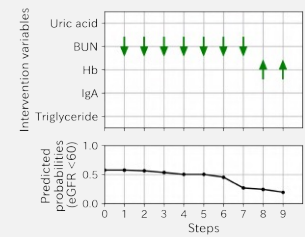
5



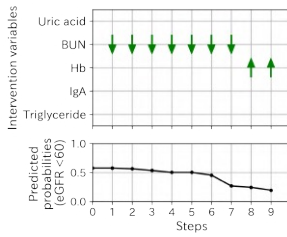
6



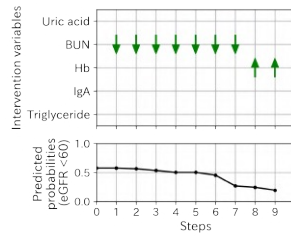
7



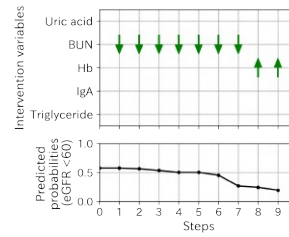
8



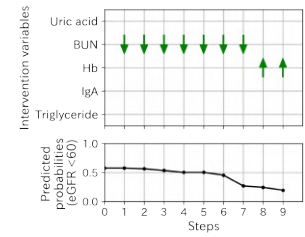
9



10

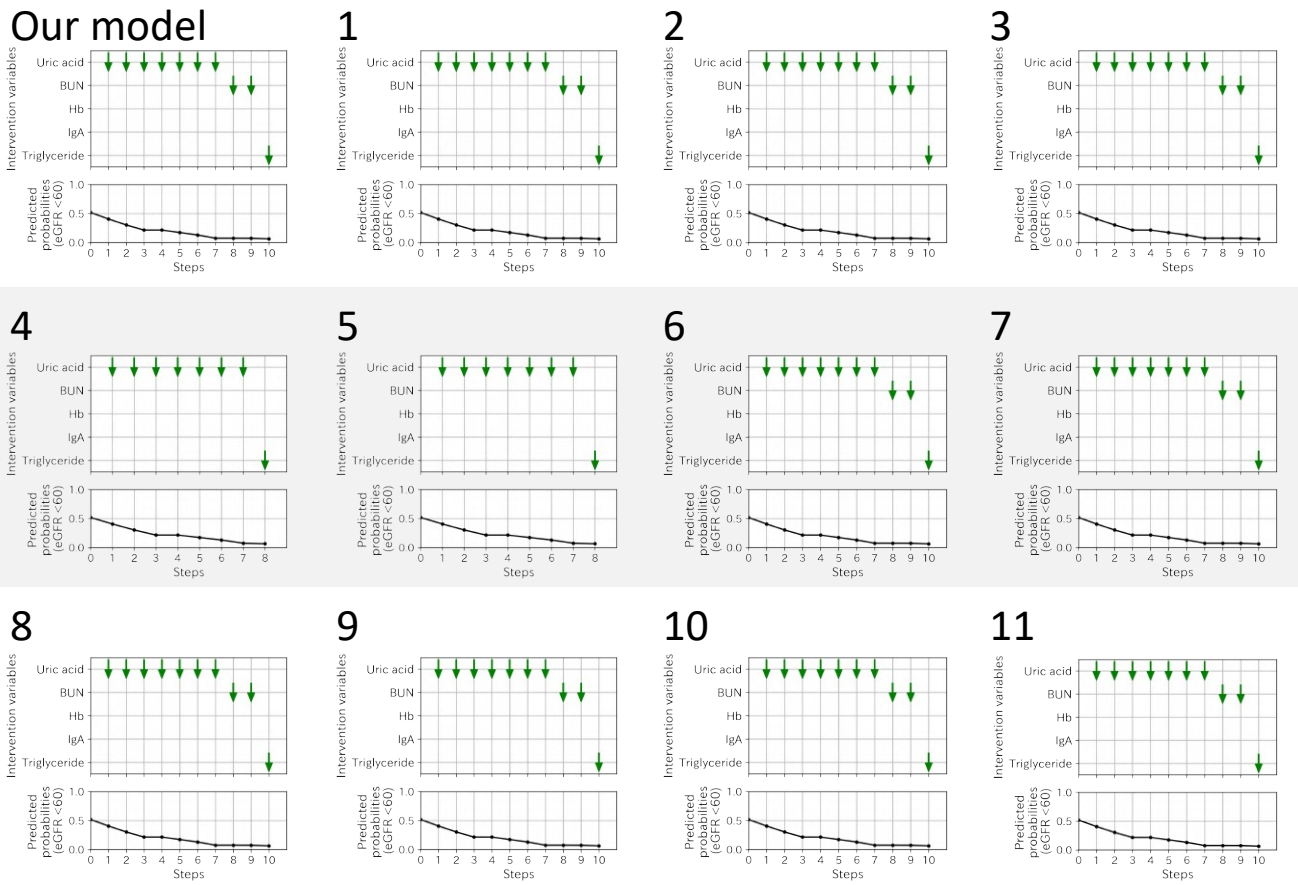


11



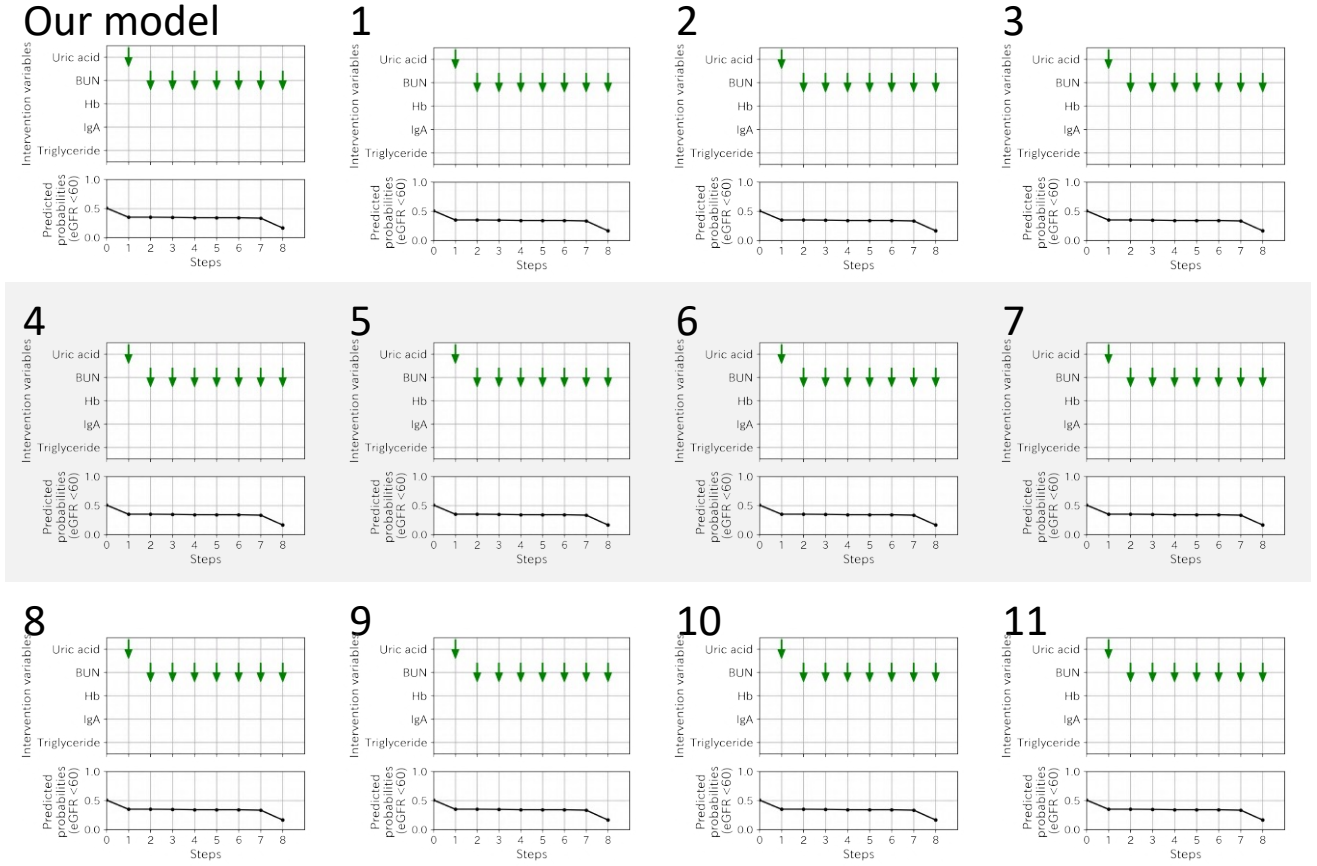
d Instance 5

Our model

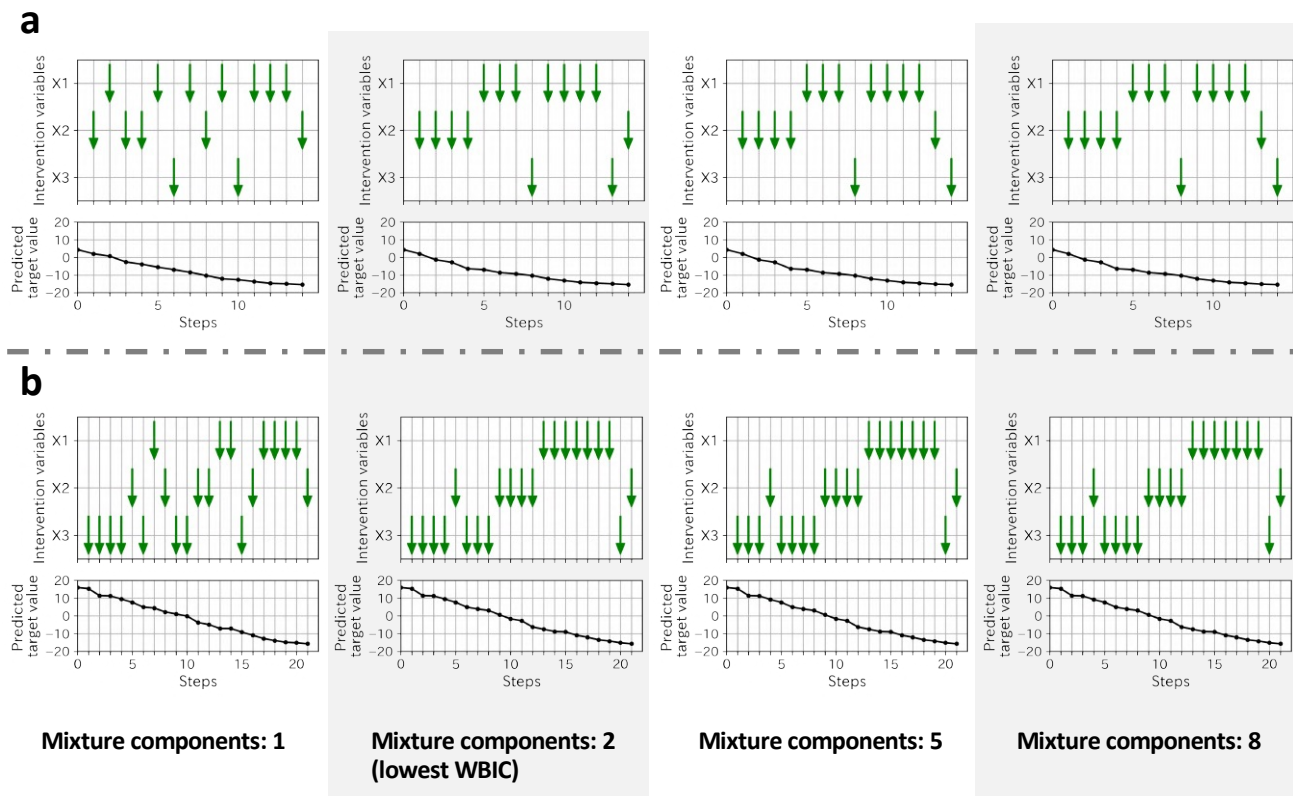


e Instance 6

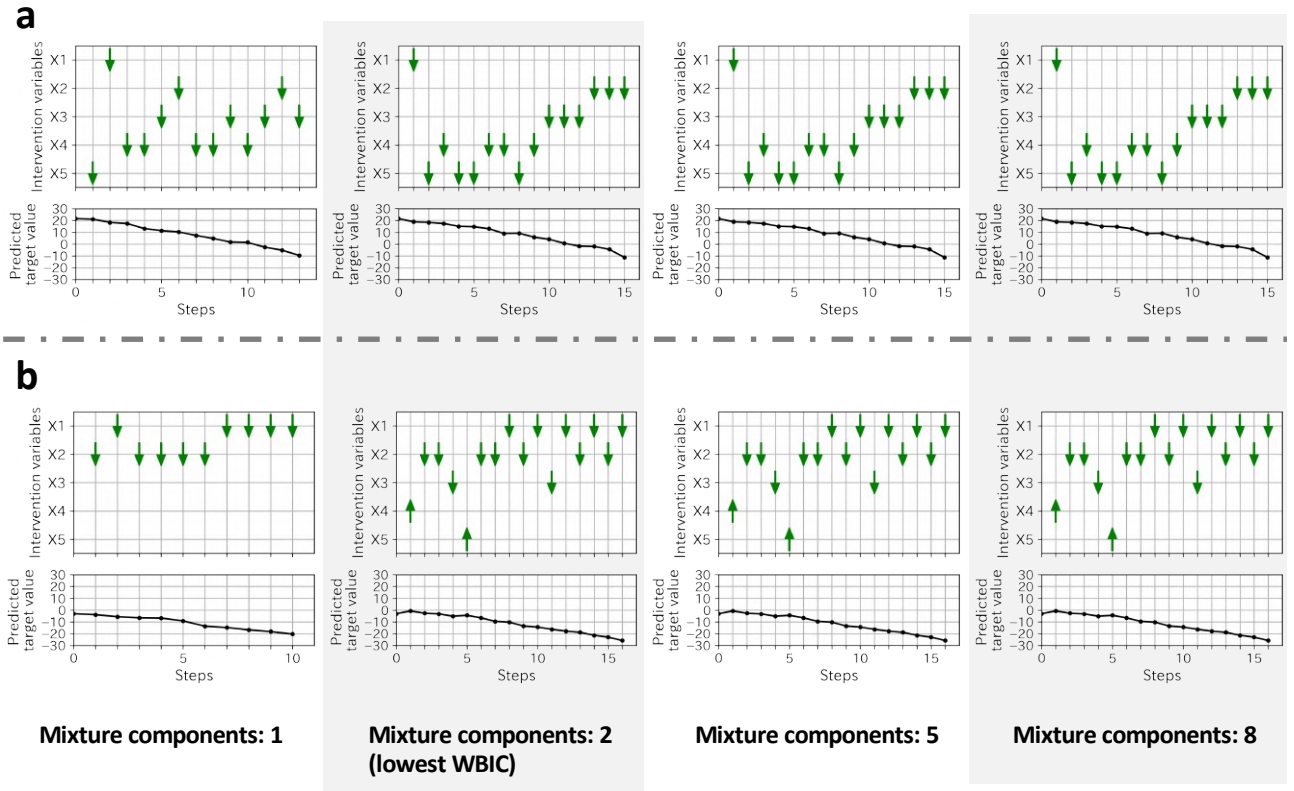
Our model



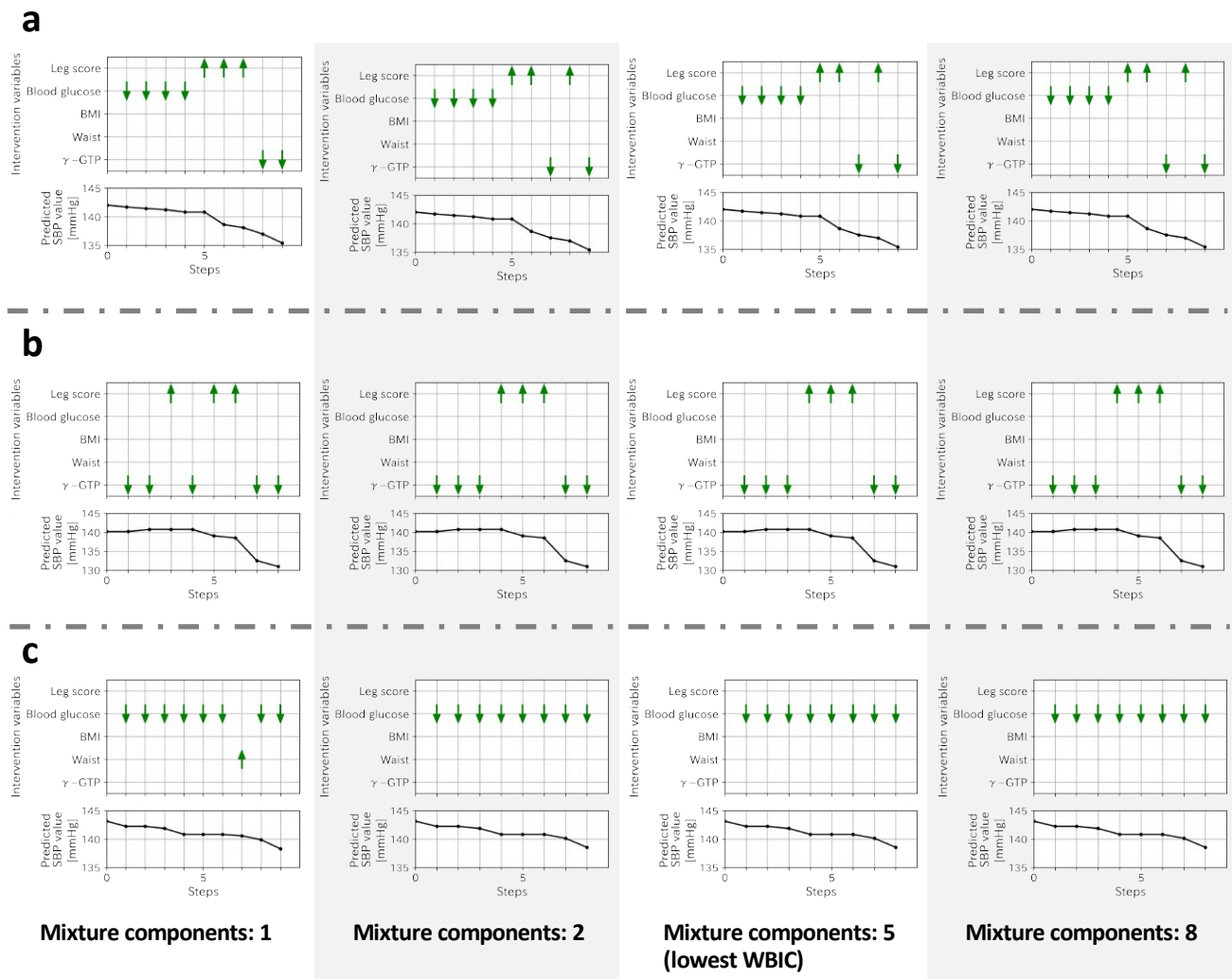
Supplementary Figure 36. Sensitivity analysis of hyperparameters in priors of hierarchical Bayesian model on chronic kidney disease (CKD) risk classification task. a List of hyperparameters examined. The hyperparameters are changed from our model setting in the main manuscript. **b** Widely applicable Bayesian information criterion (WBIC) of stochastic surrogate models with 1–8 mixture components. **c–e** Paths planned using stochastic surrogate models for each hyperparameter setting. Each subfigure corresponds to the instances described in Fig. 8 in the main text: instance 4 (**c**), instance 5 (**d**), and instance 6 (**e**).



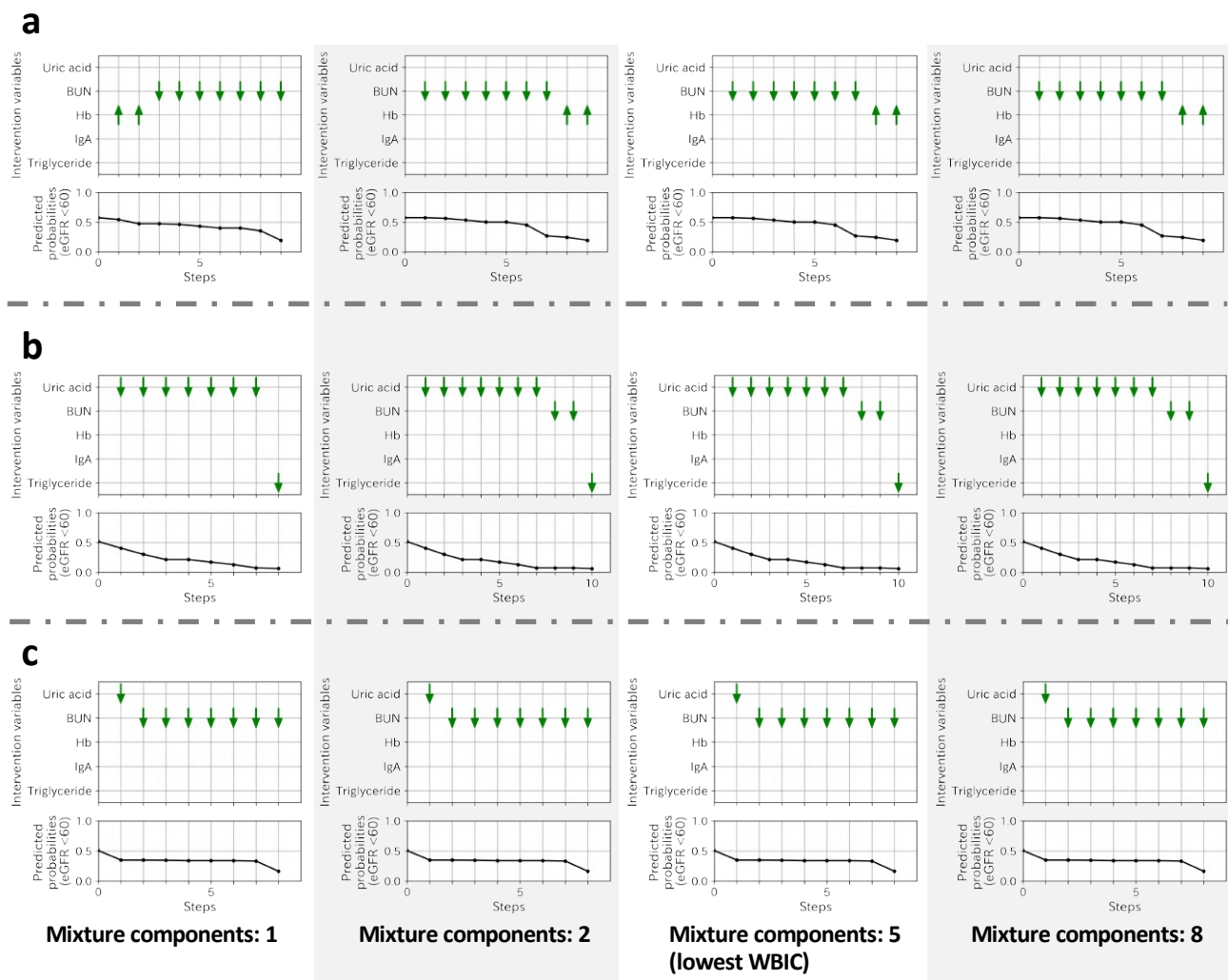
Supplementary Figure 37. Sensitivity analysis of path planning using hierarchical Bayesian model with different number of mixture components on three-dimensional (3D) synthetic dataset. The optimal paths for improving the response variable predicted by the XGBoost model were planned using hierarchical Bayesian models with different number of mixture components. Displayed instances are the same as in Fig. 4: instance A (**a**) and instance B (**b**).



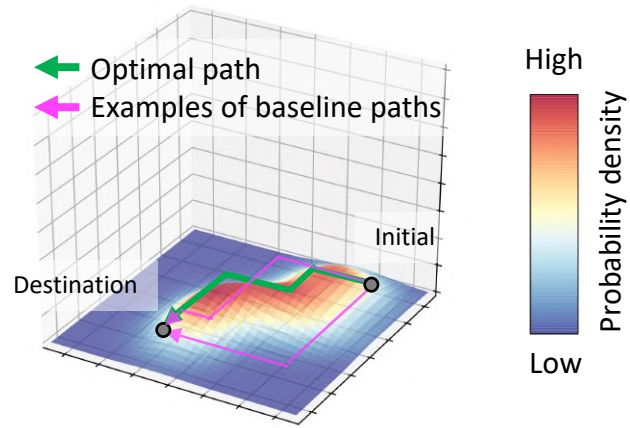
Supplementary Figure 38. Sensitivity analysis of path planning using hierarchical Bayesian model with different number of mixture components on five-dimensional (5D) synthetic dataset. The optimal paths for improving the response variable predicted by the XGBoost model were planned using hierarchical Bayesian models with different number of mixture components. Displayed instances are the same as in Supplementary Fig. 13: instance C (**a**) and instance D (**b**).



Supplementary Figure 39. Sensitivity analysis of path planning using hierarchical Bayesian model with different number of mixture components on systolic blood pressure (SBP) regression task. The optimal paths for improving the response variable predicted by the XGBoost model were planned using hierarchical Bayesian models with different number of mixture components. Displayed instances are the same as in Fig. 6: instance 1 (**a**), instance 2 (**b**), and instance 3 (**c**).



Supplementary Figure 40. Sensitivity analysis of path planning using hierarchical Bayesian model with different number of mixture components on chronic kidney disease (CKD) risk classification task. The optimal paths for improving the response variable predicted by the XGBoost model were planned using hierarchical Bayesian models with different number of mixture components. Displayed instances are the same as in Fig. 8: instance 4 (**a**), instance 5 (**b**), and instance 6 (**c**).

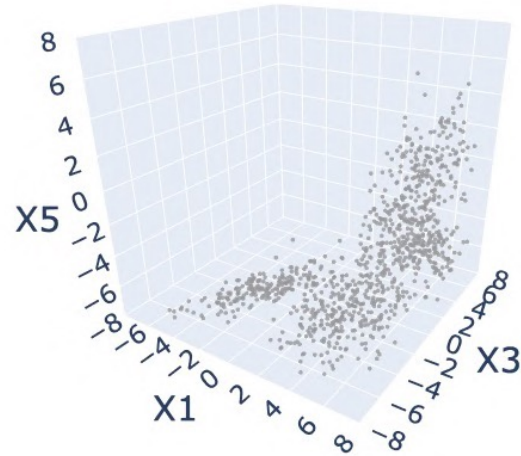
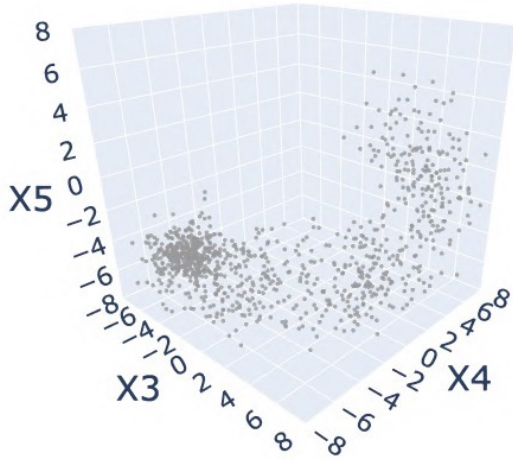
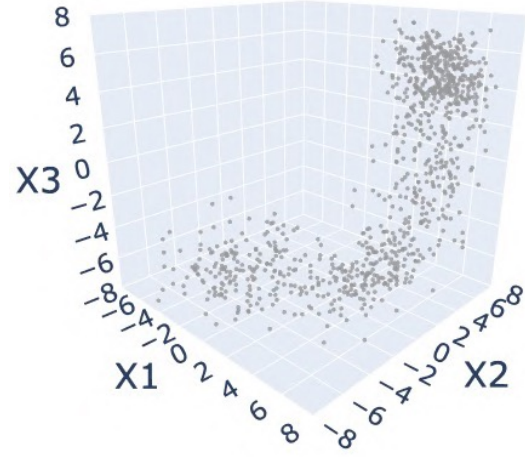


Supplementary Figure 41. Schematic explanation about actionability score. A baseline path is defined as a path that connects the initial node and destination node of the optimal path presented by the framework with a random minimum step, that is, without detour. Actionability score is defined as the logarithm of the ratio of the optimal path-actionability to the baseline path-actionability. Due to the large number of baseline paths, we used the geometric mean of the actionability of ten random baseline paths to calculate the score. The higher the actionability score, the optimal path is constructed through nodes with higher probabilities than the baseline path.

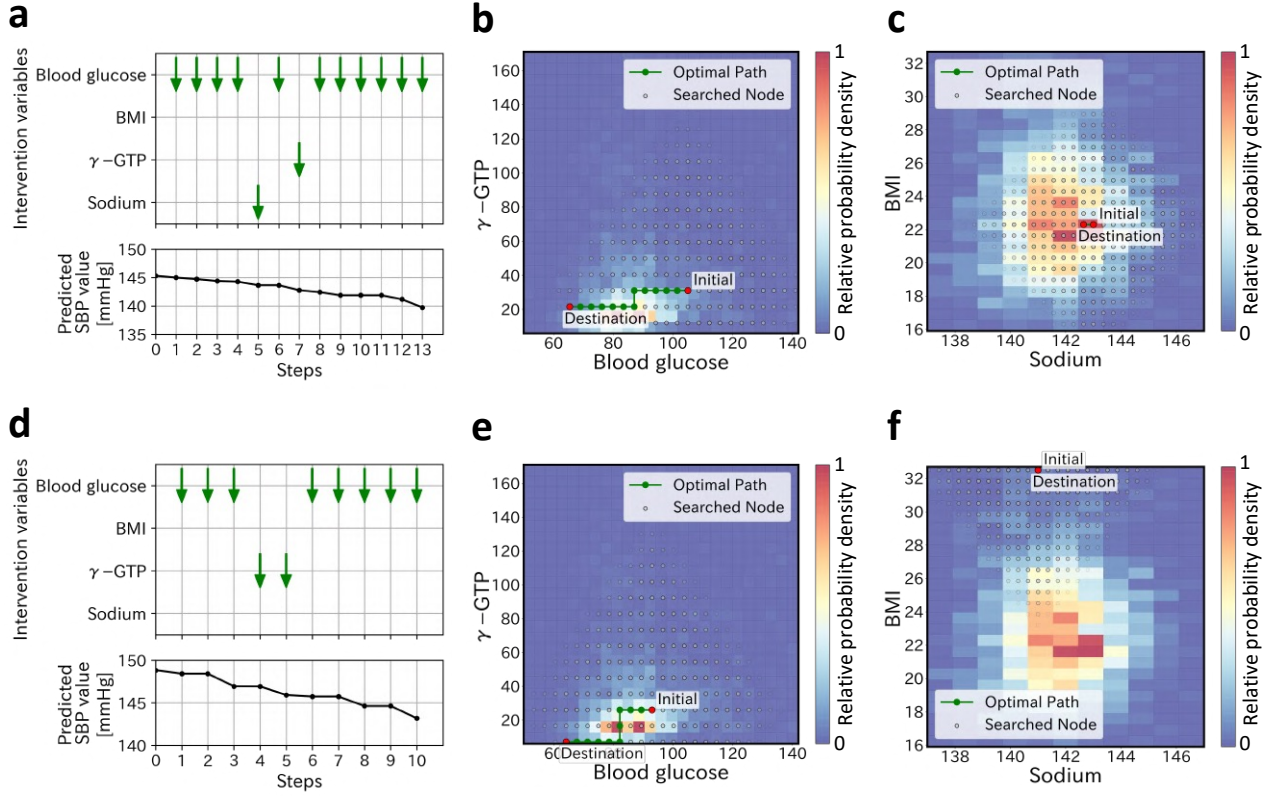
$$\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad k \in \{1, \dots, 5\}$$

$$\boldsymbol{\mu}_k = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_5 \end{bmatrix} \quad \mu_i = \begin{cases} 5 & (i < k) \\ 0 & (i = k) \\ -5 & (i > k) \end{cases}$$

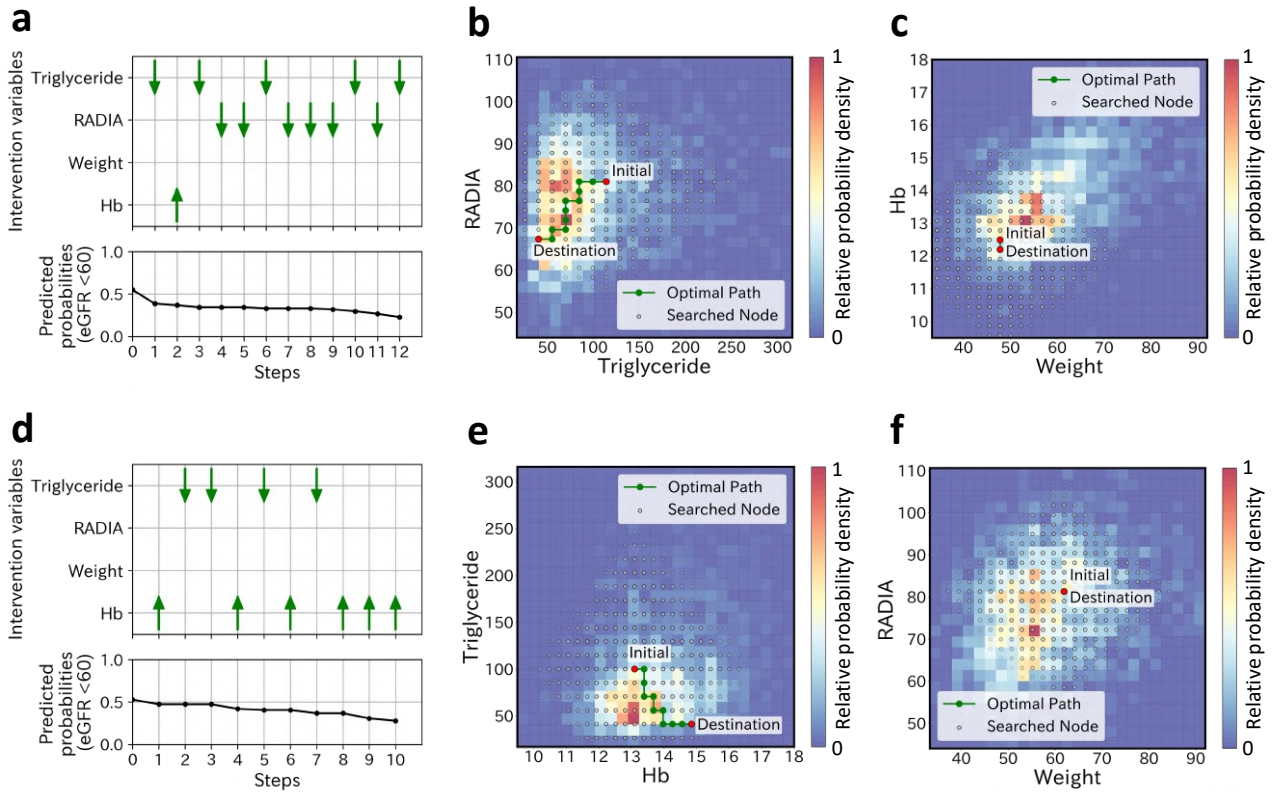
$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{15} \\ \vdots & \ddots & \vdots \\ \sigma_{51} & \dots & \sigma_{55} \end{bmatrix} \quad \sigma_{ij} = \begin{cases} 5 & (i = j = k) \\ 1 & (i = j \neq k) \\ \mathcal{N}(0, 1) & (i < j) \\ \sigma_{ji} & (i > j) \end{cases}$$



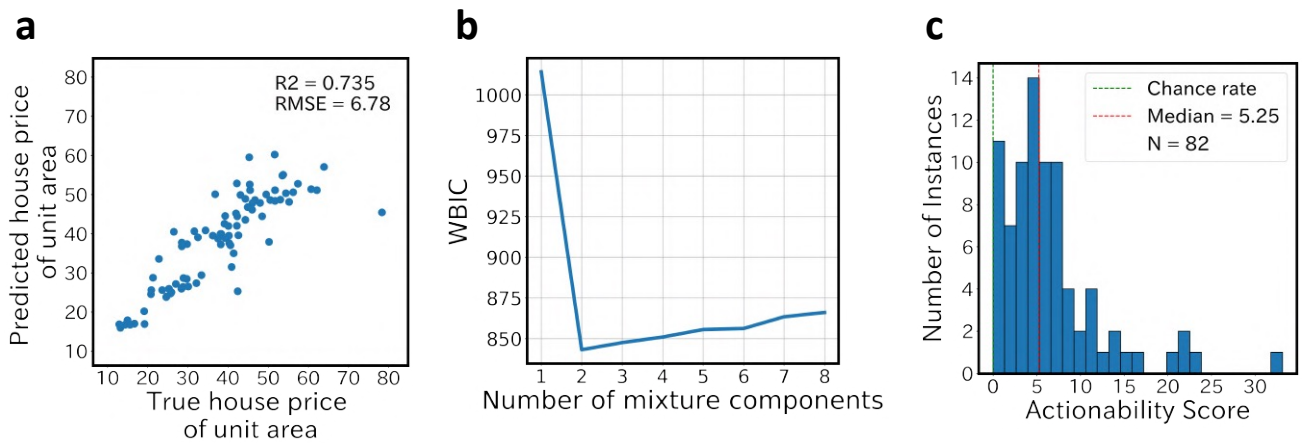
Supplementary Figure 42. Generation of five-dimensional (5D) synthetic dataset. The five 5D normal distributions generated 200 data points that consisted of x_i ($i \in \{1, \dots, 5\}$). Subsequently, a response variable was set to the sum of x_i with Gaussian noise ($\sigma = 2$). The synthetic dataset consisted of a total of 1,000 data points with explanatory variables ($X_1, \dots, X_5$) and a response variable.



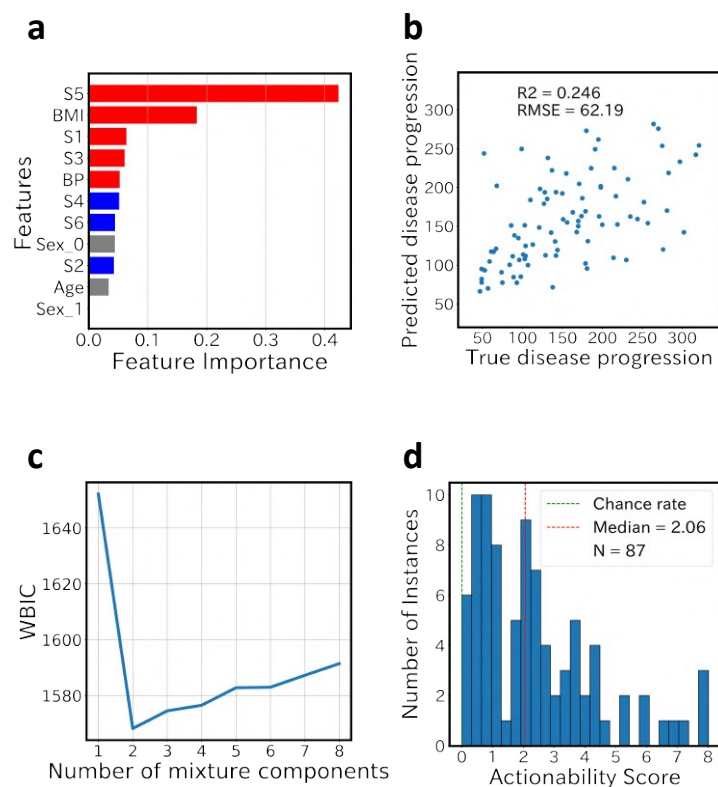
Supplementary Figure 43. Examples of personal actionable paths for treatment with intervention variables based on hypothesis-driven selection in systolic blood pressure (SBP) regression task. The optimal paths for improving the response variable predicted by the ML model are represented for two examples: instance 11 (**a–c**) and instance 12 (**d–f**). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, c, e, f** 2D plots of the path. In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



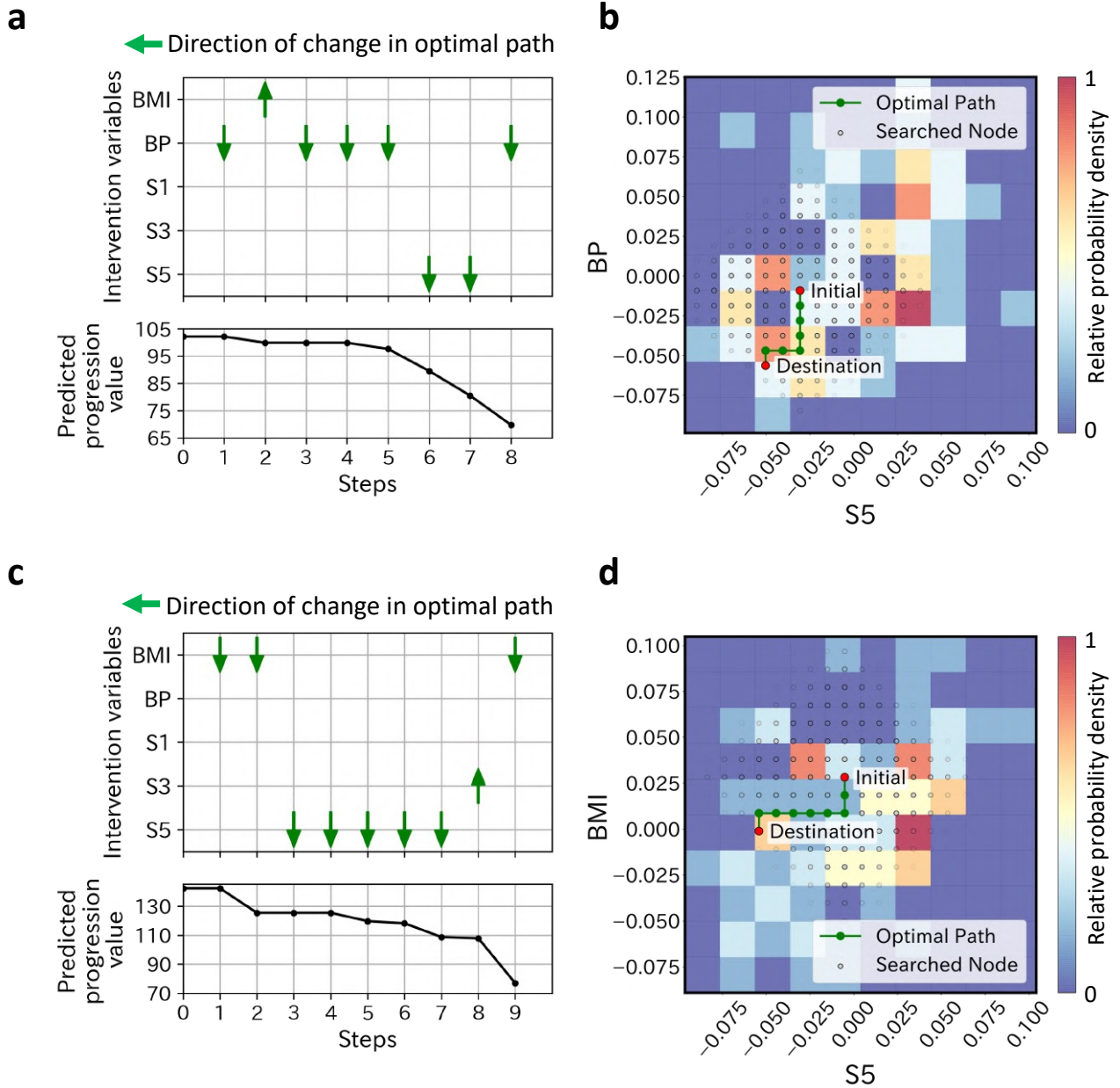
Supplementary Figure 44. Examples of personal actionable paths for treatment with intervention variables based on hypothesis-driven selection in chronic kidney disease (CKD) risk classification task. The optimal paths for improving the response variable predicted by the ML model are represented for two examples: instance 13 (a–c) and instance 14 (d–f). **a, d** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, c, e, f** 2D plots of the path. In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.



Supplementary Figure 45. Results of proposed framework on real estate dataset. **a** Plot for prediction vs. true response variable. **b** WBIC values of the stochastic surrogate models with 1–8 mixture components. **c** Histogram of actionability scores at different instances. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



Supplementary Figure 46. Results of proposed framework on diabetes progression dataset. a Feature importance of the regression model. The color of each bar represents the following: red: intervention variables in path planning, gray: variables which are difficult to be intervened, and blue: other variables. Details of features are described in Supplementary Table 7. **b** Plot for prediction vs. true response variable. **c** WBIC values of stochastic surrogate models with 1–8 mixture components. **d** Histogram of actionability scores for each instance. An actionability score of zero indicates that the actionability of the optimal path is equivalent to that of the baseline path.



Supplementary Figure 47. Examples of personal actionable paths for treatment on diabetes progression

dataset. The optimal paths needed for the improvement of the response variables predicted by the machine learning model are represented for randomly selected two examples: instance D1 (**a, b**) and instance D2 (**c, d**). **a, c** The orders of changes in the explanatory variables in the optimal path and the accompanying changes in the predicted values. In the transition steps, the upward or downward arrow represents a unit increase or decrease in the explanatory variable, respectively. **b, d** Two-dimensional (2D) plots of the path. The 2D plots are shown regarding the two influential variables: S5 and average blood pressure (BP) (**b**), and S5 and body mass index (BMI) (**d**). In the heatmaps, the probability density of the actual data, normalized by the panel with the maximum number of data, is expressed.

Supplementary Tables

Supplementary Table 1. The number of instances.

Task	Training	Test	Surrogate modeling
Regression task on SBP	10,260	2,543	2,222
Classification task on SBP (Hypertension risk)	10,308	2,495	2,185
Regression task on eGFR	10,233	2,563	2,230
Classification task on eGFR (CKD risk)	10,194	2,602	2,255

Supplementary Table 2. Correspondence between clinical guideline–recommended treatments for improving blood pressure and variables used in our framework.

Guideline recommended treatments	Examples of corresponding variables
Salt reduction	Serum sodium
Exercise	BMI Waist Leg score Blood glucose
Alcohol restriction	γ -GTP
Alleviation of obesity	BMI Waist
Nutrients and dietary patterns	γ -GTP Blood glucose

BMI: body mass index, γ -GTP: gamma-glutamyltransferase.

Supplementary Table 3. Clinicians' assessments on utility of health-improvement paths.

	Practical		Informative	
	Optimal path	Random path	Optimal path	Random path
Cardiologists	0.20 ± 0.21 [0, 5, 3, 2, 0]	0.22 ± 0.18 [1, 4, 2, 4, 0]	0.42 ± 0.30 [3, 8, 6, 4, 0]	0.38 ± 0.33 [1, 7, 4, 7, 0]
Nephrologists	0.30 ± 0.29 [1, 0, 2, 7, 5]	0.26 ± 0.23 [3, 0, 3, 1, 6]	0.52 ± 0.26 [2, 4, 6, 9, 5]	0.46 ± 0.19 [4, 2, 7, 6, 4]

Ratio of paths clinicians evaluated as practical or informative (Mean ± standard deviation). The numbers in brackets indicate the instances each clinician assessed as practical or informative in ten instances.

Supplementary Table 4. Correspondence between clinical guideline–recommended treatments for patients with lower renal functions and variables used in our framework.

Guideline recommended treatments	Examples of corresponding variables
Nutrients and dietary patterns	Weight BMR score Right arm X 50kHz Triglyceride
Lowering blood pressure (Salt reduction, anti-hypertensive drugs)	RADIA
Iron supplementation and other erythropoiesis-stimulating agents	Hb Erythrocyte count
Lipid-lowering treatment	Triglyceride
Alleviation of obesity Regular exercise	Weight BMR score Right arm X 50kHz
Lowering uric acid (Hyperuricemia drugs)	Uric acid

BMR score: basal metabolic rate score, RADIA: right ankle diastolic blood pressure, Hb: hemoglobin.

Supplementary Table 5. Classification model scores on chronic kidney disease (CKD) risk.

Model	AUC	TP	FN	FP	TN
XGBoost	0.844	35	171	19	2,377
Random Forest	0.833	9	197	4	2,392
Support vector machine	0.790	43	163	43	2,353

AUC: area under the curve, TP: true positive, FN: false negative, FP: false positive, TN: true negative.

Supplementary Table 6. Classification model scores on hypertension risk.

Model	AUC	TP	FN	FP	TN
XGBoost	0.772	127	515	98	1,755
Random Forest	0.771	72	570	59	1,794
Support vector machine	0.733	104	538	81	1,772

AUC: area under the curve, TP: true positive, FN: false negative, FP: false positive, TN: true negative.

Supplementary Table 7. Description of the diabetes progression dataset features.

Feature	Description
Response variable	
Disease progression	A quantitative measure of disease progression one year after baseline
Explanatory variables	
Age	Age in years
Sex	
BMI	Body mass index
BP	Average blood pressure
S1	T-cells (a type of white blood cells)
S2	Low-density lipoproteins
S3	High-density lipoproteins
S4	Thyroid stimulating hormone
S5	Lamotrigine
S6	Blood sugar level