CURRENT MEDICAL RESEARCH AND OPINION
2023, VOL. 39, NO. 11, 1463–1471
https://doi.org/10.1080/03007995.2023.2270420
Article ST-0541/2270420

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Prediction models for in-hospital deaths of patients with COVID-19 using electronic healthcare data

Kenichi Hiraga[a], Masato Takeuchi[a], Takeshi Kimura[b], Satomi Yoshida[a] and Koji Kawakami[a]

[a]Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan; [b]Research and Analytics Department, Real World Data Co., Ltd, Kyoto, Japan

ABSTRACT

Objective: Many models for predicting various disease prognoses have achieved high performance without laboratory test results. However, whether laboratory test results can improve performance remains unclear. This study aimed to investigate whether laboratory test results improve the model performance for coronavirus disease 2019 (COVID-19).

Methods: Prediction models were developed using data from the electronic healthcare record database in Japan. Patients aged ≥18 years hospitalized for COVID-19 after February 11, 2020, were included. Their age, sex, comorbidities, laboratory test results, and number of days from February 11, 2020, were collected. We developed a logistic regression, XGBOOST, random forest, and neural network analysis and compared the performance with and without laboratory test results. The performance of predicting in-hospital death was evaluated using the area under the curve (AUC).

Results: Data from 8,288 hospitalized patients (females, 46.5%) were analyzed. The median patient age was 71 years. A total of 6,630 patients were included in the training dataset, and 312 (4.7%) died. In the logistic regression model, the area under the curve was 0.88 (95% confidence interval [CI] = 0.83–0.93) and 0.75 (95% CI = 0.68–0.81) with and without laboratory test results, respectively. The performance was not fundamentally different between the model types, and the laboratory test results improved the performance in all cases. The variables useful for prediction were blood urea nitrogen, albumin, and lactate dehydrogenase.

Conclusions: Laboratory test results, such as blood urea nitrogen, albumin, and lactate dehydrogenase levels, along with background information, helped estimate the prognosis of patients hospitalized for COVID-19.

ARTICLE HISTORY
Received 15 July 2023
Revised 4 October 2023
Accepted 10 October 2023

KEYWORDS
Database; electronic health records; machine learning; model performance; prognosis

Introduction

Studies have attempted to develop models to predict the future onset of diseases or the prognosis of patients. Indeed, scoring systems based on studies developing prediction models, such as the Framingham coronary heart disease prediction scores and quick sequential organ failure assessment (qSOFA) scores for sepsis, are used in clinical practice[1,2].

Recently, studies on the development of prediction models have been increasing, with models using a database of real-world data, such as electronic healthcare records (EHR) or claims data[3–8]. In general, EHR and claims databases include large amounts of data on patients with various backgrounds. EHR databases usually include information on outcomes such as laboratory test results. However, when conducting studies using a database of real-world data, it is difficult for researchers to use data not included in the selected database. Therefore, it is important to select a database that contains sufficient data for each study.

Although laboratory test results are used to consider the diagnosis or prognosis of diseases in clinical practice, various prediction models predicting disease prognoses achieve high performance without laboratory test results, such as in cases of atrial fibrillation and chronic obstructive pulmonary disease[3,4]. However, this does not mean that laboratory test results are not required for high performance in all diseases. It is unclear whether laboratory test results can improve the performance of prediction models.

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2, has had a major impact worldwide[9–11]. Many models for predicting death in COVID-19 have been reported[8,12–16]. Some used laboratory test results[8,12–14], whereas others did not[15,16]. Previous studies have reported that various laboratory test results, such as elevated D-dimer levels, are risk factors for death in patients with COVID-19[17,18].

We hypothesized that the use of laboratory test results can improve the prediction model performance by using a database of real-world data and COVID-19. Accordingly, we

CONTACT Satomi Yoshida yoshida.satomi.4r@kyoto-u.ac.jp Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/03007995.2023.2270420.

aimed to quantitatively measure the usability of the information gained from laboratory test results in improving COVID-19 prediction model performance.

Methods

Study design and setting

The RWD database was used in this study. The RWD database is owned by the Health, Clinic, and Education Information Evaluation Institute (HCEI, Kyoto, Japan) and operated by Real World Data Co., Ltd. (Kyoto, Japan). The RWD database includes the demographic data, diagnoses, laboratory test results, medication prescriptions, and medical procedures of approximately 20 million patients (both inpatients and outpatients) from approximately 190 medical institutions across Japan as of July 2021[19,20]. The data collection began in 2015. The data in this database are continuously updated from the electronic medical records of each medical institution. The RWD database does not include the identification of individual information. This study was conducted in accordance with the principles of the Declaration of Helsinki. The Investigation and Ethics Committee of Kyoto University approved this study (Approval No. R2895-1; a study on prediction models for in-hospital deaths of patients with COVID-19). This study was initially approved on May 6, 2021, with a modified application approved on May 22, 2023.

Study population

Patients hospitalized for COVID-19 and aged $\geq$18 years were included. Patients with COVID-19 were defined as those assigned confirmed disease names corresponding to the International Classification of Diseases 10th Revision (ICD10) codes U071 or U072 on or after February 11, 2020 and before June 7, 2021, the administrative end of the database. Hospitalization due to COVID-19 was defined as hospitalization within 7 days before or after the ICD10 codes were assigned. Patients who met the inclusion criteria more than once were included in the study only once. Patients without any prescriptions, laboratory test results, or medical procedure data during hospitalization were excluded. Patients without a documented discharge date and those who died on the day of admission were excluded.

Variables and outcomes

We used baseline background features, including age, sex, smoking, body mass index, comorbidities, laboratory test results at the time of admission, and the number of days elapsed since the naming date of COVID-19 by the World Health Organization as exploratory variables. The detailed codes used to define comorbidities are summarized in Supplementary Table S1. We selected laboratory test items that are often measured during off-duty hours. To make the results available at the time of admission (or very early in the admission period), the explanatory variables that could be identified at that time were selected. Laboratory test results at the time of admission were defined as the results of the test performed closest to the date of admission among the tests performed from 7 days prior to admission to 3 days after admission for each item. Body mass index data were identified using the Diagnosis Procedure Combination (DPC) data. DPC is a payment approach used only in Japan and is based on case-mix classification[21]. Only acute care hospitals can choose DPC, and most acute care hospitals in Japan have adopted this payment scheme. The outcome was defined as in-hospital death in the primary analysis. In-hospital death was identified using a combination of the date of death, date of admission/discharge, and DPC data.

In the secondary analysis, outcomes were defined as admission to the intensive care unit (ICU), extracorporeal membrane oxygenation (ECMO), and invasive mechanical ventilation.

Model development

Descriptive statistics were calculated to summarize the baseline background features, comorbidities, laboratory test results, and features of the hospital in which each patient was hospitalized. The number of missing values for each variable was calculated.

We developed models to predict the outcomes of the primary and secondary analyses. We used all of the variables described above, except for the features of the hospital, as explanatory variables. Patients without values on smoking were regarded as non-smokers, and missing values in the continuous variables were complemented by a value calculated using a chained equation. In addition, to address abnormal outliers, values below the 0.05 quantile and values greater than the 0.95 quantile were rounded to the 0.05 quantile and the 0.95 quantile, respectively, for each continuous variable. We randomly split the data into training and test data at a ratio of 4:1. We developed logistic regression models, XGBOOST, random forest, and neural network models to predict the outcomes of the primary and secondary analyses with and without laboratory test values. The logistic regression model has been frequently used in studies that have developed prediction models with binary outcomes, and we assume that this model could serve as a benchmark. In the logistic regression model, the boundary between one class and another was assumed to be linear. However, this boundary is often non-linear in the real-world setting. Machine learning methods such as XGBOOST, random forest, and neural networks can solve the classification problem with a non-linear boundary. Given the possibility that outcomes may rarely occur, down-sampling for the training data combined with random forest, synthetic minority oversampling technique (SMOTE) combined with XGBOOST, and SMOTE combined with random forest were used in the analysis. The scale of the variables was standardized to develop the logistic regression and neural network models. In addition, in the case of XGBOOST and the neural network, we split all the data except the test data using a ratio of 4:1 into training data and data to evaluate early stopping to prevent overfitting. We trained the models with training data, and

the hyperparameters were tuned using grid search and random search using k-fold cross-validation ($k = 10$). K-fold cross-validation is a method in which training data are divided into k parts, one of which is used as validation data and the remaining k-1 pieces of data are used as training data, and the model performance for each set of hyperparameters is evaluated by iterating this process k times and integrating the results. The grid search searches for all combinations of enumerated parameters, whereas the random search selects combinations of parameters to search. After tuning the hyperparameters, except for the cases of models using down-sampling or SMOTE, the models were trained again using the entirety of the training data with hyperparameters corresponding to the best performance in the process of hyperparameter tuning.

We evaluated the performance of the models using the area under the receiver operating characteristic (ROC) curve (AUC) for the test data and its 95% confidence interval (CI). Moreover, the calibration curves of all models developed in the primary analysis are shown, and the Shapley additive explanation (SHAP) values are shown to visualize the contribution of each variable to the results. The SHAP is a popular framework for interpreting predictions and evaluating the predictive importance of each variable[22]. One of the main criticisms of machine learning is that it is unclear which variables contributed to the prediction. SHAP is a recently proposed method to overcome this problem, analogous to the odds ratios of logistic regression. In general, the higher the absolute value of the SHAP value, the more important the variable[22].

Finally, subgroup analyses were performed according to age and sex as females aged <50 years, males aged <50 years, females aged ≥50 years, and males aged ≥50 years, and we tried to define cutoff values of laboratory test items for each group. Laboratory test items corresponding to the top five absolute values of SHAP values in the case of at least one type of model were selected. The cutoff value was defined as the value corresponding to a point on the ROC curve nearest to the coordinate (0, 1) (upper left corner) for each laboratory test item. In addition, for each part of the split data, the sensitivity and specificity corresponding to the number of laboratory test items exceeding the cutoff values were calculated.

Python 3.7.6 was used for all data analyses. Additional information regarding the Python code used is provided in Supplementary Appendix 1.

Results

Characteristics of the study population

Patients with a diagnosis of COVID-19 between March 7, 2020 and May 27, 2021 were included. In total, 9,065 patients fulfilled the inclusion criteria and 777 patients were excluded. Finally, 8,288 patients were included in the analysis, of whom 3,854 (46.5%) were female; the median age was 71 years (interquartile range = 51–82 years). The number of patients for whom data were available after 2021 was 5,338 (64.4%). A summary of the explanatory variables and hospital characteristics is presented in Table 1.

Results of the primary analysis

In the case of XGBOOST and the neural network, data from 5,304 patients were assigned to the training data (1,326 to validation data and 1,658 to test data), and 250 (4.7%) patients in the training dataset died. For the other models, data from 6,630 patients were assigned to the training data (1,658 for test data), and 312 (4.7%) patients in the training dataset died.

First, we observed the results of the models that included laboratory test results. The AUC was 0.88 (95% CI = 0.83–0.93) in the case of logistic regression. The AUC for each model is shown in Table 2 and the ROC curve for each model is shown in Figure 1. The calibration curves for each model are shown in Figure 2. The SHAP values of each variable in the case of logistic regression, XGBOOST, random forest, and neural networks are shown in Figure 3.

Second, we presented the results of the models without laboratory test results as exploratory variables. The AUC was 0.75 (95% CI = 0.68–0.81) in the case of logistic regression, and the value of the AUC for each model is shown in Table 3.

Results of the secondary analysis

Finally, the results of the secondary analysis are presented. The number of patients assigned to the training data was 6,630, and the number of patients admitted to the ICU and those who received invasive mechanical ventilation were 5 (0.07%) and 25 (0.37%), respectively. None of the patients had received ECMO. In the cases of prediction of admission or transfer to the ICU and the induction of invasive mechanical ventilation, the values of the AUC were 0.92 (95% CI = 0.54–1.00) and 0.81 (95% CI = 0.61–1.00) with logistic regression, respectively.

Cut-off values of laboratory test results

Laboratory test items with the top five absolute values of SHAP values in at least one type of model were blood urea nitrogen (BUN), albumin, lactate dehydrogenase (LDH), blood glucose, aspartate aminotransferase (AST), D-dimer, prothrombin time, and C-reactive protein (CRP). The calculated cutoff values for these items in each age and sex subgroup are shown in Table 4. In addition, the ROC curve for each laboratory test item is shown in Supplementary Figure S1 in Supplementary Appendix 1. The combinations of sensitivity and specificity when three items among BUN, albumin, LDH, blood glucose, AST, D-dimer, prothrombin time, and CRP exceed the cutoff were calculated. For females aged <50 years, males aged <50 years, females aged ≥50 years, and males aged ≥50 years, the sensitivity and specificity of the combinations were 1.00 and 0.70; 1.00 and 0.69; 0.89 and 0.51; and 0.93 and 0.49, respectively. When six items exceeded the cutoff, the sensitivity and specificity of the

Table 1. Patient's background characteristics.

		Overall	Missing
Total number of patients		8,288	0
Age in years, median (Q1, Q3)		71.0 (51.0, 82.0)	0
Cancer, *n* (%)		2,254 (27.2)	0
Cardiovascular disease, *n* (%)		4,612 (55.6)	0
Chronic respiratory disease, *n* (%)		956 (11.5)	0
Liver disease, *n* (%)		1,113 (13.4)	0
Kidney disease, *n* (%)		578 (7.0)	0
Obesity, *n* (%)		36 (0.4)	0
Diabetes, *n* (%)		2,613 (31.5)	0
Women, *n* (%)		3,854 (46.5)	0
Smokers, *n* (%)		942 (11.4)	7,657
BMI*, median (Q1, Q3)		22.3 (19.8, 25.0)	5,681
CRP* (mg/dL), median (Q1, Q3)		1.4 (0.2, 6.1)	1,106
D-dimer (μg/mL), median (Q1, Q3)		1.6 (0.6, 4.7)	3,632
γ-GTP* (U/L), median (Q1, Q3)		26.0 (16.0, 53.0)	1,545
AST (GOT) * (U/L), median (Q1, Q3)		25.0 (19.0, 38.0)	886
ALT (GPT) * (U/L), median (Q1, Q3)		18.0 (12.0, 31.0)	895
ALP* (U/L), median (Q1, Q3)		226.0 (171.0, 304.0)	2,027
Albumin (g/dL), median (Q1, Q3)		3.7 (3.2, 4.1)	1,425
Potassium (mEq/L), median (Q1, Q3)		4.1 (3.8, 4.4)	891
Calcium (mg/dL), median (Q1, Q3)		8.8 (8.4, 9.2)	2,639
Creatinine (mg/dL), median (Q1, Q3)		0.8 (0.6, 1.1)	873
Creatinine kinase (U/L), median (Q1, Q3)		86.0 (54.0, 158.0)	1,547
Chloride (mEq/L), median (Q1, Q3)		104.0 (100.0, 106.0)	939
Blood glucose (mg/dL), median (Q1, Q3)		118.0 (98.0, 150.0)	2,248
Sodium (mEq/L), median (Q1, Q3)		139.0 (137.0, 141.0)	900
Prothrombin activity (%), median (Q1, Q3)		92.0 (78.6, 103.6)	2,492
Prothrombin time (s), median (Q1, Q3)		12.2 (11.4, 13.2)	3,542
Hematocrit (%), median (Q1, Q3)		38.3 (33.7, 42.5)	798
Hemoglobin (g/dL), median (Q1, Q3)		12.7 (11.0, 14.3)	798
LDH* (U/L), median (Q1, Q3)		225.0 (182.0, 292.8)	1,762
BUN* (mg/dL), median (Q1, Q3)		15.9 (11.2, 22.7)	880
Uric Acid (mg/dL), median (Q1, Q3)		5.1 (4.0, 6.4)	3,654
eGFR* (mL/min/1.73 m^2), median (Q1, Q3)		68.8 (51.0, 86.0)	930
Activated partial thromboplastin time (s), median (Q1, Q3)		30.4 (27.6, 34.2)	2,641
Lymphocytes (/μL), median (Q1, Q3)		1,027.8 (600.0, 1,620.0)	1,243
Monocytes (/μL), median (Q1, Q3)		402.5 (258.0, 584.1)	1,261
Neutrophils (/μL), median (Q1, Q3)		5,842.9 (3,747.1, 8,756.5)	2,310
Basophils (/μL), median (Q1, Q3)		20.0 (9.8, 40.0)	1,711
Eosinophils (/μL), median (Q1, Q3)		40.0 (10.0, 115.2)	1,397
White blood cell count (/μL), median (Q1, Q3)		7,100.0 (4,600.0, 9,900.0)	882
Total bilirubin (mg/dL), median (Q1, Q3)		0.7 (0.5, 1.0)	1,016
Total protein (g/dL), median (Q1, Q3)		6.8 (6.2, 7.2)	1,272
Platelet count (10^4/μL), median (Q1, Q3)		22.7 (17.2, 31.1)	795
Red blood cell count (10^4/μL), median (Q1, Q3)		408.0 (332.0, 465.0)	795
Number of days from the date of naming by WHO, median (Q1, Q3)		324.0 (220.0, 376.0)	0
Beds number, *n* (%)	100–299	700 (8.4)	0
	20–99	7 (0.1)	0
	300–499	2,402 (29.0)	0
	<20	1 (0.0)	0
	≥500	5,178 (62.5)	0
Hospital regions, *n* (%)	Chubu	944 (11.4)	0
	Chugoku	36 (0.4)	0
	Hokkaido	212 (2.6)	0
	Kanto	549 (6.6)	0
	Kinki	4,919 (59.4)	0
	Kyushu, Okinawa	1,345 (16.2)	0
	Shikoku	24 (0.3)	0
	Tohoku	259 (3.1)	0

Abbreviations: BMI, body mass index; CRP, C-reactive protein; γ-GTP, γ-glutamyl transpeptidase; AST (GOT), aspartate aminotransferase (glutamate oxaloacetate transaminase); ALT (GPT), alanine aminotransferase (glutamic pyruvic transaminase); ALP, alkaline phosphatase; LDH, lactate dehydrogenase; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate.

Table 2. AUC values of models with laboratory test data: estimated value (95% confidence interval).

	Logistic regression	XGBOOST (Grid search)	XGBOOST (Random search)	Random forest	Random forest and down-sampling	XGBOOST and SMOTE	Random forest and SMOTE	Neural network
Training data	0.87 (0.85–0.90)	0.93 (0.91–0.95)	0.97 (0.95–0.98)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.86 (0.83–0.89)
Test data	0.88 (0.83–0.93)	0.88 (0.83–0.93)	0.88 (0.83–0.93)	0.87 (0.82–0.92)	0.88 (0.83–0.93)	0.82 (0.77–0.88)	0.85 (0.79–0.90)	0.87 (0.82–0.92)

Abbreviations: AUC, area under the curve; SMOTE, synthetic minority over-sampling technique.

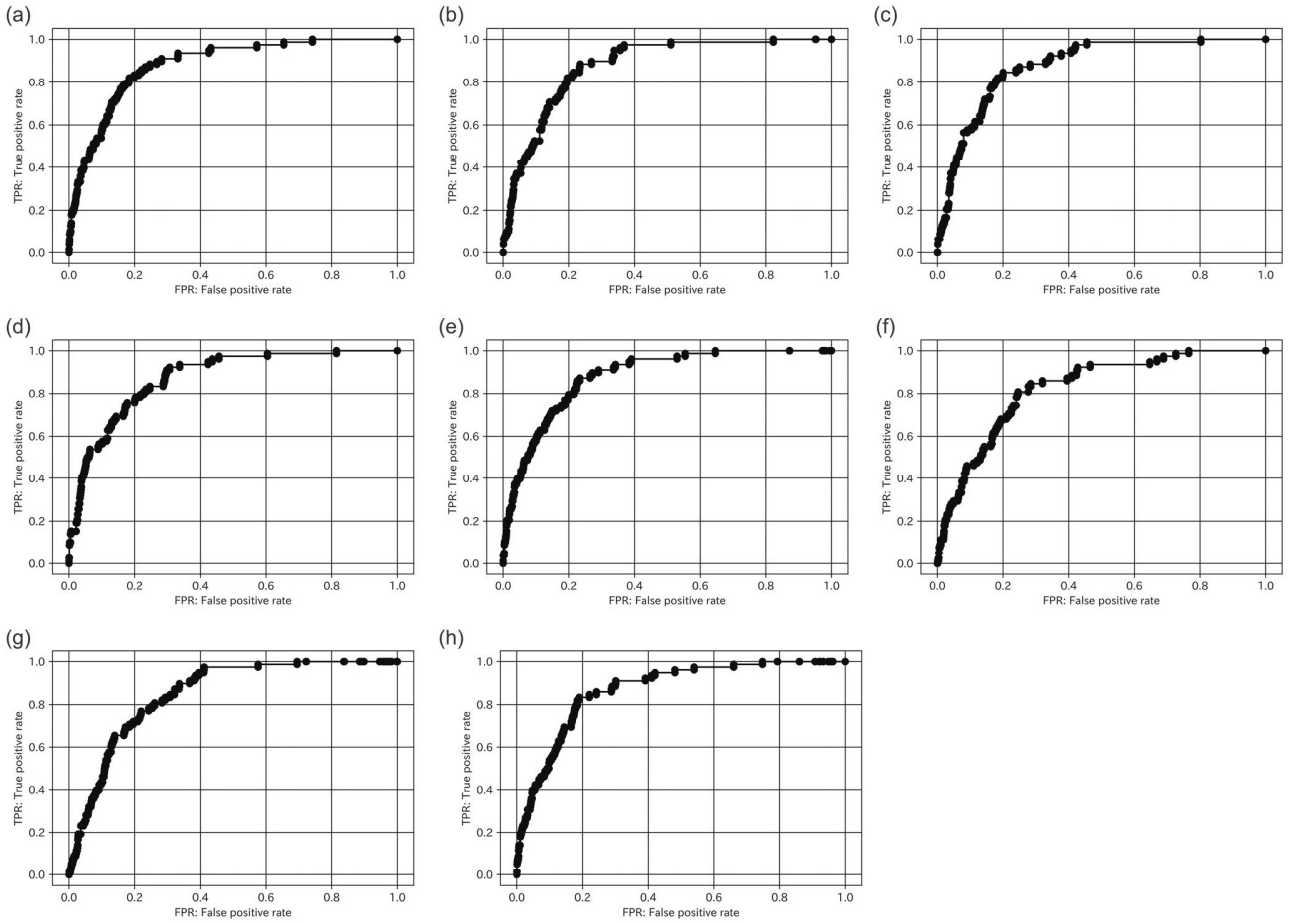


Figure 1. ROC curves of the mortality prediction models with laboratory test data in the case of (a) logistic regression, (b) XGBOOST alone with grid search, (c) XGBOOST alone with random search, (d) random forest alone, (e) down-sampling and random forest, (f) SMOTE and XGBOOST, (g) SMOTE and random forest, and (h) neural network.

Abbreviations. SMOTE, synthetic minority oversampling technique; ROC, receiver operating characteristic curve.

combinations were 0.75 and 0.98; 0.75 and 0.97; 0.47 and 0.93; and 0.41 and 0.91, respectively. The remaining results are shown in Supplementary Table S2.

Discussion

This study aimed to compare the performance of various prediction models (logistic regression, XGBOOST, random forest, and neural network) with and without laboratory test values, such as the mortality prediction of COVID-19. The performance of the models with laboratory test results was better than that of the models without laboratory test results, regardless of the model type. In addition, some laboratory test results reported in previous studies as risk factors for mortality in patients with COVID-19, including blood urea nitrogen (BUN), albumin, and lactate dehydrogenase (LDH), showed high absolute values of SHAP values in this study.

The AUCs of the models with laboratory test results as explanatory variables were higher than those of models without laboratory test results as explanatory variables in all cases: XGBOOST, random forest, neural network, a combination of down-sampling and random forest, a combination of SMOTE and random forest, and a combination of SMOTE and XGBOOST. Furthermore, although an excessive number of variables can cause overfitting to the training data, hyperparameters were tuned using the cross-validation method to prevent overfitting, and models with laboratory test data achieved higher performance than models without laboratory test results, even in the test data. Obtaining information regarding laboratory test results can change the impression about the diagnosis or prognosis of patients in clinical practice. The results of this study supported this intuition. Unlike other background information, such as age, sex, smoking, body mass index, and comorbidities, laboratory test results can markedly change within a short period of time. Thus, it is possible that laboratory test results can more sensitively capture changes in the patients' medical conditions, and this sensitivity leads to a high prediction performance.

Albumin, age, LDH, and BUN levels, in addition to cardiovascular disease, were included among the variables with the top five absolute SHAP values in the logistic regression analysis. Although in different orders, BUN, albumin, and LDH were among the variables with the top five absolute SHAP values in all cases of logistic regression, XGBOOST, random forest, and neural network. In XGBOOST with grid search, XGBOOST with random search, and random forest, age was among the top five variables in all cases, and the remaining

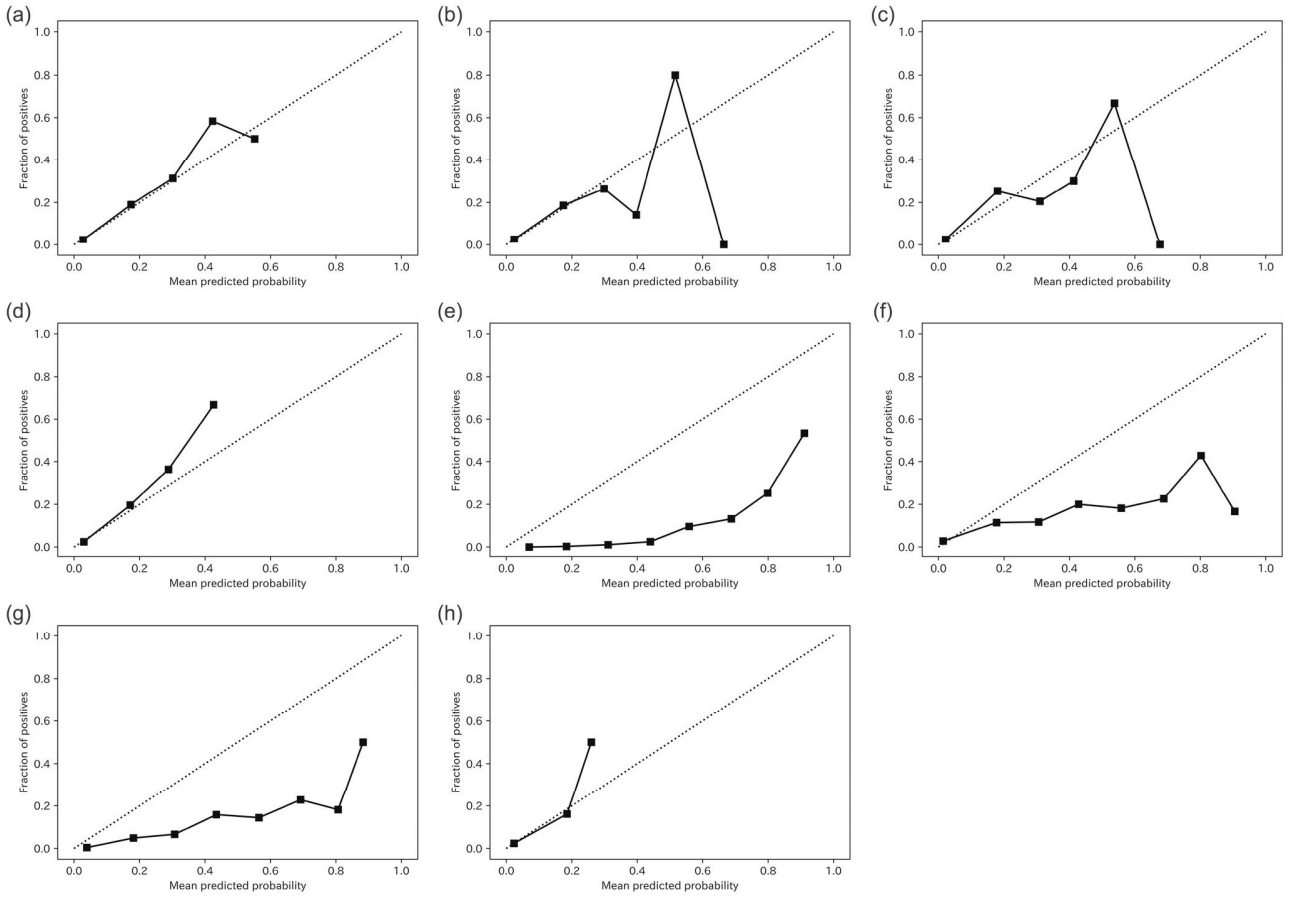


Figure 2. Calibration curves of the mortality prediction models with laboratory test data in the case of (a) logistic regression, (b) XGBOOST alone with grid search, (c) XGBOOST alone with random search, (d) random forest alone, (e) down-sampling and random forest, (f) SMOTE and XGBOOST, (g) SMOTE and random forest, and (h) neural network.
Abbreviation. SMOTE, synthetic minority oversampling technique.

ones were blood glucose level, aspartate aminotransferase (AST), and D-dimer. The remaining two of the top five variables in the neural network were prothrombin time and C-reactive protein (CRP). High BUN, low albumin, high LDH, high blood glucose, high AST, high D-dimer, and high CRP levels have been reported as risk factors for mortality in patients with COVID-19, and the directions of change in values (higher or lower than the normal range) were consistent with those previously reported[17,18]. The median patient age was 71 years, and the mortality rate in the training data was 4.7%. This mortality rate was higher than that of all patients in Japan reported in a survey by the Japanese Ministry of Health, Labor, and Welfare (approximately 0.2% as of September 10, 2022)[23]. Restriction to inpatients only and the high proportion of older patients may be responsible for the high mortality rate observed in this study.

Because death was a rare outcome in the training data (4.7%, 312/6,630), we combined down-sampling and random forest, SMOTE and random forest, and SMOTE and XGBOOST, which are generally suitable for rare outcomes[24,25]. However, the performances of these methods were not materially different from those of XGBOOST alone or random forest alone, and the calibration curves of these combined strategies deviated from a line representing accurate probability predictions. This feature of the calibration curve may be because under-sampling can change the prior distribution of the outcome and distort the probability estimates[26,27]. In general, bagging algorithms, such as random forest, have been reported to work better than boosting when combined with re-samplings, such as down-sampling or SMOTE[25]; however, no substantial difference was found in this study.

Some international collaboration studies have been conducted in settings similar to this study[28–31]. These studies reported that the patients' clinical features differed between the first and the second wave of the pandemic. This difference can cause the change in the important factors for predicting the prognosis of COVID-19 patients. However, a study for developing prediction models for death of COVID-19 patients using only data in 2020 showed that age, albumin level, AST level, creatine level, CRP level, and white blood cell count were important predictors of mortality. These factors overlapped with those in our study although we included later data. This similarity can imply that factors important for predicting COVID-19 mortality have not changed over time. Some study focused on the trend of laboratory test values and the symptoms. Laboratory test results only at the admission are used in this study, and the RWD database does not have information about symptoms.

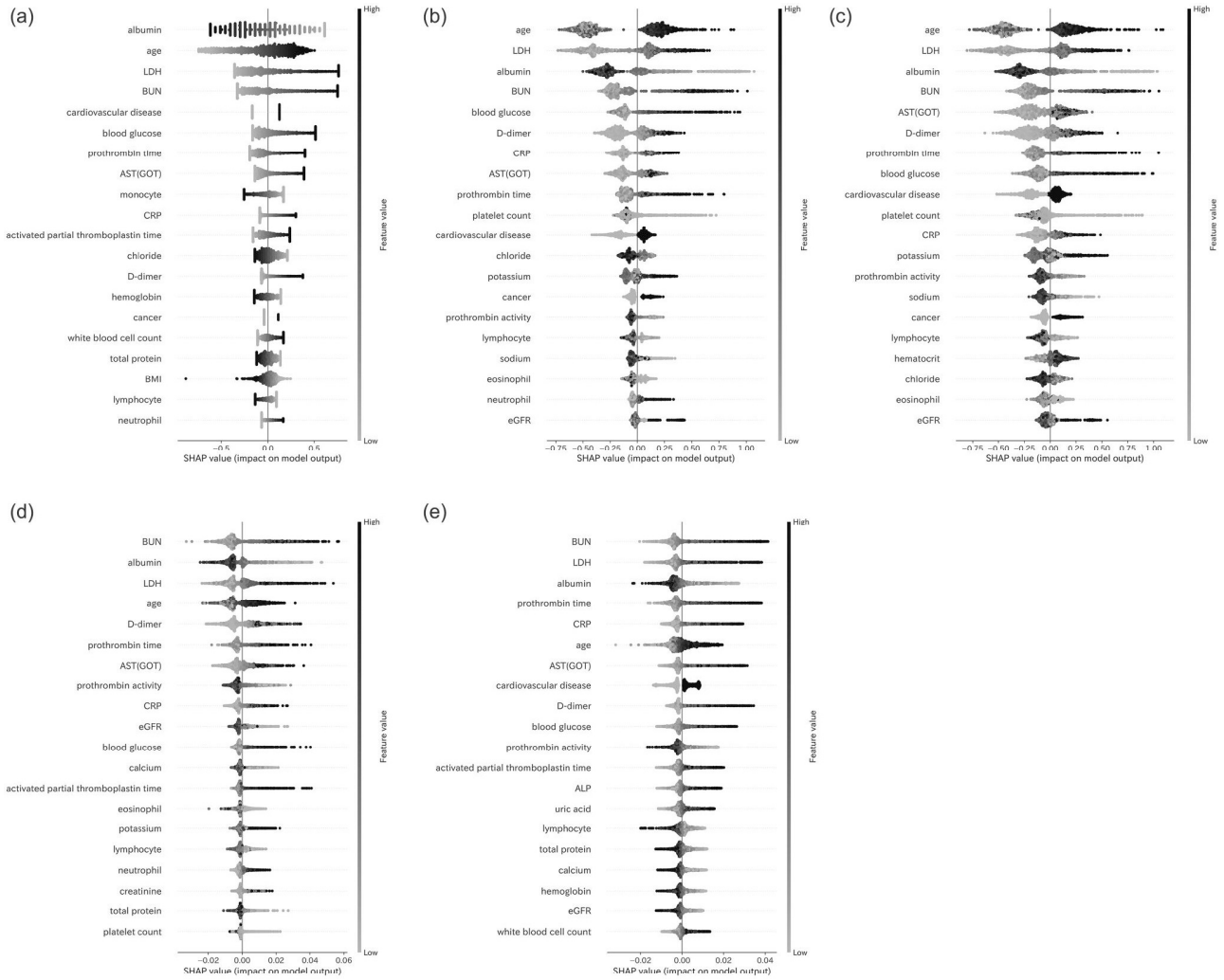


Figure 3. SHAP values of the mortality prediction models with laboratory test data in the case of (a) logistic regression, (b) XGBOOST with grid search, (c) XGBOOST with random search, (d) random forest, and (e) neural networks.
Abbreviations. SHAP, Shapley additive explanation; LDH, lactate dehydrogenase; BUN, blood urea nitrogen; AST (GOT), aspartate aminotransferase (glutamate oxaloacetate transaminase); CRP, C-reactive protein; BMI, body mass index; eGFR, estimated glomerular filtration rate; ALP, alkaline phosphatase.

Table 3. AUC values of models without laboratory test data: estimated value (95% confidence interval).

	Logistic regression	XGBOOST (Random search)	Random forest	Random forest and down-sampling	XGBOOST and SMOTE	Random forest and SMOTE	Neural network
Training data	0.74 (0.71–0.77)	0.78 (0.75–0.81)	0.84 (0.81–0.87)	0.95 (0.93–0.97)	1.00 (1.00–1.00)	0.97 (0.97–0.98)	0.73 (0.70–0.77)
Test data	0.75 (0.68–0.81)	0.74 (0.68–0.81)	0.74 (0.67–0.80)	0.75 (0.69–0.81)	0.72 (0.65–0.78)	0.72 (0.66–0.79)	0.74 (0.67–0.80)

Abbreviations: AUC, area under the curve; SMOTE, synthetic minority over-sampling technique.

However, whether these factors improve the performance of models is one of the topics to be examined next.

This study has several limitations. First, a selection bias may exist because not all medical institutions in Japan participate in the RWD database. Thus, external validity should be tested using data collected in different settings. Second, it was difficult to collect additional items that were not included in the database, such as vital signs, radiological imaging results, and patients' social backgrounds, because this study used existing data. However, high performance can be achieved using only the available data.

Conclusions

The prediction models of COVID-19 mortality show better performance when laboratory test results are included in the model than when they are not. Moreover, high BUN, low albumin, high LDH, high blood glucose, high AST, high D-dimer, high prothrombin time, and high CRP levels were identified as risk factors that can significantly contribute to mortality in patients with COVID-19. It might be possible to estimate the prognosis of patients hospitalized for COVID-19 with higher performance if laboratory test results, such as

Table 4. Cut-off values of laboratory test values.

	Females aged <50 years ($n = 1,081$)			Males aged <50 years ($n = 877$)		
	Number of deaths = 4			Number of deaths = 8		
	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
BUN (mg/dL)	17.2	0.75	0.95	19.2	0.50	0.91
albumin (g/dL)	3.4	0.75	0.65	3.9	0.88	0.75
LDH (U/L)	223.0	1.00	0.76	232.0	0.88	0.68
blood glucose (mg/dL)	96.0	0.50	0.47	153.0	0.75	0.90
AST (GOT) (U/L)	31.0	1.00	0.84	34.0	0.88	0.70
D-dimer (μg/mL)	1.2	0.75	0.54	1.6	0.88	0.80
prothrombin time (s)	14.1	0.75	0.95	13.0	0.88	0.84
CRP (mg/dL)	9.11	1.00	0.94	0.83	0.38	0.57
	Females aged ≥50 years ($n = 2,773$)			Males aged ≥50 years ($n = 3,557$)		
	Number of deaths = 150			Number of deaths = 228		
	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
BUN (mg/dL)	21.4	0.68	0.72	22.0	0.66	0.67
albumin (g/dL)	3.15	0.60	0.75	3.3	0.67	0.67
LDH (U/L)	260.0	0.64	0.65	259.0	0.68	0.63
blood glucose (mg/dL)	141.0	0.52	0.67	134.0	0.50	0.61
AST(GOT) (U/L)	32.0	0.61	0.68	39.0	0.54	0.73
D-dimer (μg/mL)	2.8	0.74	0.58	2.8	0.68	0.65
prothrombin time (s)	12.4	0.63	0.64	13.0	0.64	0.66
CRP (mg/dL)	3.77	0.61	0.67	4.86	0.64	0.66

Cut-off: For albumin, the value in the cell is the maximum value in the abnormal range, and for others, it is the minimum value in the abnormal range. For example, for females aged <50 years, BUN ≥17.2 mg/dL and albumin ≤3.4 g/dL are abnormal ranges.
Abbreviations: CRP, C-reactive protein; AST (GOT), aspartate aminotransferase (glutamate oxaloacetate transaminase); LDH, lactate dehydrogenase; BUN, blood urea nitrogen.

BUN, albumin, LDH, blood glucose, AST, D-dimer, prothrombin time, and CRP, are considered in clinical practice.

Transparency

Declaration of funding

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

Declaration of financial/other relationship

Kenichi Hiraga is a paid consultant of Real World Data Co., Ltd. Masato Takeuchi declares no conflicts of interest. Takeshi Kimura is an employee of Real World Data Co., Ltd. Satomi Yoshida was employed by the Department of Digital Health and Epidemiology, an Industry-Academia Collaboration Course supported by Eisai Co., Ltd., Kyowa Kirin Co., Ltd., Real World Data Co., Ltd., and Mitsubishi Corporation. Satomi Yoshida has also received consulting fees from Real World Data Co., Ltd. Koji Kawakami has received research funds from Eisai Co., Ltd., Kyowa Kirin Co., Ltd., Mitsubishi Corporation, OMRON Corporation, Real World Data Co., Ltd., Sumitomo Pharma Co., Ltd., and Toppan Inc.; consulting fees from Advanced Medical Care Inc., JMDC Inc., LEBER Inc., and Shin Nippon Biomedical Laboratories Ltd.; executive compensation from Cancer Intelligence Care Systems, Inc.; honoraria from Chugai Pharmaceutical Co., Ltd., Kaken Pharmaceutical Co., Ltd., Mitsubishi Chemical Holdings Corporation, Mitsubishi Corporation, Pharma Business Academy, and Toppan Inc.; and held stock in Real World Data Co., Ltd. The data used in this study were provided free of charge by the Real World Data Co., Ltd. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

KH, MT, and TK contributed to the planning, conducting, and reporting of the work described in the article. SY and KK ensured that the reporting was standard and scientific. KH analyzed the data, and all authors interpreted the results. All authors collaborated in the drafting and critical revision of the manuscript. All authors approved the final version of the manuscript and vouched for the accuracy of the analyses and adherence to the protocol.

Acknowledgements

Authors would like to thank Editage (www.editage.com) for the English language editing. We thank the Health, Clinic, and Education Information Evaluation Institute for the database development and operation.

Data availability statement

Data distribution is not allowed by the data provider.

ORCID

Satomi Yoshida http://orcid.org/0000-0002-3411-9427

References

[1] D'Agostino RB, Sr, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA. 2001;286(2):180–187. doi: 10.1001/jama.286.2.180.

[2] Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315(8):801–810. doi: 10.1001/jama.2016.0287.

[3] Hulme OL, Khurshid S, Weng LC, et al. Development and validation of a prediction model for atrial fibrillation using electronic health records. JACC Clin Electrophysiol. 2019;5(11):1331–1341. doi: 10.1016/j.jacep.2019.07.016.

[4] Macaulay D, Sun SX, Sorg RA, et al. Development and validation of a claims-based prediction model for COPD severity. Respir Med. 2013;107(10):1568–1577. doi: 10.1016/j.rmed.2013.05.012.

[5] Nguyen OK, Washington C, Clark CR, et al. Man vs. machine: comparing physician vs. electronic health record-based model predictions for 30-day hospital readmissions. J Gen Intern Med. 2021; 36(9):2555–2562. doi: 10.1007/s11606-020-06355-3.

[6] Le S, Allen A, Calvert J, et al. Convolutional neural network model for intensive care unit acute kidney injury prediction. Kidney Int Rep. 2021;6(5):1289–1298. doi: 10.1016/j.ekir.2021.02.031.

[7] Panahiazar M, Taslimitehrani V, Pereira N, et al. Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inform. 2015;216:40–44.

[8] He F, Page JH, Weinberg KR, et al. The development and validation of simplified machine learning algorithms to predict prognosis of patients with COVID-19: multicenter, retrospective study. J Med Internet Res. 2022;24(1):e31549. doi: 10.2196/31549.

[9] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–1242. doi: 10.1001/jama.2020.2648.

[10] Khan M, Adil SF, Alkhathlan HZ, et al. COVID-19: a global challenge with old history, epidemiology and progress so far. Molecules. 2020;26(1):39. doi: 10.3390/molecules26010039.

[11] Sharma A, Farouk IA, Lal SK. COVID-19: a review on the novel coronavirus disease evolution, transmission, detection, control and prevention. Viruses. 2021;13(2):202. doi: 10.3390/v13020202.

[12] Berenguer J, Borobia AM, Ryan P, et al. Development and validation of a prediction model for 30-day mortality in hospitalised patients with COVID-19: the COVID-19 SEIMC score. Thorax. 2021; 76(9):920–929. doi: 10.1136/thoraxjnl-2020-216001.

[13] Gao Y, Cai GY, Fang W, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun. 2020;11(1):5033. doi: 10.1038/s41467-020-18684-2.

[14] Guan X, Zhang B, Fu M, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study. Ann Med. 2021;53(1):257–266. doi: 10.1080/07853890.2020.1868564.

[15] Moon HJ, Kim K, Kang EK, et al. Prediction of COVID-19-related mortality and 30-day and 60-day survival probabilities using a nomogram. J Korean Med Sci. 2021;36(35):e248. doi: 10.3346/jkms.2021.36.e248.

[16] Yadaw AS, Li YC, Bose S, et al. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. Lancet Digit Health. 2020;2(10):e516–e525. doi: 10.1016/S2589-7500(20)30217-X.

[17] Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. BMC Infect Dis. 2021;21(1):855. doi: 10.1186/s12879-021-06536-3.

[18] Marin BG, Aghagoli G, Lavine K, et al. Predictors of COVID-19 severity: a literature review. Rev Med Virol. 2021;31(1):1–10. doi: 10.1002/rmv.2146.

[19] Hashimoto H, Takeuchi M, Kawakami K. Association between biopsies for anti-neutrophil cytoplasmic antibody-associated vasculitis and prognosis: a retrospective cohort study. Clin Rheumatol. 2022;41(2):541–548. doi: 10.1007/s10067-021-05889-z.

[20] Takeuchi M, Ogura M, Inagaki N, et al. Initiating SGLT2 inhibitor therapy to improve renal outcomes for persons with diabetes eligible for an intensified glucose-lowering regimen: hypothetical intervention using parametric g-formula modeling. BMJ Open Diabetes Res Care. 2022;10(3):e002636. doi: 10.1136/bmjdrc-2021-002636.

[21] Matsuda S. Development of case mix based evaluation system in Japan. Jpn Hosp. 2016;35:35–44.

[22] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Von Luxburg U, Bengio S, et al., editors. Advances in neural information processing systems. Long Beach Convention Center, Long Beach, (CA): the Neural Information Processing Systems Foundation, Inc. (NeurIPS); 2017. p. 4768–4777.

[23] The website of the Japanese Ministry of Health, Labor and Welfare [Internet]. Visualizing the data: information on COVID-19 infections; [cited 2022 Sep 10]. Available from: https://covid19.mhlw.go.jp/en/

[24] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–357. doi: 10.1613/jair.953.

[25] Galar M, Fernandez A, Barrenechea E, et al. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid- based approaches. IEEE Trans Syst Man Cybern C. 2012; 42(4):463–484. doi: 10.1109/TSMCC.2011.2161285.

[26] Dal Pozzolo A, Caelen O, Johnson RA, et al. Calibrating probability with undersampling for unbalanced classification. IEEE Symposium on Computational Intelligence in Data Mining, Cape Town, South Africa, 2015; p. 159–66.

[27] Pozzolo AD, Caelen O, Bontempi G. When is undersampling effective in unbalanced classification tasks?. Conference Proceedings on Machine Learning and Knowledge Discovery in Databases, Brussels, Belgium, September 2015.

[28] Weber GM, Zhang HG, L'Yi S, et al. International changes in COVID-19 clinical trajectories across 315 hospitals and 6 countries: retrospective cohort study. J Med Internet Res. 2021;23(10): e31400. doi: 10.2196/31400.

[29] Weber GM, Hong C, Xia Z, et al. International comparisons of laboratory values from the 4CE collaborative to predict COVID-19 mortality. NPJ Digit Med. 2022;5(1):74. doi: 10.1038/s41746-022-00601-0.

[30] Bourgeois FT, Gutiérrez-Sacristán A, Keller MS, et al. International analysis of electronic health records of children and youth hospitalized with COVID-19 infection in 6 countries. JAMA Netw Open. 2021;4(6):e2112596. doi: 10.1001/jamanetworkopen.2021.12596.

[31] Estiri H, Strasser ZH, Brat GA, et al. Evolving phenotypes of non-hospitalized patients that indicate long COVID. BMC Med. 2021; 19(1):249. doi: 10.1186/s12916-021-02115-0.

Prediction Models for In-Hospital Deaths of Patients with COVID-19 Using Electronic Healthcare Data

Kenichi Hiraga M.D.[1], Masato Takeuchi M.D.[1], Takeshi Kimura M.D.[2], Satomi Yoshida Ph.D.[1*], Koji Kawakami M.D.[1]

[1]Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan; [2]Research and Analytics Department, Real World Data Co., Ltd., Kyoto, Japan

*Corresponding author:

Satomi Yoshida (ORCID ID:0000-0002-3411-9427)

Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

Email ID: yoshida.satomi.4r@kyoto-u.ac.jp

Phone: +81-75-753-9469
Fax: +81-75-753-4469

TRIPOD Checklist: Prediction Model Development

Section/Topic		Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	2
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3–4
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	4
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	4
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	4–5
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	4
	5b	Describe eligibility criteria for participants.	5
	5c	Give details of treatments received, if relevant.	Not applicable
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	5–6
	6b	Report any actions to blind assessment of the outcome to be predicted.	6
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	5–7, Table 1
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	6-7
Sample size	8	Explain how the study size was arrived at.	4–5
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	6, Table 1
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	6–8
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	6–8
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	6–8
Risk groups	11	Provide details on how risk groups were created, if done.	Not applicable
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	8–9
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	8–9, Table 1
Model development	14a	Specify the number of participants and outcome events in each analysis.	8–10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	None
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	None

	15b	Explain how to the use the prediction model.	9-10
Model performance	16	Report performance measures (with CIs) for the prediction model.	9–10, Figure 1, Figure 2, Table 2, Table 3
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	11–12
Implications	20	Discuss the potential clinical use of the model and implications for future research.	11–12
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Supplementary Material
Funding	22	Give the source of funding and the role of the funders for the present study.	15

eTable 1. Code list (ICD10)

Name	Code
COVID-19	U071 or U072
Cancer	C00–C97
Cardiovascular disease	I00–99
Chronic respiratory disease	J40–47
Liver disease	K70–77, C22, B15–19
Kidney disease	N17–19
Obesity	E65–68
Diabetes mellitus	E10–14

COVID-19, coronavirus disease; ICD10 International Classification of Diseases 10th Revision

eTable 2. Sensitivity and specificity corresponding to number of items among BUN,

albumin, LDH, blood glucose, AST, D-dimer, prothrombin time, and CRP exceeding the cut-off value.

	Females aged <50 years		Males aged <50 years		Females aged ≥50 years		Males aged ≥50 years	
Number of items	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
1	1.00	0.16	1.00	0.25	1.00	0.13	1.00	0.13
2	1.00	0.44	1.00	0.49	0.96	0.31	0.97	0.30
3	1.00	0.70	1.00	0.69	0.89	0.51	0.93	0.49
4	1.00	0.88	0.88	0.83	0.77	0.69	0.80	0.67
5	0.75	0.96	0.88	0.92	0.65	0.83	0.61	0.81
6	0.75	0.98	0.75	0.97	0.47	0.93	0.41	0.91
7	0.75	1.00	0.50	0.99	0.23	0.97	0.21	0.97
8	0.25	1.00	0.00	1.00	0.06	1.00	0.08	0.99

Abbreviations

CRP: C-reactive protein, AST (GOT): aspartate aminotransferase (glutamate oxaloacetate transaminase), LDH: lactate dehydrogenase, BUN: blood urea nitrogen

Range of searched hyperparameters and tuning method (Python code)

Logistic Regression

Model: Model = sklearn.linear_model.LogisticRegression (max_iter=10000, random_state=42)

Hyperparameter Ranges: {"C": [10 ** i for i in range(-5, 6)]}

Method for Cross Validation: GridSearchCV(model, Hyperparameter Ranges, cv=KFold(10, random_state = 42, shuffle=True), scoring="neg_log_loss")

XGBOOST (GridSearch)

Model: Model = xgboost.XGBClassifier (random_state=42)

Hyperparameter Ranges:

```
{ 'objective':['binary:logistic'],
   'max_depth': [4, 8, 16],
   'alpha': [0.1, 1, 10, 100],
   'learning_rate': [0.1],
   'n_estimators': [200],
   'eval_metric': ['logloss'],
   'random_state': [42],
   'use_label_encoder':[False]  }
```

Method for Cross Validation: GridSearchCV(Model, Hyperparameter Ranges, cv=KFold(10, random_state = 42, shuffle=True), scoring="neg_log_loss")

XGBOOST (XGBOOST + SMOTE) (Random Search)

Model: Model = xgboost.XGBClassifier(random_state=42)

Hyperparameter Ranges:

{ 'objective':['binary:logistic'],

'max_depth': [i for i in range (1,21)],

'alpha': [0.01*2*i for i in range (1, 501)],

'learning_rate': [0.1],

'n_estimators':[200],

'eval_metric': ['logloss'],

'random_state' : [42],

'use_label_encoder':[False] }

Method for Cross Validation:

RandomizedSearchCV(Model, Hyperparameter Ranges, cv=KFold(10, random_state = 42, shuffle=True), n_iter = 30, scoring="neg_log_loss", random_state=42)

Random Forest (Random Forest + SMOTE, Random Forest + Down Sampling)

Model: Model = sklearn.ensemble. RandomForestClassifier(random_state=42)

Hyperparameter Ranges:

{ 'criterion' : ['gini', 'entropy'],

'n_estimators': [10, 50, 100, 200],

'max_depth' : [3, 5, 7, 9, 11] }

Method for Cross Validation:

RandomizedSearchCV(Model, Hyperparameter Ranges, cv=KFold(10, random_state = 42, shuffle=True), n_iter = 30, scoring="neg_log_loss", random_state=42)

Neural network

Model:

```
def build_model (n_hidden=1, n_neurons=10, learning_rate=1e-3, input_shape=XX, activation="relu", optimizer_type="adam"):
    model = tf.keras.models.Sequential()
    model.add(tf.keras.layers.Dense(n_neurons, activation=activation, input_dim=input_shape))
    for layer in range(n_hidden):
        model.add(tf.keras.layers.Dense(n_neurons, activation=activation))
    model.add(tf.keras.layers.Dense(1, activation='sigmoid'))

    if optimizer_type=="adam":
        optimizer = tf.keras.optimizers.Adam(learning_rate=learning_rate)
    else:
        optimizer = tf.keras.optimizers.SGD(learning_rate=learning_rate)

    model.compile(optimizer=optimizer,
                  loss='binary_crossentropy',
                  metrics=[tf.keras.metrics.AUC()])

    return model
Model = tensorflow.keras.wrappers.scikit_learn.KerasClassifier(build_model)
```

Hyperparameter Ranges:

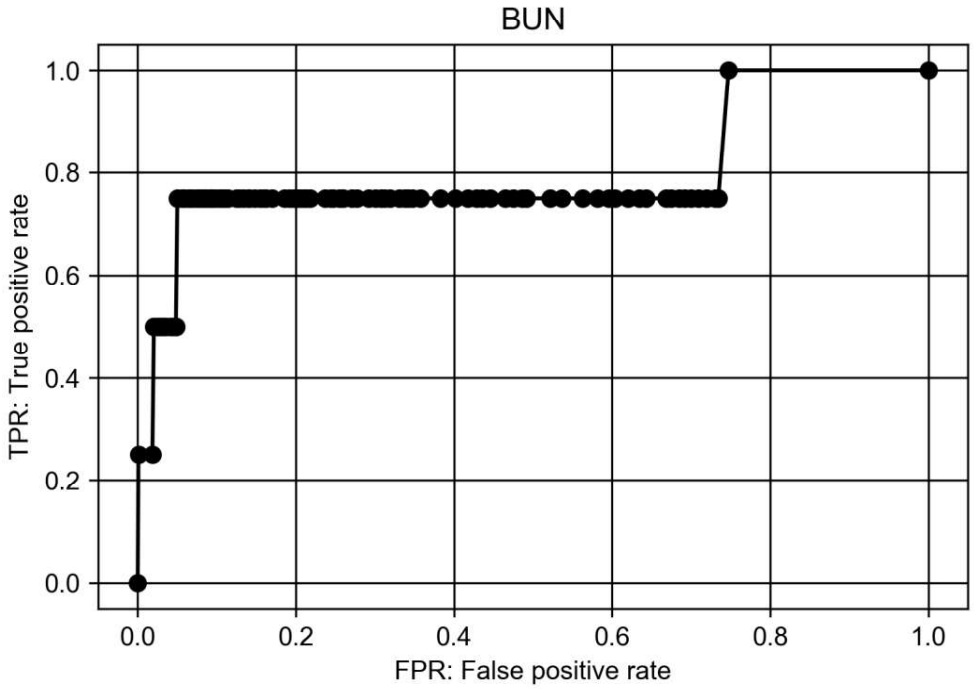
```
{ "n_hidden" : list(range(1, 6)),
   "n_neurons" : [10, 20, 50, 100],
   "learning_rate" : [10**(-i) for i in range(1, 6)],
   "activation" : ["relu", "sigmoid"],
   "optimizer_type" : ["adam"],
   "input_shape" : [X_train.shape[1]] }
```

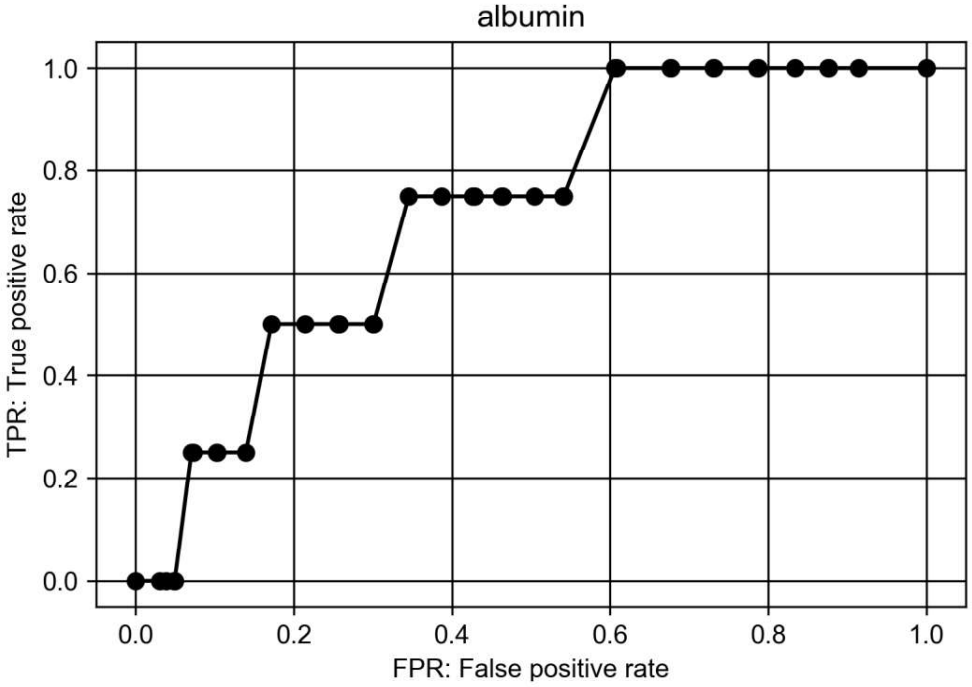
Method for Cross Validation:

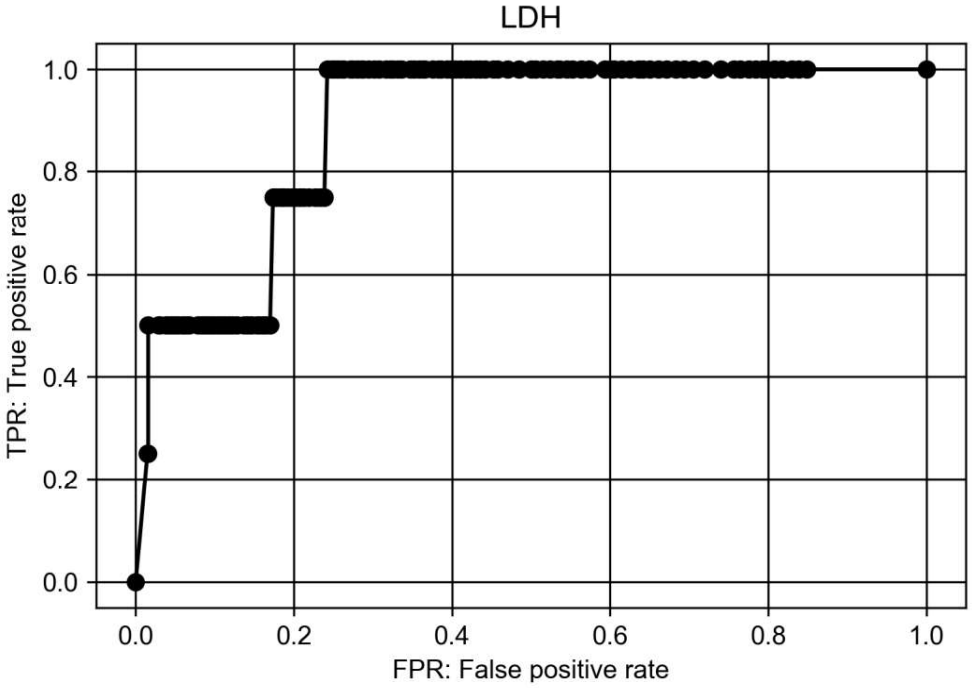
RandomizedSearchCV(Model, Hyperparameter Ranges, cv=KFold(10, random_state = 42, shuffle=True), n_iter = 10, scoring="neg_log_loss", random_state=42)

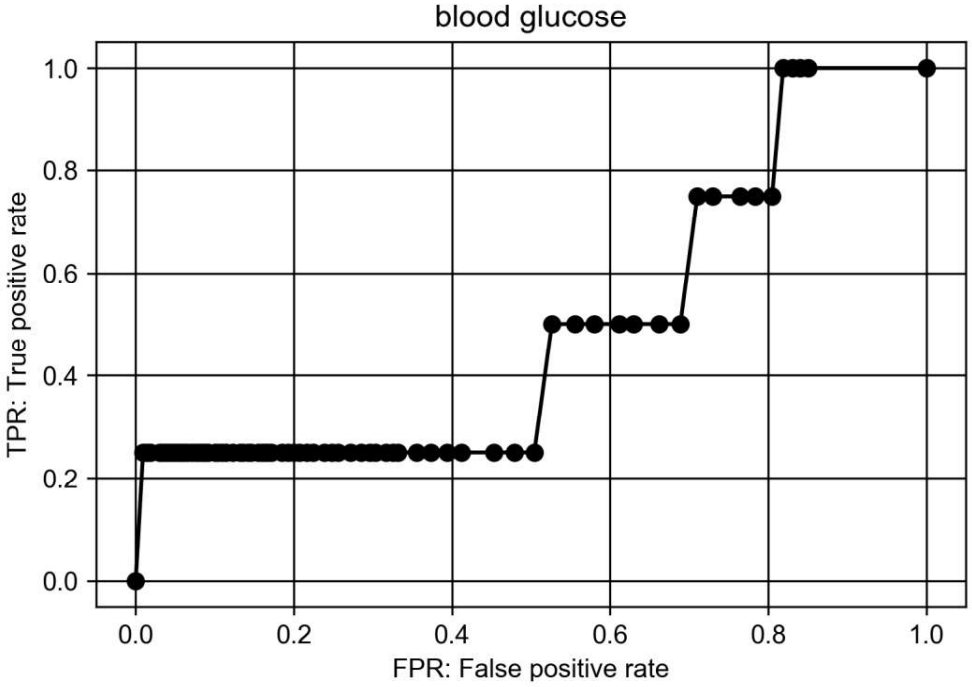
Supplementary Figure 1. ROC curve for each laboratory test item

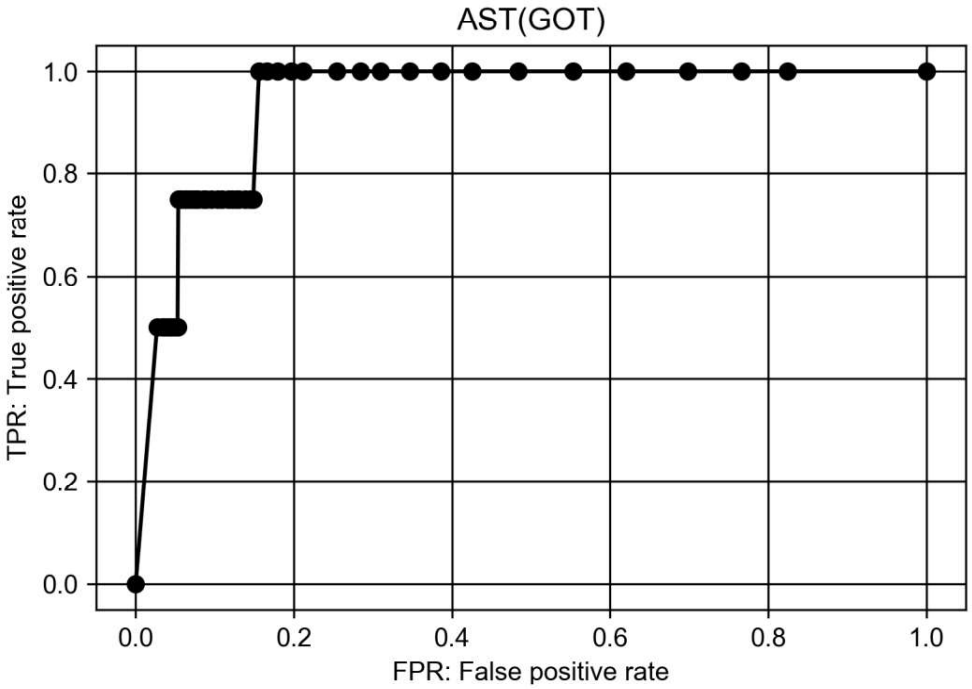
(a) Female aged <50 years

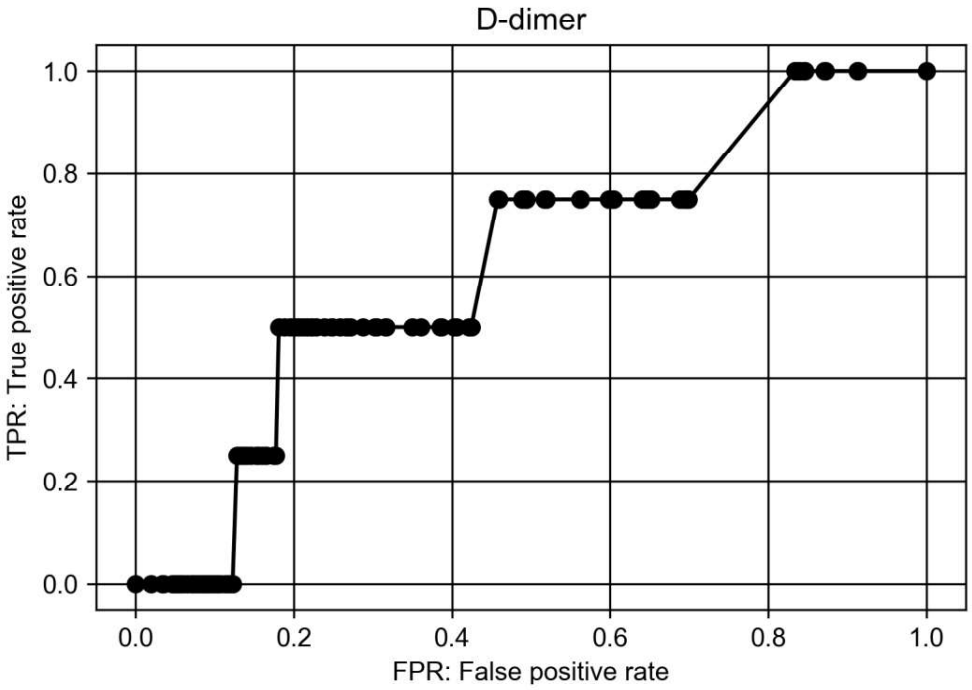


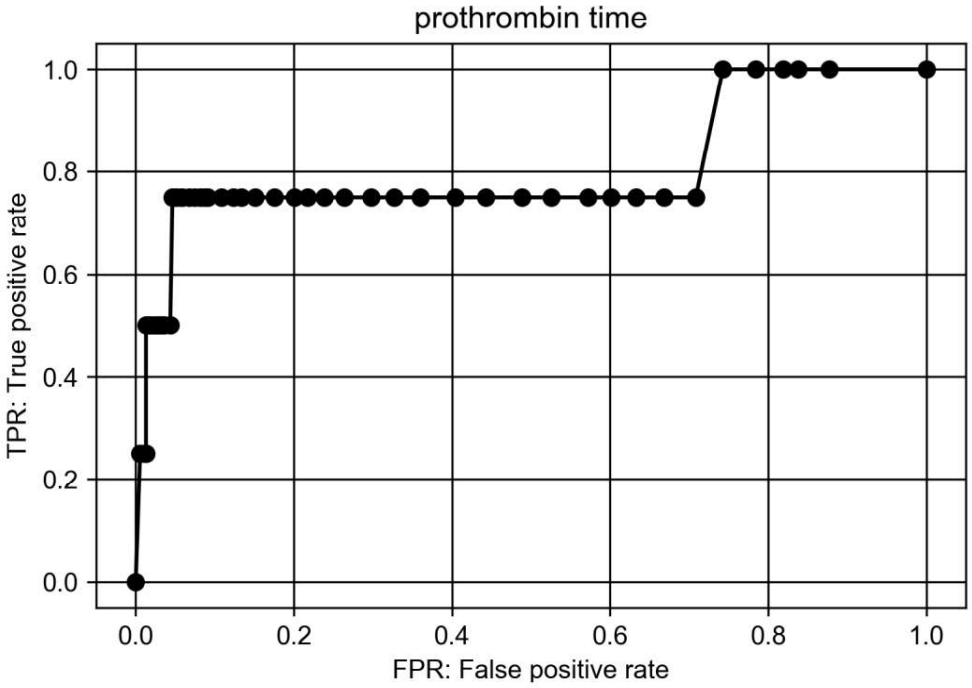

albumin
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

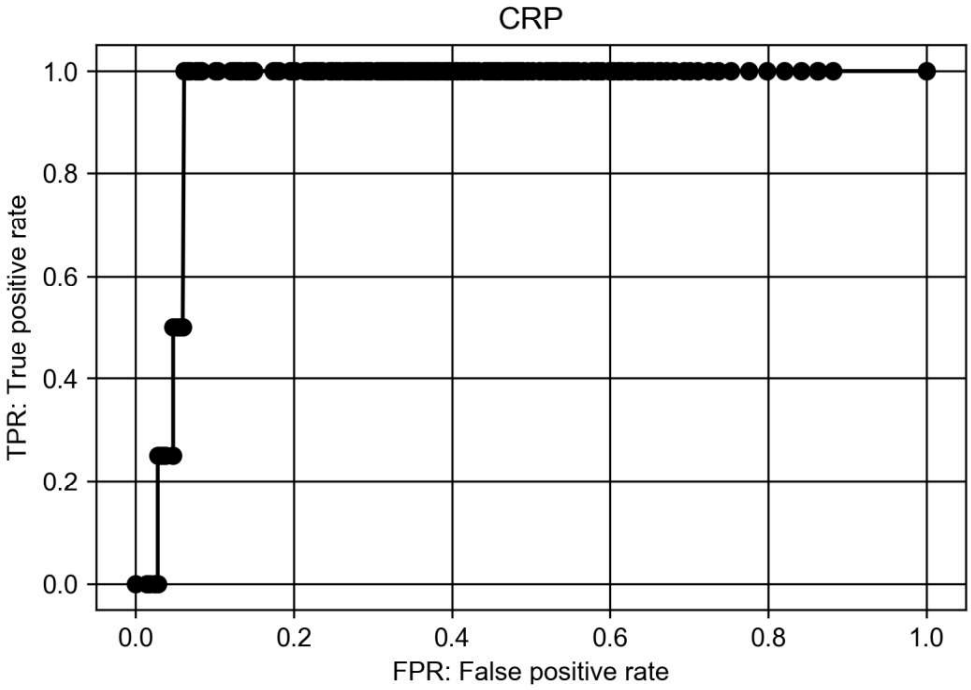

LDH
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


blood glucose
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

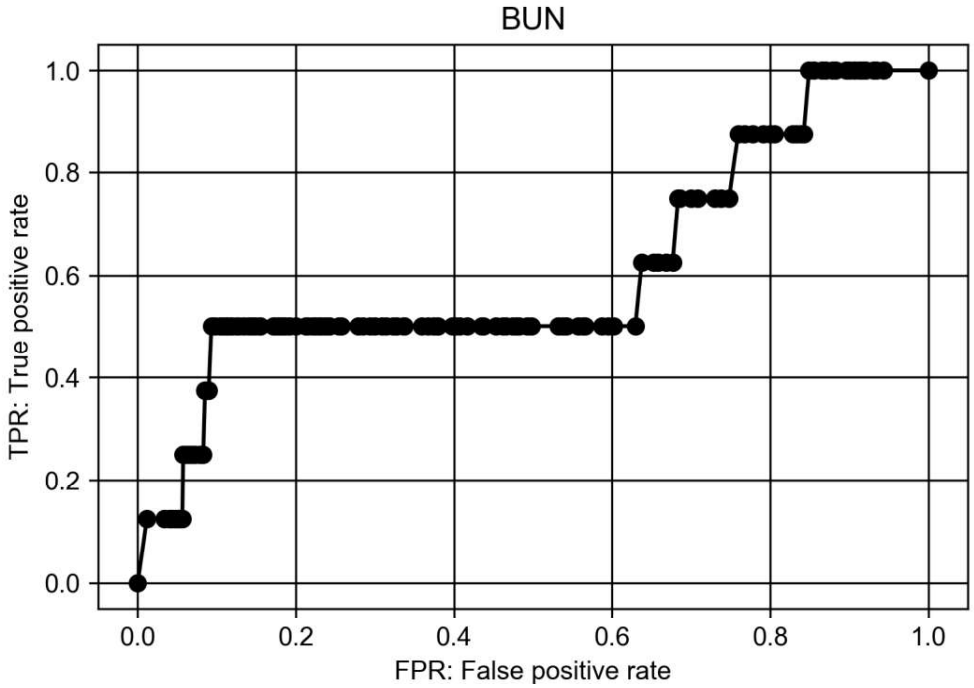

AST(GOT)
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

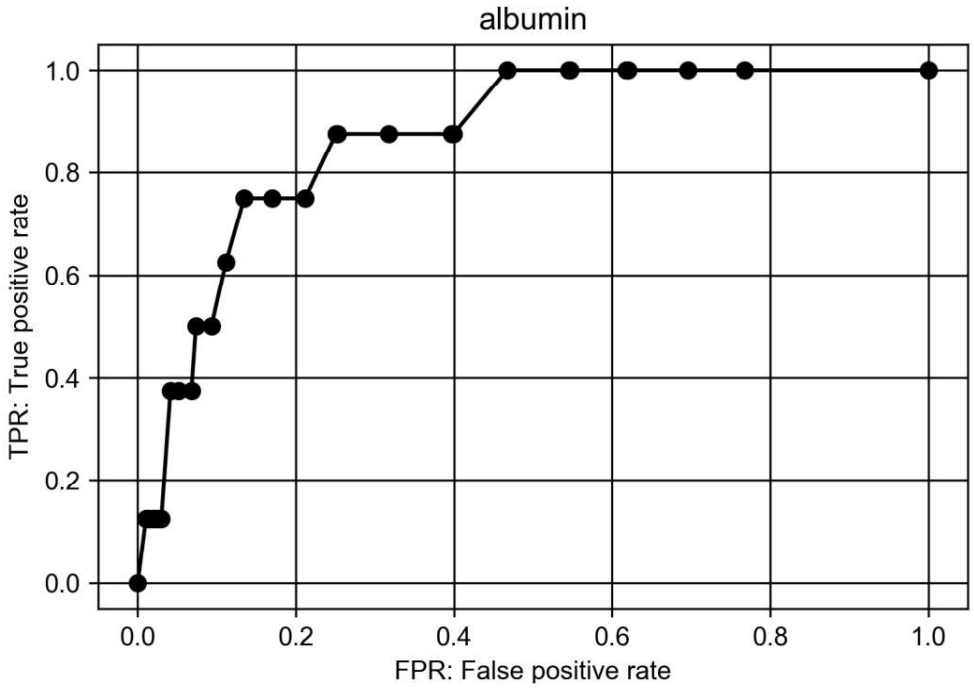

D-dimer
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

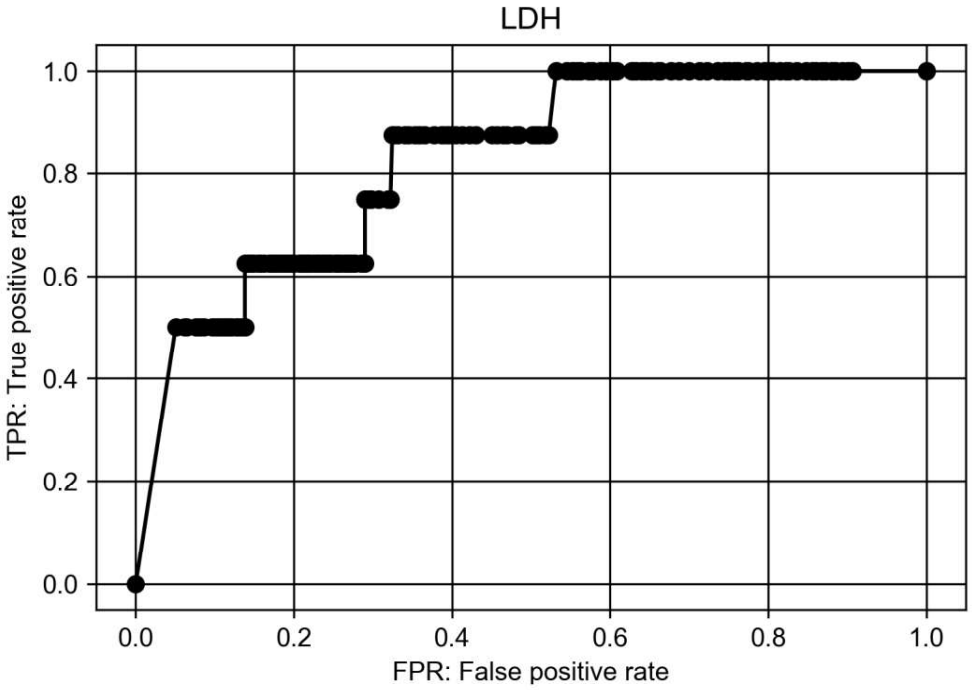

prothrombin time
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

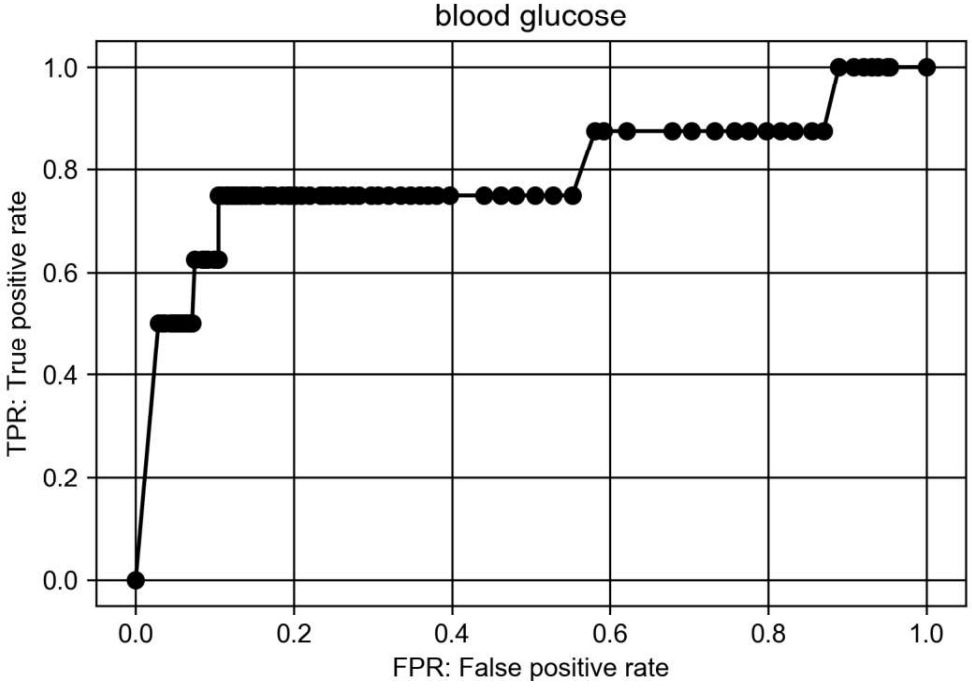

CRP
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

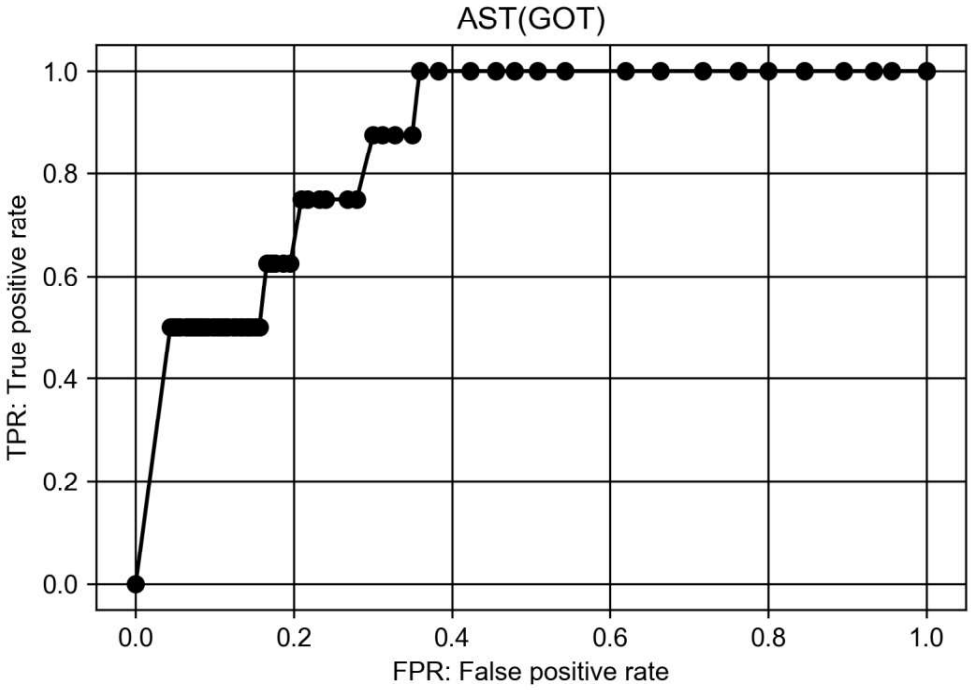
(b) Males aged <50 years

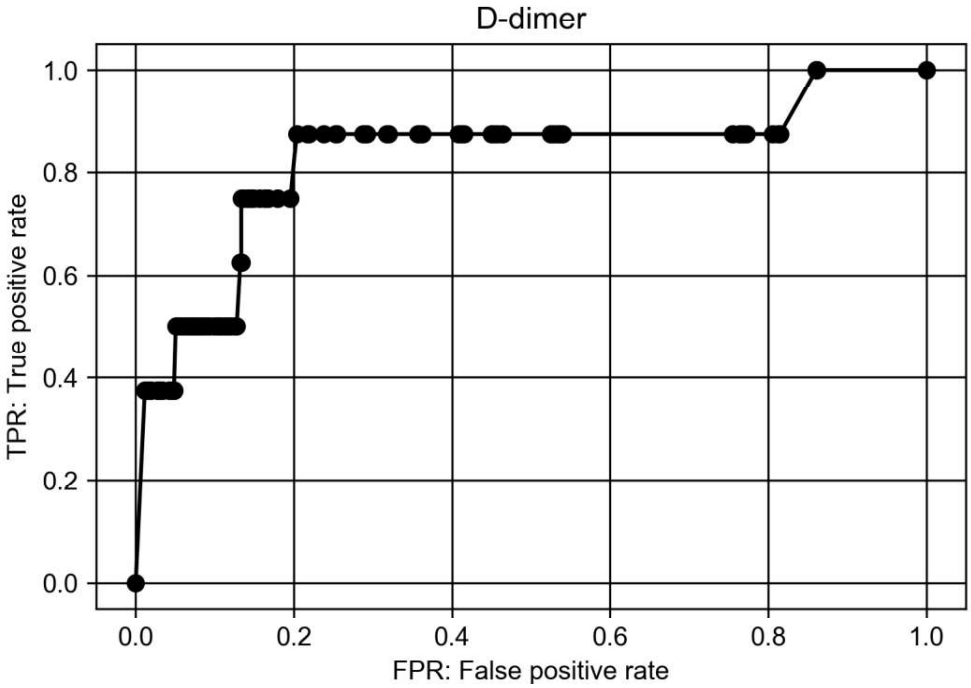


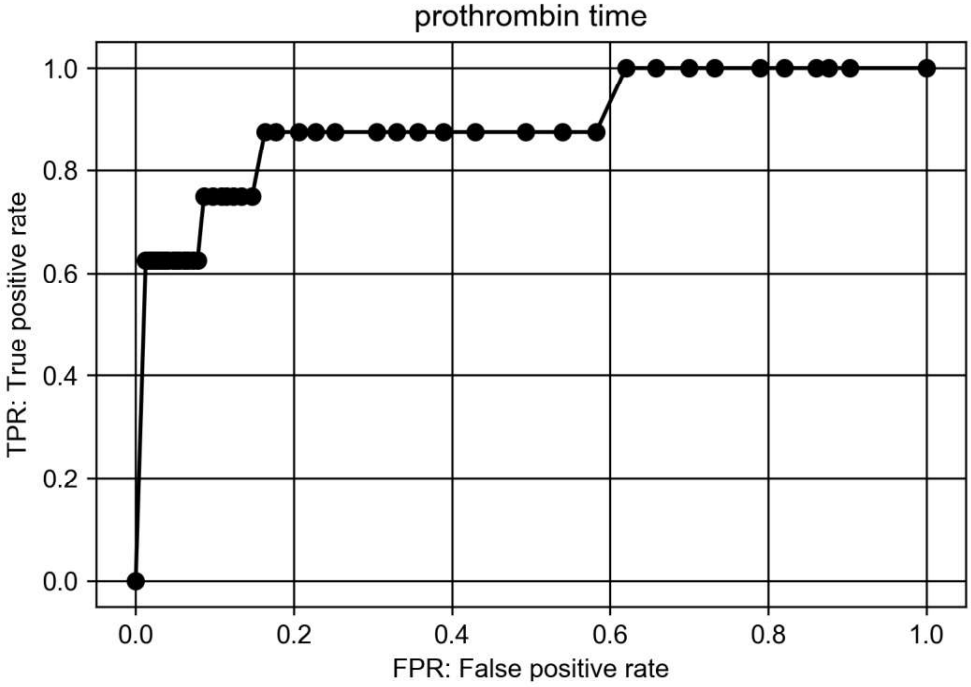


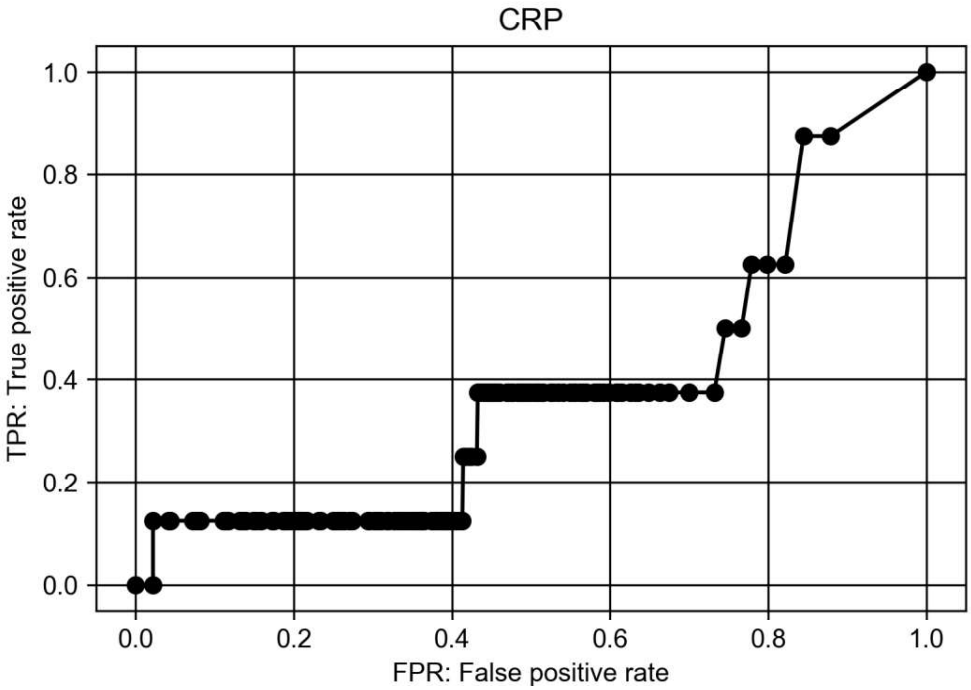

LDH
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


blood glucose
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

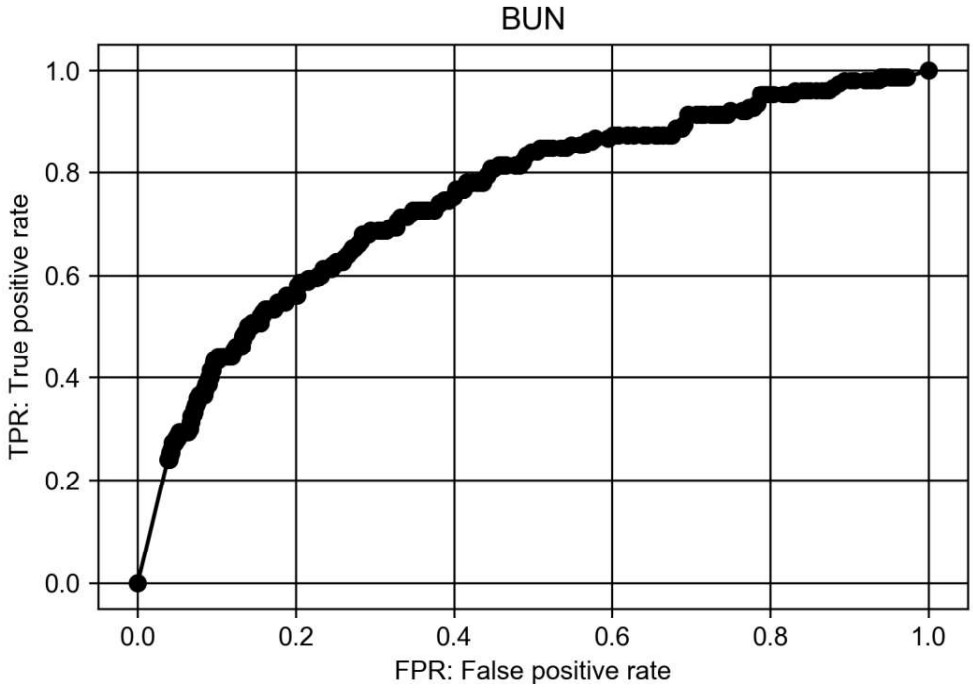

AST(GOT)
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

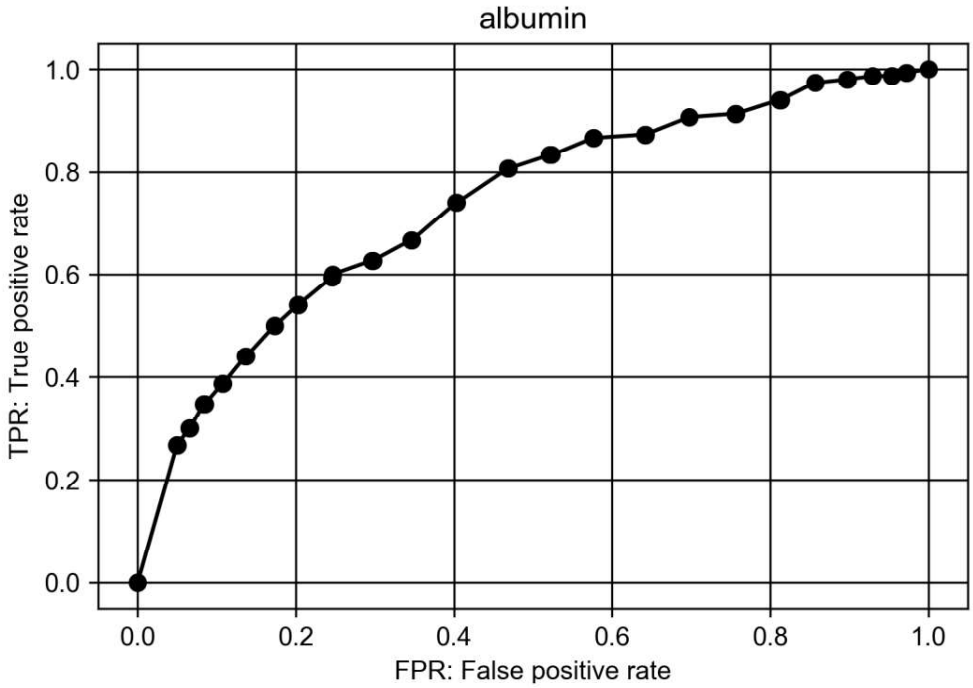

D-dimer
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

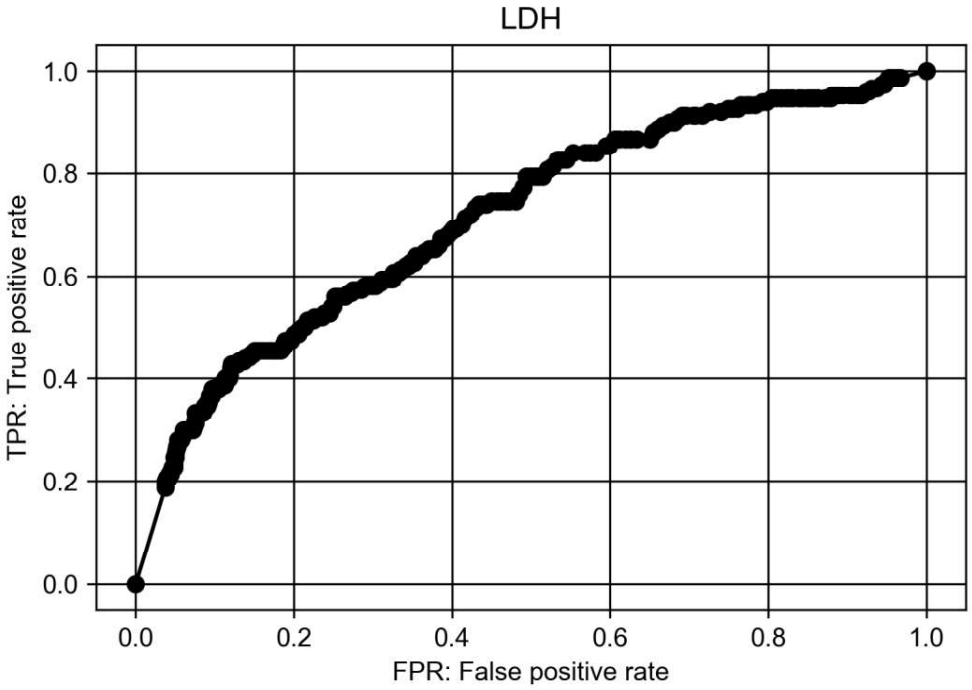

prothrombin time
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

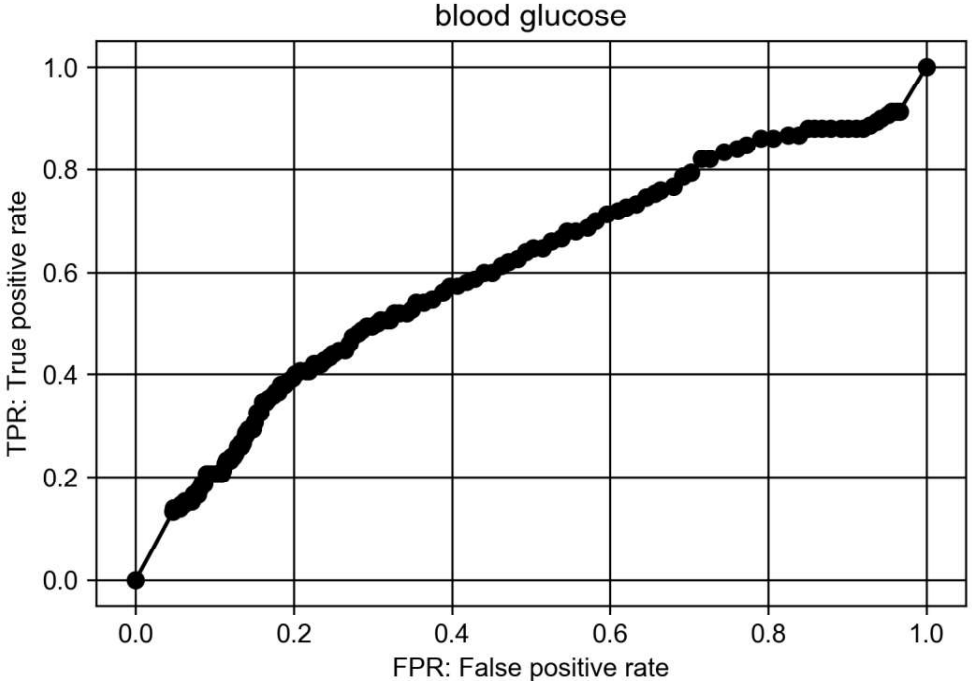

CRP
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

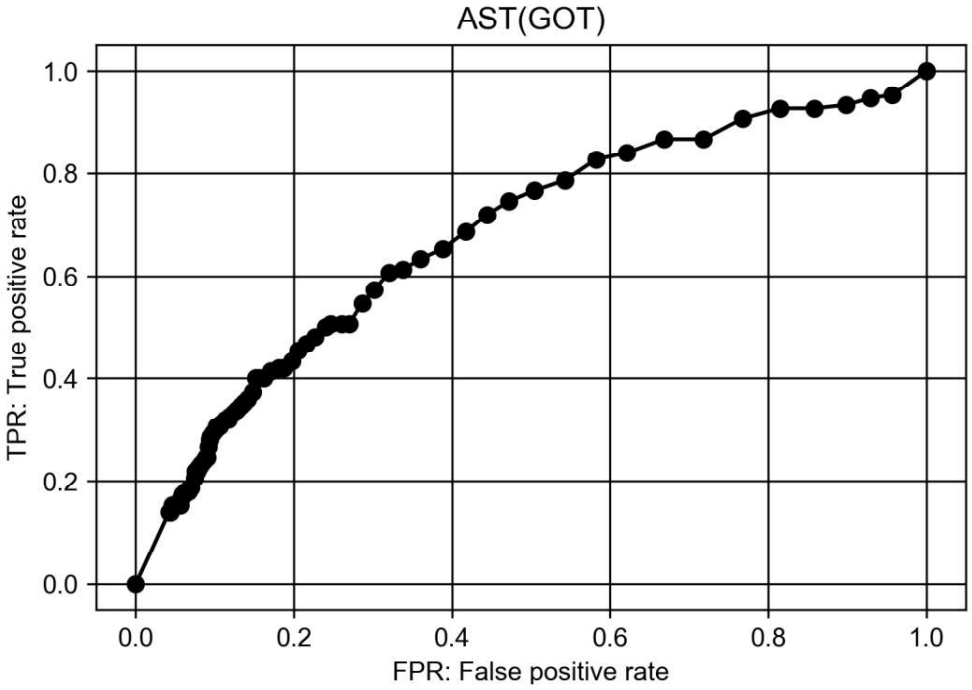
(c) Females aged ≥ 50 years

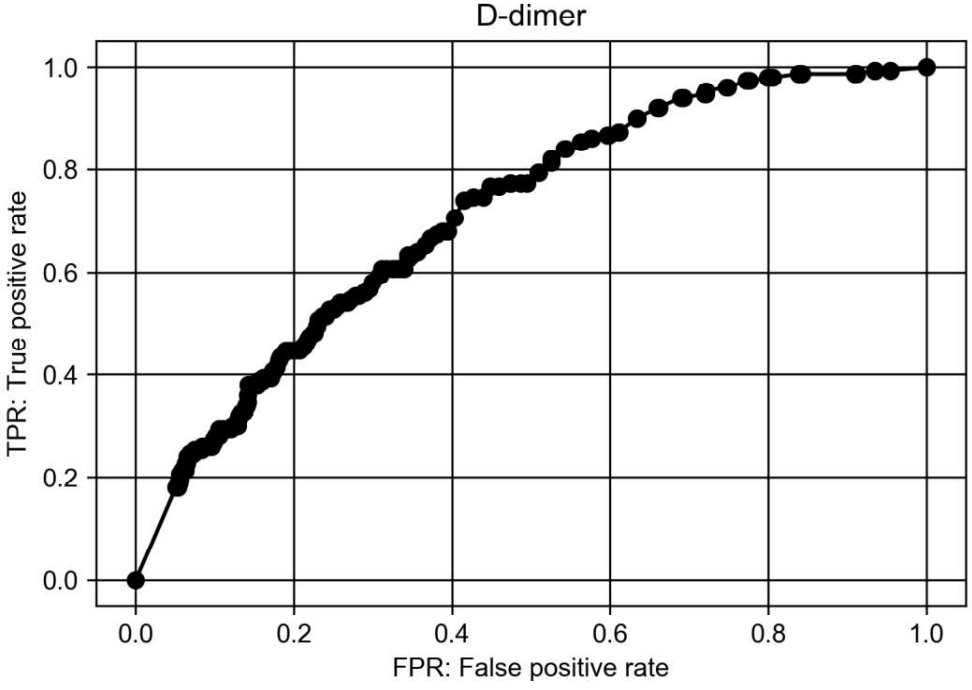


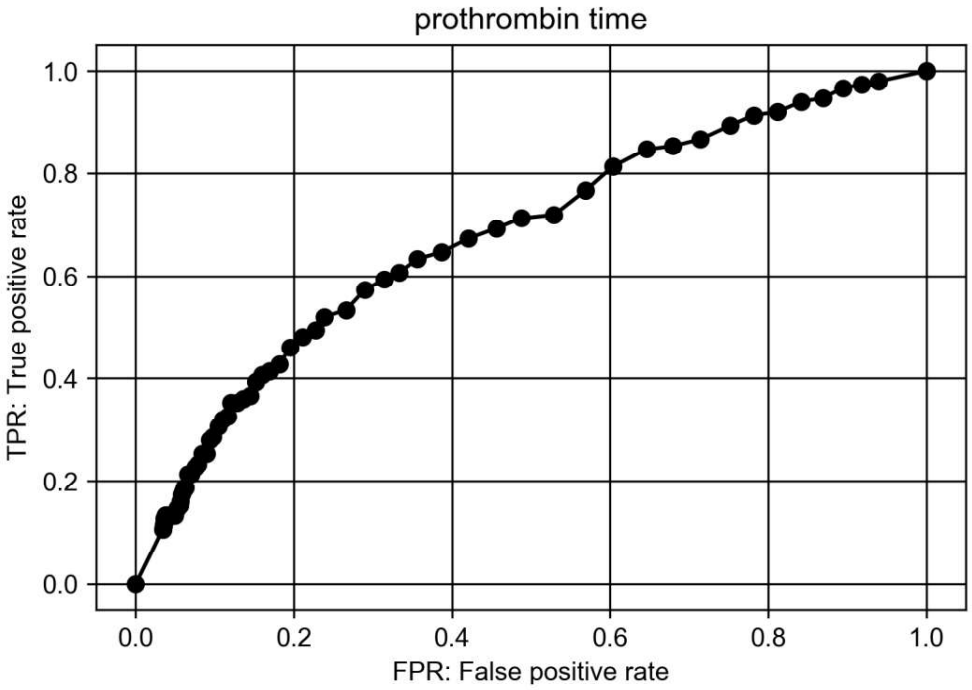


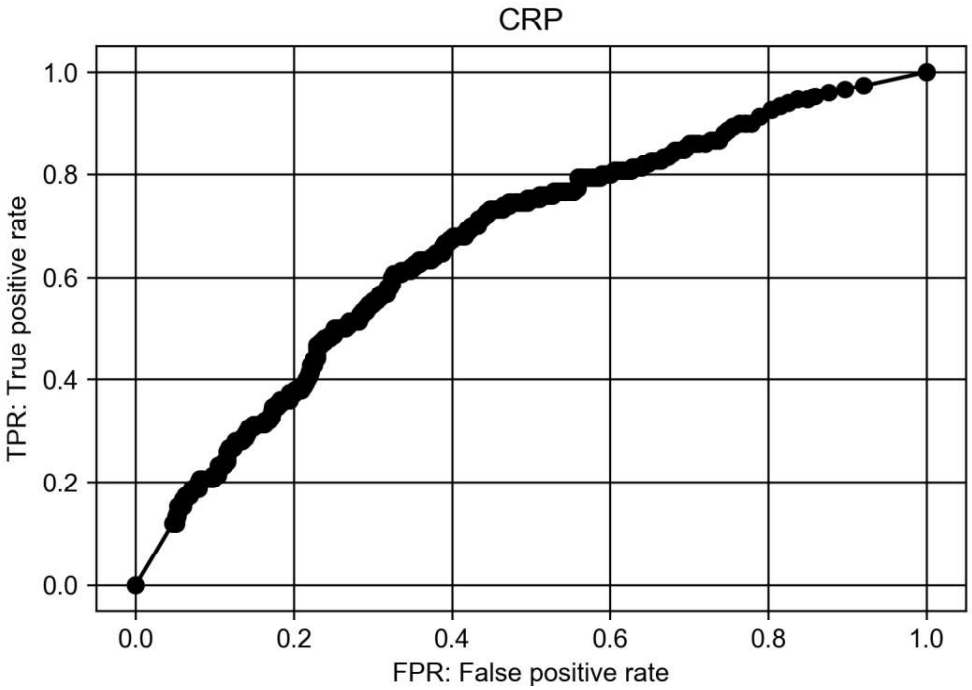

LDH
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


blood glucose
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

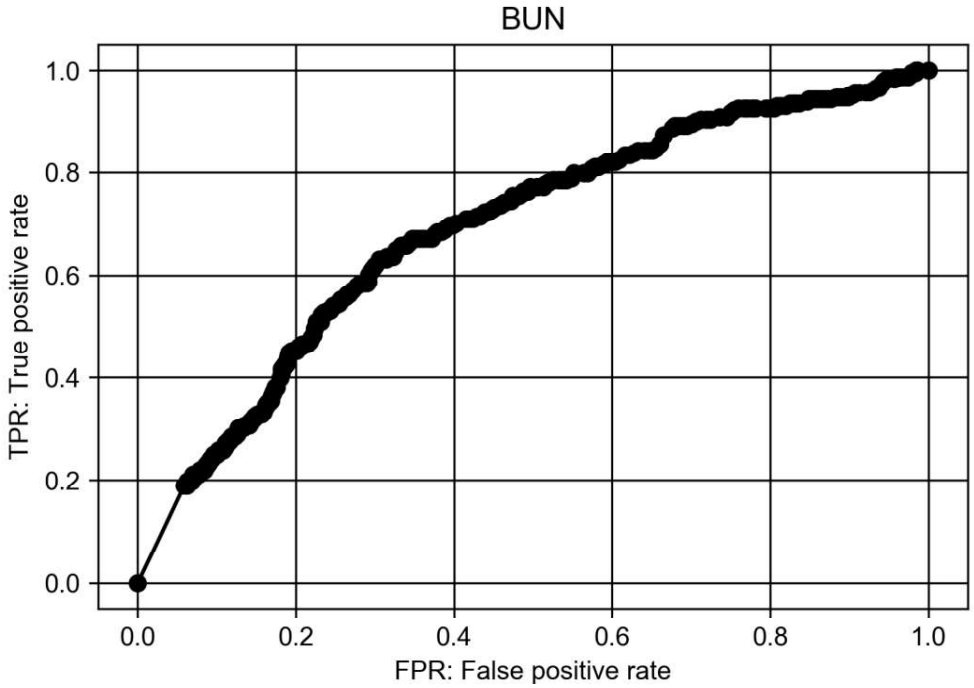

AST(GOT)
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

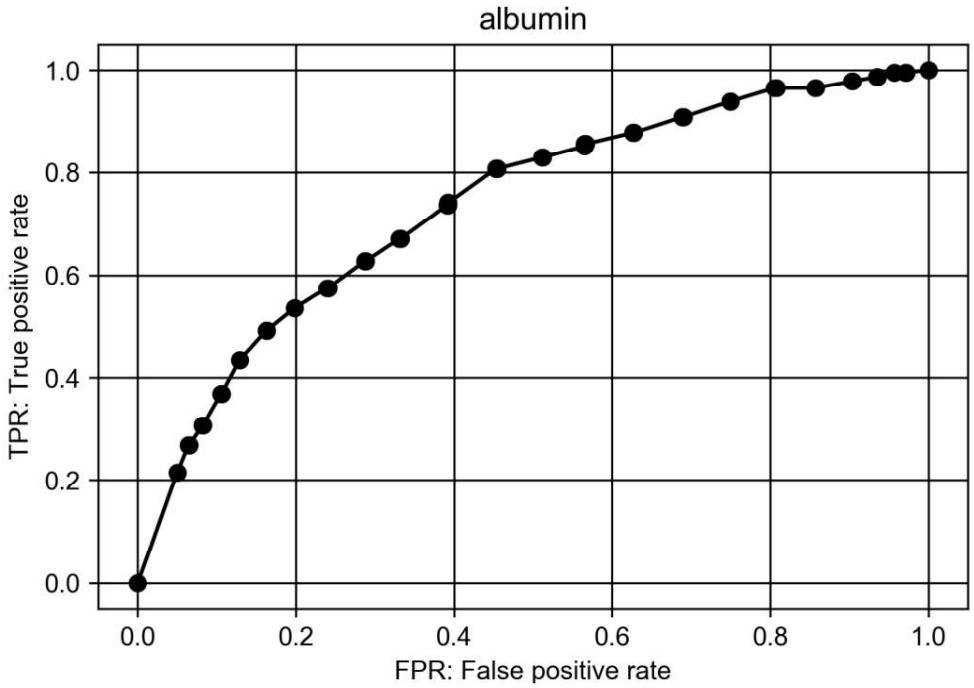

D-dimer
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

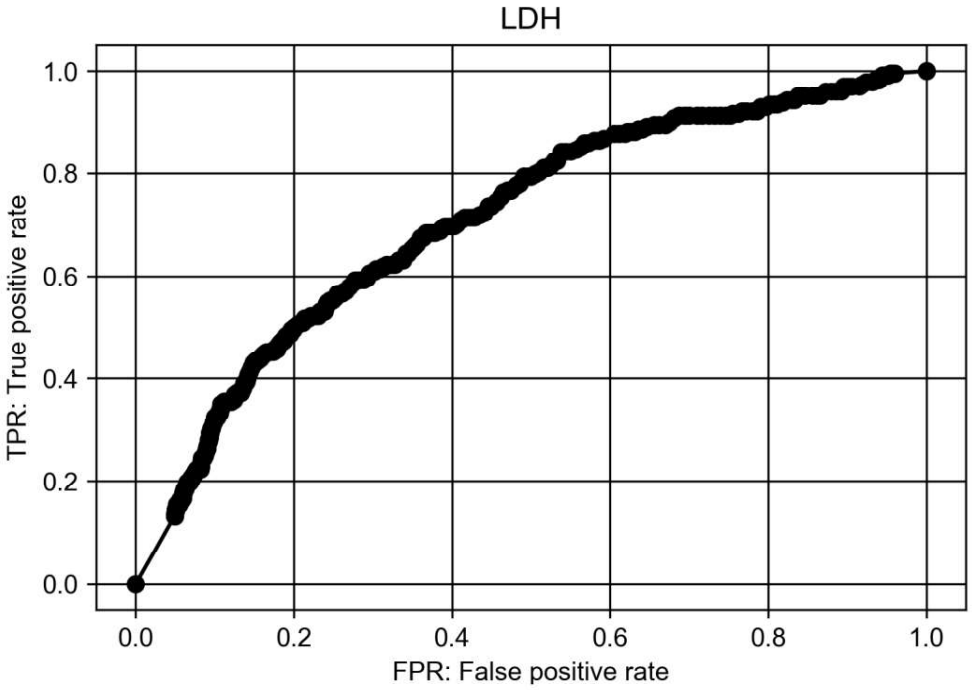

prothrombin time
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

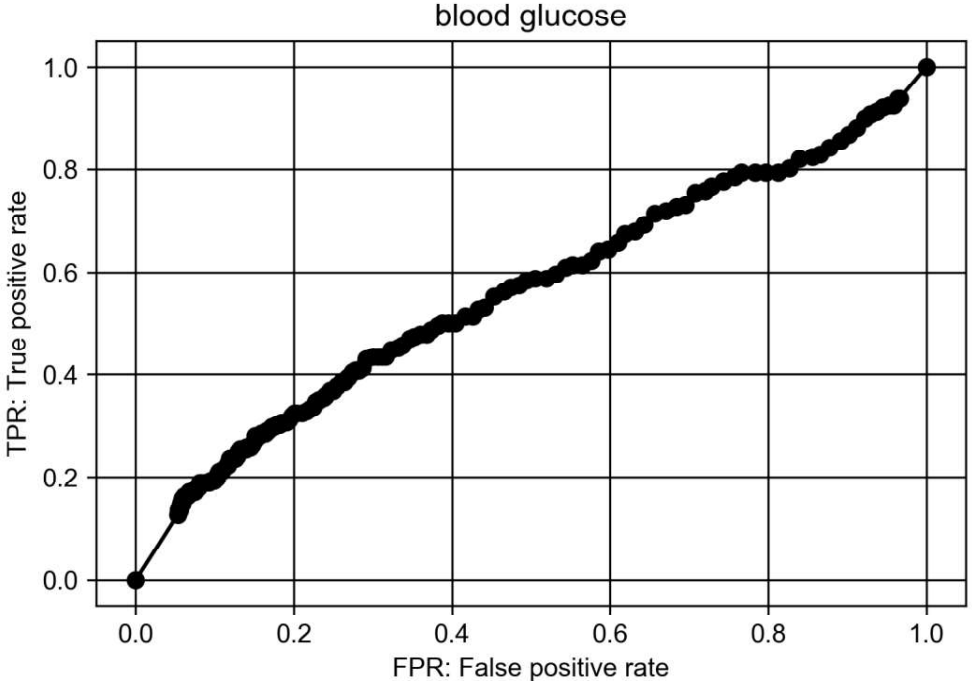

CRP
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

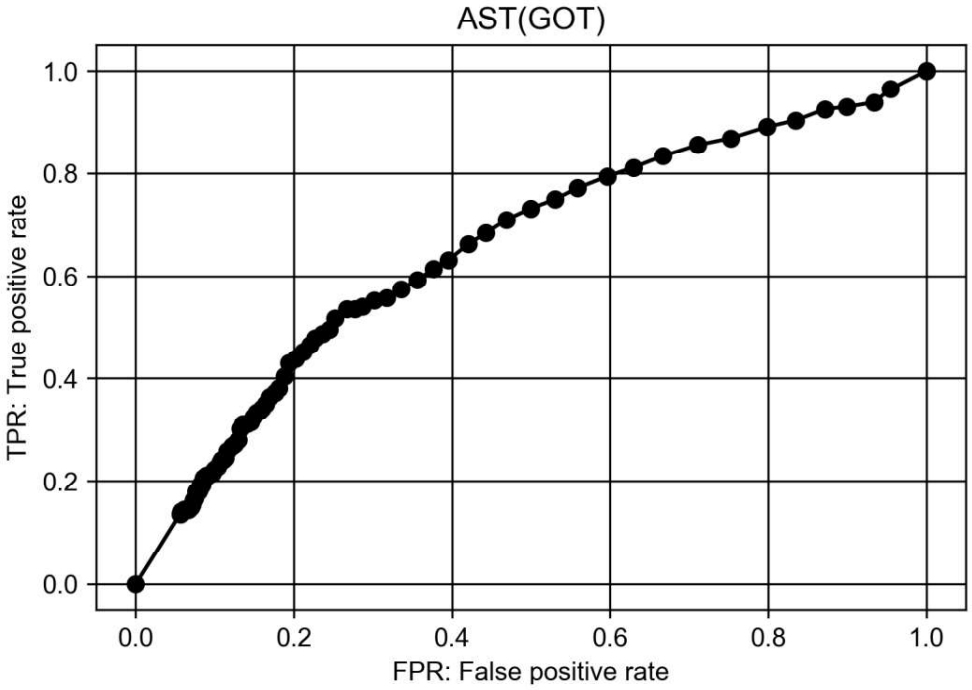
(d) Males aged ≥ 50 years

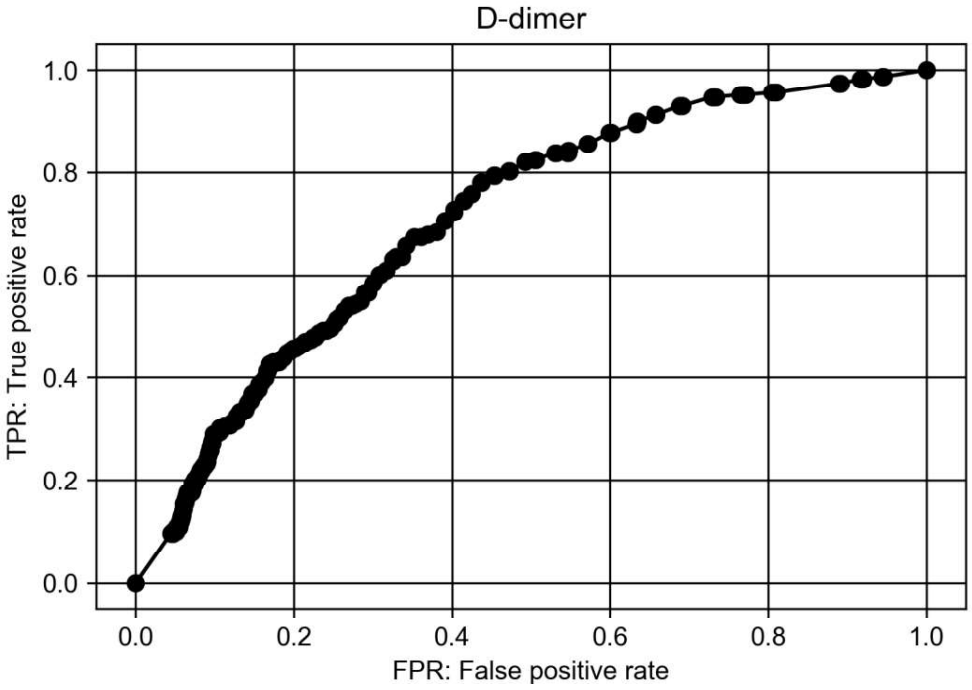


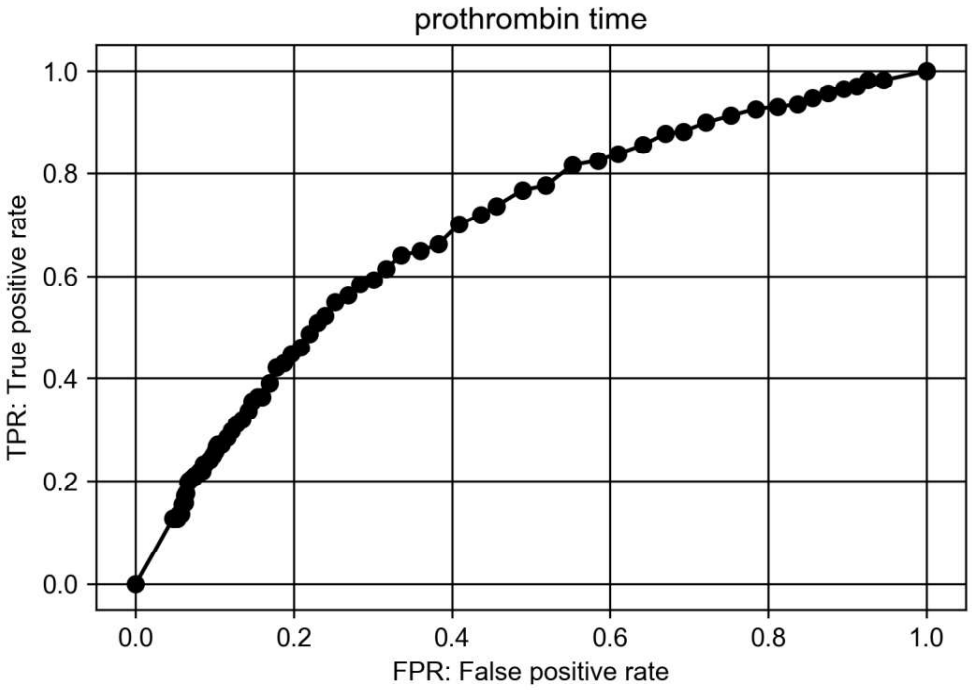


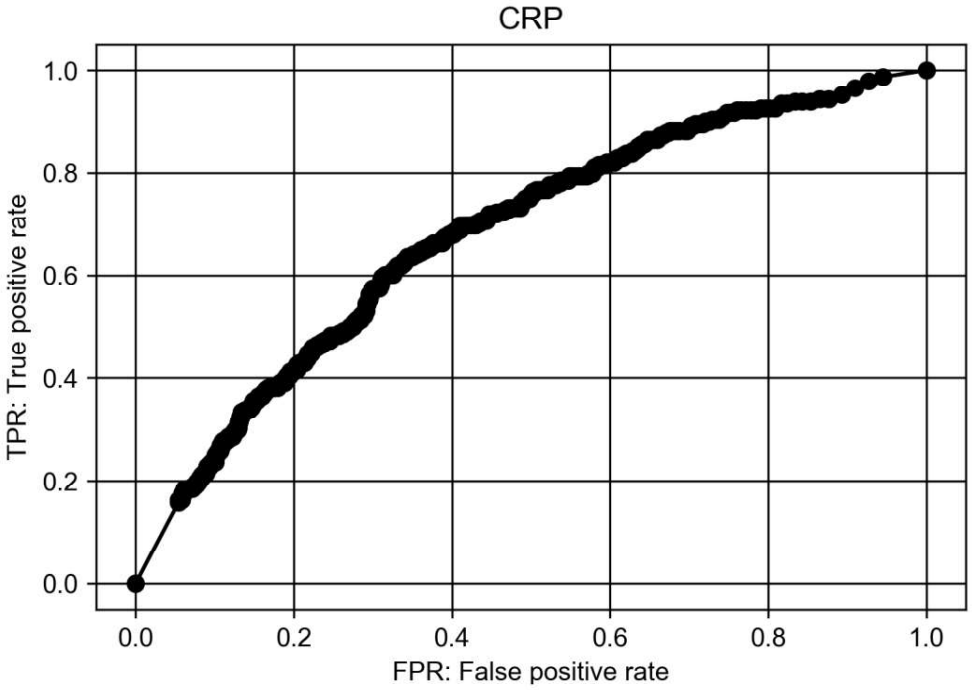

LDH
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


blood glucose
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


AST(GOT)
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


D-dimer
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


prothrombin time
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


CRP
TPR: True positive rate
FPR: False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

CRP: C-reactive protein, AST (GOT): aspartate aminotransferase (glutamate oxaloacetate transaminase), LDH: lactate dehydrogenase, BUN: blood urea nitrogen ROC: receiver operating characteristic curve

Correction

Article title: Prediction models for in-hospital deaths of patients with COVID-19 using electronic healthcare data
Authors: Kenichi Hiraga, Masato Takeuchi, Takeshi Kimura, Satomi Yoshida and Koji Kawakami
Journal: *Current Medical Research and Opinion*
Bibliometrics: Volume 39, Number 11, pages 1463–1471
DOI: https://doi.org/10.1080/03007995.2023.2270420

When this article was first published online, the value in the "Missing" value column of the "Smokers" row of the Table 1 was incorrect. The correct value should be 5,699, not 7,657. This is now been corrected and article has been re-published online.