Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(1):103-130

Original Article

Variable Selection for Recurrent Events Using Heuristic Approaches: Identifying Informative Variables for Rehospitalization in Schizophrenia Patients

Mahya Arayeshgari[1], Leili Tapak[2,3*], Sharareh Parami[1], Behnaz Alafchi[3]

[1]Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran.
[2]Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.
[3]Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.

ARTICLE INFO

Received 26.11.2022
Revised 24.01.2023
Accepted 22.02.2023
Published 15.03.2023

Key words:
Random forest;
Recursive feature elimination;
Deviance residual;
Recurrent event datasets;
Variable selection;
Schizophrenia

ABSTRACT

Introduction: Recurrent event data, as a generalization of survival data, are frequently observed in various areas of medical research, including sequential hospitalizations in patients with schizophrenia. As experiencing multiple relapses during schizophrenia can have many implications, such as self-harm or harm to others, loss of education or employment, or other adverse outcomes, identifying and determining the most critical factors related to relapses in this disorder is essential. This study aimed to utilize heuristic approaches for selecting predictor variables in the field of recurrent events with an application to schizophrenia disorder.
Methods: A two-step algorithm was employed to apply a combination of two variable selection methods, recursive feature elimination (RFE) and genetic algorithm feature selection (GAFS), and four modeling techniques: Gradient boosting (GB), artificial neural network (ANN), random forest (RF), and support vector machine (SVM) to simulated recurrent event datasets.
Results: In most simulation scenarios, the results indicated that the combination of RFE and RF applied to the deviance residual (DR) outperforms the other methods. The RFE-RF-DR selected the following predictor variables: Number of children, age, marital status, and history of substance abuse.
Conclusion: Our findings revealed that the proposed machine learning-based model is a promising technique for selecting predictor variables associated with a recurrent outcome when analyzing multivariate time-to-event data with recurrent events.

Introduction

In recurrent event datasets, events occur more than once for each subject. In other words, the occurrence of an event does not remove the subject from the risk set after that.[1] These data types are frequently used in different areas, such as medical studies.[2] For instance, the majority of schizophrenia patients experience frequent recurrences of symptoms during the illness. There are several severe implications for recurrences of symptoms of this chronic and disabling illness, such as self-harm or harm to others, jeopardizing employment or

*.Corresponding Author: l.tapak@umsha.ac.ir

education status of the patients, and biological threats (for example, returning to the previous level of function may not be possible for some patients).[3] Tumor recurrences after surgeries, successive heart attacks, asthma attacks, and epileptic seizures are other examples of recurrent event datasets.[1, 2]

Typically, there are many candidate predictor variables when we want to model response variables. However, there is often no prior knowledge of which variables should be entered into a statistical model. In addition to making model training expensive and limiting the model's interpretability, non-informative or redundant features can significantly lower the performance of prediction models. Various methods for selecting predictor variables are available such as recursive feature elimination (RFE), and genetic algorithms feature selection (GAFS) which can be used together with machine learning methods. While GAFS searches for the best maximum number of generations, RFE searches for the best subset size. Recognizing and removing the uninformative features are frequently contingent on the learning algorithms used.[4] Therefore, the comparison of different variable selection methods may provide us with valuable information about the ability of different statistical models in terms of selecting important predictor variables and correctly predicting new data.

In the last few decades, machine learning theories have been used extensively to extract proper knowledge from a wide variety of problems with vast amounts of information.[5] Popular methods of machine learning, such as the artificial neural network (ANN) and support vector machine (SVM), can utilize to select informative explanatory variables, but they cannot be applicable to recurrent events data directly. Despite the extensive application and development of machine learning models for analyzing different outcomes, including time to a predefined event (single survival outcome), few attempts have been made to extend machine learning methods in analyzing multivariate survival data. For example, a two-step algorithm has been introduced to deal with this restriction of recurrent event datasets.6 In this algorithm, in the first step, for each subject, different types of residuals (risk indices) are obtained by fitting a Cox model without any covariates (null model). In the second step, residuals are used as outcomes in gradient boosting (GB) and random forest (RF) algorithms.

As the performance of machine learning methods for different outcomes is data-dependent, investigating and comparing different techniques for various data is vitally important. In the present study, we used the same two-step algorithm proposed by Duan and Fu, with the difference that instead of estimating the correct order of variables' importance, we conducted a simulation study to compare four machine learning models (GB, ANN, RF, and SVM) in terms of selecting predictor variables (using RFE and GAFS methods) and their predictive power based on selected variables. In addition, we used the most performed combination of machine learning and variable selection methods applied to risk indices for a real data set related to the rehospitalization of patients with schizophrenia disorder.

Materials and Methods

Risk Indices

We need a single summary statistic (risk indices) for the response variable of each subject to apply machine learning methods such as GB to recurrent events data. Before stating risk indices, we begin with some notations. Consider n independent subjects experiencing several events over a time course (i.e., recurrent events). Thus, for the i th subject, the counting process

$$N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \wedge C_{ij} \leq t)$$

is defined which is a stochastic process that accounts for the number of cumulative events over the [0, t] interval, where T_{ij} and C_i stand for the j th event time and individual's censoring time, respectively. Then, the intensity function of this recurrent event counting process for subject i can be defined as

$$\begin{aligned}\lambda_i(t) &= \lim_{ds\to 0} \frac{P\{N_i(t, t+ds) = 1 \mid \Im_{t^-}\}}{ds} \\ &= Y_i(t)\lambda_0(t)\exp\{\beta' Z_i(t)\}\end{aligned} \qquad (1)$$

Where N_i (t, t + ds) denotes the number of events that occur over the $\Im_{t^-}$ [t, t + ds] interval for the i th subject, and (σ-algebra) represents all that occurred up to time t. Y_i (t) indicates a predictable process that gets a value of 1 if the i th subject is under observation at time t and 0 elsewhere. The baseline intensity function and regression coefficients vector are represented by $\lambda_0(.)$ and β, respectively, and also for the i th subject, a covariate process with p dimensions is shown with Z_i (t).

Three risk indices will be introduced in the following sections. A higher value of the risk indices represents a higher chance of experiencing more events during a certain period of time.

Martingale Residual (MR)

The martingale residual for the ith subject, over [0, t] time interval, is defined as the difference between the counting process and the integrated intensity function.[7] Now, in recurrent event datasets,

$$\hat{M}_i(\tau_i) = N_i(\tau_i) - e^{\hat{\beta}' Z_i}\,\hat{\Lambda}_0(\tau_i)$$

indicates the martingale residual for the i th subject. Here, is the last observation time for subject, τ_i is the maximum partial likelihood estimator for β and $\hat{\Lambda}_0(\tau_i)$ is the Breslow estimator[8, 9] for the baseline cumulative mean function $\hat{\Lambda}_0(\tau_i)$. Henceforth, the covariate Z_i is assumed not to be time-dependent. The risk index will be derived from the martingale residual by fitting a null model on the intensity function (1) stated in section 2.1:

$$\hat{M}_i = N_i(\tau_i) - \hat{\Lambda}_0(\tau_i) \qquad (2)$$

$\hat{M}_i$ was replaced with $\hat{M}_i(\tau_i)$ to show the risk index.

Adjust Martingale Residual (AMR)

We can standardize equation (2) to adjust the martingale residual:

$$\hat{M}_i^a = \frac{N_i(\tau_i) - \hat{\Lambda}_0(\tau_i)}{\sqrt{\hat{\Lambda}_0(\tau_i)}} \qquad (3)$$

The asymptotic distribution of this modified residual is independent of time τ_i.[8]

Deviance Residual (DR)

Unlike residuals in linear models, martingale residuals are not symmetrically distributed around zero. To obtain symmetry, deviance residuals which are simply a normalizing transformation of the martingale residuals, are sometimes used.[10] For the *i*th subject, the deviance residual (D_i^1) is the square root of the deviance (D_i) augmented by the sign of the martingale residual ($\hat{M}_i$):

$$D_i^1 = sign(\hat{M}_i) \times \sqrt{D_i} \qquad (4)$$

where

$$sign(\hat{M}_i) = \begin{cases} +1 & \hat{M}_i > 0 \\ 0 & \hat{M}_i = 0 \\ -1 & \hat{M}_i < 0 \end{cases}$$

and

$$D_i = -2\left[\hat{M}_i + N_i(\tau_i)\log\left\{\frac{N_i(\tau_i) - \hat{M}_i}{N_i(\tau_i)}\right\}\right]$$

Equations 2, 3, and 4 are called risk indices.

Variable Selection Methods

Recursive Feature Elimination (RFE)

RFE is a backward selection algorithm.[11] In this method, in order to rank the variables from most important to least, a measure of variable importance is calculated after the full model has been fitted. The model is rebuilt after repeatedly eliminating the least important predictors at each search step. The objective function is calculated for a new model after it has been built. After repeating this process for some predefined sequences, the final model is created using the subset size that corresponds to the highest value of the objective function.[12]

Genetic Algorithms Feature Selection (GAFS)

It has been demonstrated that a genetic algorithm (GA), which relies on population biology's evolutionary principles is successful in finding optimal solutions for complicated, multivariate functions. Feature selection is also a challenging optimization issue in which we look for the set of features that will produce the best possible prediction of the response. In the context of feature selection using GA, the length of the chromosome, a binary array, is equal to the number of predictors in the data set. The existence or lack of each predictor in the data is represented by a binary record on a chromosome, or gene. The model uses the variables suggested by the binary vector to calculate the chromosome's fitness. Therefore, among the 2^n various combos of predictor sets, GAs are charged with finding optimal solutions.[12]

Two-step Algorithm

Step #1

a) Fit a model without any covariates on the intensity function of recurrent event datasets:

$$\lambda(t \mid \Im_{t^-}) = \lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\beta' Z_i) = Y_i(t)\lambda_0(t) \qquad (5)$$

b) Estimate the baseline intensity function by the Breslow estimator.

c) Obtain risk indices including MR, AMR, and DR from the fitted model for each subject.

Step #2

Apply RFE-GB, RFE-ANN, RFE-RF, RFE-SVM, GAFS-GB, GAFS-ANN, GAFS-RF, and GAFS-SVM to risk indices generated from step 1 in order to select informative predictor variables and assess models' predictive power.

Evaluation Metrics

The suggested predictive models were quantitatively assessed using some performance measures, including sensitivity, specificity, total accuracy, and root-mean-squared error (RMSE) as follows:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (7)$$

$$\text{Total Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (9)$$

Where a false positive (FP) indicates non-informative variables that were incorrectly identified as informative, a true positive (TP) indicates informative variables that were correctly diagnosed as informative, a true negative (TN) indicates non-informative variables correctly identified as non-informative, and a false negative (FN) indicates informative variables incorrectly identified as non-informative. In addition, n denotes the total number of data points. The output values, denoted as $\hat{y}_i$, represent the predicted values generated by the model, while the real values, denoted as $\hat{y}_i$, are the true values of the response variable.

Software

Statistical analysis was performed using R-4.2.1 software by several packages such as simrec, caret, kernlab, gbm, plyr, randomForest, and nnet.

Simulation Study

We set up 20 different scenarios to carry out the simulation study to evaluate and compare the performance of different combinations of machine learning and variable selection methods applied to different risk indices. In the simulation study, we generated 50 data sets with sample sizes of 300 and nine covariates. Three out of nine covariates were considered effective (informative). All covariates were generated from the standard normal distribution. We used four different sets of coefficient values ((2,1.2,0.6,0,0,0,0,0,0), (1.9,1.5,1.3,0,0,0,0,0,0), (1.9,1.5,0.1,0,0,0,0,0,0), and (1,0.8,0.1,0,0,0,0,0,0)). The event time and individual censoring time were generated from an exponential distribution and a uniform distribution, respectively. The censoring rate was set to 0, 0.2, 0.5, and 0.8. We assumed patients were followed up for a maximum of three years from the onset of the study. Each time the simulation data was generated, we used the 70/30 proportion to split the data into training and test set. Then, the 10-fold cross-validation strategy was performed on the training set to optimize the tuning parameters of machine learning methods and compute sensitivity, specificity, and total accuracy. Furthermore, in order to compare the prediction of the models, we used the test set to calculate the RMSE.

Simulation Results

Evaluation metrics based on different machine learning and variable selection methods for each risk index are provided in Tables 1 to 6 (for a 1-year and 3-year follow-up). According to Tables 1 and 2, the method of RFE-RF-MR (recursive feature elimination and random forest algorithm applied to martingale residual) gave the most considerable sensitivity, specificity, total accuracy, and smallest RMSE in 5, 12, 12, and 6 scenarios, respectively. In addition, RFE-GB-MR provided smaller RMSEs in 12 scenarios; however, the values of RMSE in these cases for RFE-RF-MR were not so different from those for RFE-GB-MR. Similarly, according to Tables 3 to 6, in most scenarios, RFE-RF-AMR and RFE-RF-DR performed well in the evaluation metrics. Finally, RFE-RF-DR, with the most considerable value for sensitivity, specificity, and total accuracy in most scenarios, was recognized as the best method compared to RFE-RF-MR and RFE-RF-AMR. On the other hand, most of the scenarios showed that the GAFS method performs well only in terms of sensitivity (Tables 1 to 6).

Table 1. Evaluation metrics for different combinations of machine learning and variable selection methods applied to martingale residual with a 1-year follow-up

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-MR	1	0	2,1.2,0.6	4.40±1.95	1.00±0.00	0.76±0.32	0.84±0.21	0.75±0.10
	2	0	1.9,1.5,1.3	4.15±1.87	1.00±0.00	0.80±0.31	0.87±0.20	0.86±0.08
	3	0	1.9,1.5,0.1	2.85±1.75	0.71±0.12	0.88±0.24	0.82±0.14	0.77±0.07
	4	0	1,0.8,0.1	3.05±1.66	0.70±0.10	0.84±0.26	0.79±0.17	0.64±0.04
	5	0.2	2,1.2,0.6	3.70±1.45	0.96±0.10	0.86±0.23	0.90±0.15	0.77±0.07
	6	0.2	1.9,1.5,1.3	3.95±1.31	1.00±0.00	0.84±0.21	0.89±0.14	0.83±0.07
	7	0.2	1.9,1.5,0.1	2.85±1.49	0.73±0.13	0.89±0.21	0.83±0.12	0.76±0.07
	8	0.2	1,0.8,0.1	3.45±2.13	0.76±0.15	0.80±0.29	0.79±0.16	0.65±0.05
	9	0.8	2,1.2,0.6	3.95±1.76	0.96±0.10	0.82±0.27	0.87±0.18	0.72±0.08
	10	0.8	1.9,1.5,1.3	4.25±1.83	1.00±0.00	0.79±0.30	0.86±0.20	0.77±0.09
	11	0.8	1.9,1.5,0.1	2.60±1.18	0.71±0.12	0.92±0.16	0.85±0.10	0.70±0.09
	12	0.8	1,0.8,0.1	4.20±2.60	0.83±0.17	0.71±0.37	0.75±0.21	0.58±0.05
RFE-ANN-MR	1	0	2,1.2,0.6	4.75±1.94	0.98±0.07	0.70±0.31	0.79±0.20	1.62±0.16
	2	0	1.9,1.5,1.3	3.95±1.60	1.00±0.00	0.84±0.26	0.89±0.17	1.79±0.16
	3	0	1.9,1.5,0.1	3.00±1.41	0.70±0.10	0.85±0.22	0.80±0.15	1.63±0.16
	4	0	1,0.8,0.1	4.00±2.36	0.76±0.15	0.71±0.33	0.73±0.18	1.07±0.06
	5	0.2	2,1.2,0.6	4.35±1.63	0.96±0.10	0.75±0.26	0.82±0.17	1.52±0.13
	6	0.2	1.9,1.5,1.3	4.15±1.30	1.00±0.00	0.80±0.21	0.87±0.14	1.63±0.15
	7	0.2	1.9,1.5,0.1	3.30±1.86	0.71±0.12	0.80±0.28	0.77±0.18	1.48±0.10
	8	0.2	1,0.8,0.1	4.10±1.74	0.80±0.16	0.71±0.25	0.74±0.16	1.02±0.06
	9	0.8	2,1.2,0.6	4.85±1.89	0.96±0.10	0.67±0.29	0.77±0.18	1.16±0.14
	10	0.8	1.9,1.5,1.3	5.00±1.74	1.00±0.00	0.66±0.29	0.77±0.19	1.26±0.14
	11	0.8	1.9,1.5,0.1	3.95±1.84	0.80±0.16	0.74±0.26	0.76±0.15	1.16±0.13
	12	0.8	1,0.8,0.1	4.60±2.28	0.85±0.17	0.65±0.31	0.72±0.17	0.79±0.06

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; RFE-GB-MR, Recursive feature elimination and gradient boosting algorithm applied to martingale residual; RFE-ANN-MR, Recursive feature elimination and artificial neural network algorithm applied to martingale residual;

Table 1 (continued)

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-RF-MR	1	0	2,1.2,0.6	3.15±0.74	0.98±0.07	0.96±0.11	0.97±0.07	0.74±0.08
	2	0	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.87±0.08
	3	0	1.9,1.5,0.1	2.05±0.22	0.68±0.07	1.00±0.00	0.89±0.02	0.77±0.07
	4	0	1,0.8,0.1	3.85±2.27	0.76±0.15	0.74±0.31	0.75±0.17	0.66±0.05
	5	0.2	2,1.2,0.6	3.30±1.55	0.95±0.12	0.92±0.24	0.93±0.16	0.78±0.07
	6	0.2	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.82±0.07
	7	0.2	1.9,1.5,0.1	2.05±0.22	0.66±0.00	0.99±0.03	0.88±0.02	0.77±0.08
	8	0.2	1,0.8,0.1	4.90±2.88	0.80±0.16	0.58±0.41	0.65±0.24	0.66±0.05
	9	0.8	2,1.2,0.6	4.35±2.18	0.96±0.10	0.75±0.34	0.82±0.22	0.72±0.09
	10	0.8	1.9,1.5,1.3	3.30±1.34	1.00±0.00	0.95±0.22	0.96±0.14	0.76±0.08
	11	0.8	1.9,1.5,0.1	3.40±1.78	0.76±0.15	0.81±0.24	0.80±0.13	0.71±0.08
	12	0.8	1,0.8,0.1	5.25±2.44	0.83±0.17	0.54±0.34	0.63±0.19	0.59±0.05
RFE-SVM-MR	1	0	2,1.2,0.6	4.60±1.81	0.98±0.07	0.72±0.29	0.81±0.19	0.73±0.10
	2	0	1.9,1.5,1.3	3.30±0.57	1.00±0.00	0.95±0.09	0.96±0.06	0.81±0.09
	3	0	1.9,1.5,0.1	2.45±0.68	0.70±0.10	0.94±0.11	0.86±0.08	0.77±0.07
	4	0	1,0.8,0.1	4.75±2.84	0.81±0.17	0.61±0.40	0.68±0.23	0.66±0.05
	5	0.2	2,1.2,0.6	4.00±1.68	0.96±0.10	0.81±0.26	0.86±0.17	0.78±0.08
	6	0.2	1.9,1.5,1.3	4.05±1.73	1.00±0.00	0.82±0.28	0.88±0.19	0.78±0.07
	7	0.2	1.9,1.5,0.1	3.05±1.35	0.70±0.10	0.84±0.19	0.79±0.12	0.78±0.08
	8	0.2	1,0.8,0.1	5.90±2.88	0.88±0.16	0.45±0.42	0.60±0.25	0.67±0.05
	9	0.8	2,1.2,0.6	3.95±1.53	0.91±0.14	0.80±0.23	0.83±0.15	0.74±0.09
	10	0.8	1.9,1.5,1.3	3.70±1.26	1.00±0.00	0.88±0.21	0.92±0.14	0.75±0.08
	11	0.8	1.9,1.5,0.1	2.95±1.05	0.76±0.15	0.89±0.12	0.85±0.07	0.73±0.10
	12	0.8	1,0.8,0.1	4.40±2.01	0.83±0.17	0.68±0.28	0.73±0.16	0.59±0.06
GAFS-GB-MR	1	0	2,1.2,0.6	6.65±2.15	0.96±0.10	0.37±0.32	0.57±0.20	0.76±0.09
	2	0	1.9,1.5,1.3	7.05±1.39	1.00±0.00	0.32±0.23	0.55±0.15	0.87±0.08
	3	0	1.9,1.5,0.1	5.05±2.03	0.83±0.17	0.57±0.29	0.66±0.17	0.77±0.07
	4	0	1,0.8,0.1	5.65±1.08	0.85±0.17	0.48±0.16	0.60±0.12	0.64±0.04
	5	0.2	2,1.2,0.6	6.80±1.79	0.96±0.10	0.35±0.26	0.55±0.15	0.78±0.06
	6	0.2	1.9,1.5,1.3	7.05±1.84	1.00±0.00	0.32±0.30	0.55±0.20	0.84±0.06
	7	0.2	1.9,1.5,0.1	5.95±2.21	0.86±0.16	0.44±0.31	0.58±0.17	0.78±0.07
	8	0.2	1,0.8,0.1	5.85±1.78	0.86±0.16	0.45±0.27	0.59±0.18	0.65±0.05
	9	0.8	2,1.2,0.6	6.30±1.75	0.98±0.07	0.44±0.27	0.62±0.17	0.73±0.09
	10	0.8	1.9,1.5,1.3	6.10±1.58	1.00±0.00	0.48±0.26	0.65±0.17	0.77±0.08
	11	0.8	1.9,1.5,0.1	6.30±1.97	0.93±0.13	0.41±0.29	0.58±0.18	0.71±0.08
	12	0.8	1,0.8,0.1	6.25±1.48	0.93±0.13	0.42±0.23	0.59±0.16	0.58±0.06

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; RFE-RF-MR, Recursive feature elimination and random forest algorithm applied to martingale residual; RFE-SVM-MR, Recursive feature elimination and support vector machine algorithm applied to martingale residual; GAFS-GB-MR, Genetic algorithm feature selection and gradient boosting algorithm applied to martingale residual;

Table 1 (continued)

GAFS-ANN-MR	1	0	2,1.2,0.6	6.00±1.55	0.98±0.07	0.49±0.24	0.65±0.15	1.62±0.16
	2	0	1.9,1.5,1.3	6.35±1.46	1.00±0.00	0.44±0.24	0.62±0.16	1.79±0.16
	3	0	1.9,1.5,0.1	5.80±1.70	0.80±0.16	0.43±0.25	0.55±0.16	1.63±0.16
	4	0	1,0.8,0.1	6.05±2.28	0.75±0.26	0.36±0.30	0.49±0.18	1.07±0.06
	5	0.2	2,1.2,0.6	6.60±1.66	1.00±0.00	0.40±0.27	0.60±0.18	1.52±0.13
	6	0.2	1.9,1.5,1.3	6.75±1.61	1.00±0.00	0.37±0.26	0.58±0.17	1.63±0.15
	7	0.2	1.9,1.5,0.1	5.90±1.33	0.88±0.16	0.45±0.22	0.60±0.17	1.48±0.10
	8	0.2	1,0.8,0.1	6.25±1.71	0.83±0.22	0.37±0.23	0.52±0.15	1.02±0.06
	9	0.8	2,1.2,0.6	6.55±1.60	0.96±0.10	0.39±0.24	0.58±0.16	1.16±0.14
	10	0.8	1.9,1.5,1.3	7.05±1.35	1.00±0.00	0.32±0.22	0.55±0.15	1.26±0.14
	11	0.8	1.9,1.5,0.1	6.00±1.33	0.81±0.17	0.40±0.23	0.54±0.17	1.16±0.13
	12	0.8	1,0.8,0.1	5.70±1.89	0.78±0.22	0.44±0.26	0.55±0.16	0.80±0.06
GAFS-RF-MR	1	0	2,1.2,0.6	6.15±1.84	0.95±0.12	0.45±0.27	0.61±0.17	0.78±0.09
	2	0	1.9,1.5,1.3	5.85±1.18	1.00±0.00	0.52±0.19	0.68±0.13	0.91±0.09
	3	0	1.9,1.5,0.1	5.05±1.73	0.78±0.16	0.55±0.27	0.62±0.18	0.80±0.07
	4	0	1,0.8,0.1	6.30±1.55	0.85±0.17	0.37±0.21	0.53±0.13	0.65±0.03
	5	0.2	2,1.2,0.6	5.90±1.65	0.96±0.10	0.50±0.26	0.65±0.17	0.80±0.06
	6	0.2	1.9,1.5,1.3	6.00±1.52	1.00±0.00	0.50±0.25	0.66±0.16	0.87±0.06
	7	0.2	1.9,1.5,0.1	5.30±1.97	0.81±0.17	0.52±0.27	0.62±0.16	0.79±0.06
	8	0.2	1,0.8,0.1	6.05±1.27	0.86±0.16	0.42±0.21	0.57±0.16	0.66±0.05
	9	0.8	2,1.2,0.6	6.35±1.56	1.00±0.00	0.44±0.26	0.62±0.17	0.73±0.08
	10	0.8	1.9,1.5,1.3	5.80±1.19	1.00±0.00	0.53±0.19	0.68±0.13	0.78±0.09
	11	0.8	1.9,1.5,0.1	5.85±1.63	0.85±0.17	0.45±0.23	0.58±0.14	0.72±0.09
	12	0.8	1,0.8,0.1	5.85±1.46	0.85±0.17	0.45±0.23	0.58±0.16	0.58±0.05
GAFS-SVM-MR	1	0	2,1.2,0.6	5.30±1.49	0.98±0.07	0.60±0.23	0.73±0.14	0.75±0.08
	2	0	1.9,1.5,1.3	5.30±1.45	1.00±0.00	0.61±0.24	0.74±0.16	0.82±0.09
	3	0	1.9,1.5,0.1	4.95±1.66	0.80±0.16	0.57±0.25	0.65±0.16	0.78±0.07
	4	0	1,0.8,0.1	6.25±1.48	0.81±0.17	0.36±0.22	0.51±0.15	0.65±0.04
	5	0.2	2,1.2,0.6	6.15±1.59	0.98±0.07	0.46±0.24	0.63±0.15	0.79±0.07
	6	0.2	1.9,1.5,1.3	5.80±1.32	1.00±0.00	0.53±0.22	0.68±0.14	0.83±0.05
	7	0.2	1.9,1.5,0.1	5.05±1.79	0.83±0.17	0.57±0.29	0.66±0.20	0.79±0.08
	8	0.2	1,0.8,0.1	6.75±1.97	0.91±0.14	0.33±0.30	0.52±0.19	0.67±0.05
	9	0.8	2,1.2,0.6	5.85±1.56	0.98±0.07	0.51±0.25	0.67±0.16	0.74±0.10
	10	0.8	1.9,1.5,1.3	5.90±1.58	1.00±0.00	0.51±0.26	0.67±0.17	0.76±0.08
	11	0.8	1.9,1.5,0.1	5.40±1.53	0.86±0.16	0.53±0.22	0.64±0.15	0.73±0.10
	12	0.8	1,0.8,0.1	6.10±1.51	0.88±0.16	0.42±0.23	0.57±0.15	0.59±0.06

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; GAFS-ANN-MR, Genetic algorithm feature selection and artificial neural network algorithm applied to martingale residual; GAFS-RF-MR, Genetic algorithm feature selection and random forest algorithm applied to martingale residual; GAFS-SVM-MR, Genetic algorithm feature selection and support vector machine algorithm applied to martingale residual

Table 2. Evaluation metrics for different combinations of machine learning and variable selection methods applied to martingale residual with a 3-year follow-up

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-MR	13	0.5	2,1.2,0.6	4.35±2.10	0.95±0.12	0.75±0.32	0.81±0.20	1.22±0.08
	14	0.5	1.9,1.5,1.3	3.95±1.84	1.00±0.00	0.84±0.30	0.89±0.20	1.26±0.15
	15	0.5	1.9,1.5,0.1	3.00±1.74	0.73±0.13	0.86±0.23	0.82±0.13	1.20±0.12
	16	0.5	1,0.8,0.1	3.40±1.95	0.70±0.10	0.78±0.29	0.75±0.17	1.11±0.12
	17	0.8	2,1.2,0.6	3.90±2.02	0.93±0.13	0.81±0.31	0.85±0.19	1.12±0.12
	18	0.8	1.9,1.5,1.3	4.80±2.19	1.00±0.00	0.70±0.36	0.80±0.24	1.15±0.12
	19	0.8	1.9,1.5,0.1	3.35±2.41	0.75±0.14	0.81±0.33	0.79±0.17	1.09±0.11
	20	0.8	1,0.8,0.1	3.15±1.87	0.73±0.13	0.84±0.26	0.80±0.15	0.93±0.08
RFE-ANN-MR	13	0.5	2,1.2,0.6	3.85±1.18	1.00±0.00	0.85±0.19	0.90±0.13	2.08±0.14
	14	0.5	1.9,1.5,1.3	3.75±1.25	1.00±0.00	0.87±0.20	0.91±0.13	2.18±0.15
	15	0.5	1.9,1.5,0.1	2.75±1.29	0.68±0.07	0.88±0.19	0.81±0.12	2.09±0.15
	16	0.5	1,0.8,0.1	2.95±1.27	0.73±0.13	0.87±0.21	0.82±0.15	1.56±0.14
	17	0.8	2,1.2,0.6	4.80±2.06	0.96±0.10	0.68±0.32	0.77±0.20	1.79±0.17
	18	0.8	1.9,1.5,1.3	3.55±0.94	1.00±0.00	0.90±0.15	0.93±0.10	1.86±0.13
	19	0.8	1.9,1.5,0.1	3.25±2.26	0.75±0.14	0.83±0.31	0.80±0.17	1.81±0.16
	20	0.8	1,0.8,0.1	3.95±2.06	0.76±0.15	0.72±0.30	0.73±0.18	1.35±0.11
RFE-RF-MR	13	0.5	2,1.2,0.6	3.10±1.25	0.91±0.14	0.94±0.18	0.93±0.12	1.21±0.08
	14	0.5	1.9,1.5,1.3	3.05±0.22	1.00±0.00	0.99±0.03	0.99±0.02	1.22±0.15
	15	0.5	1.9,1.5,0.1	2.20±0.41	0.66±0.00	0.96±0.06	0.86±0.04	1.20±0.13
	16	0.5	1,0.8,0.1	3.85±2.32	0.76±0.15	0.74±0.31	0.75±0.16	1.14±0.11
	17	0.8	2,1.2,0.6	3.80±1.93	0.98±0.07	0.85±0.31	0.90±0.20	1.10±0.12
	18	0.8	1.9,1.5,1.3	3.10±0.30	1.00±0.00	0.98±0.05	0.98±0.03	1.13±0.11
	19	0.8	1.9,1.5,0.1	3.05±2.25	0.73±0.13	0.85±0.31	0.81±0.17	1.11±0.11
	20	0.8	1,0.8,0.1	5.00±2.80	0.78±0.16	0.55±0.39	0.63±0.21	0.96±0.09
RFE-SVM-MR	13	0.5	2,1.2,0.6	4.40±2.01	0.93±0.13	0.73±0.31	0.80±0.20	1.23±0.08
	14	0.5	1.9,1.5,1.3	3.50±0.68	1.00±0.00	0.91±0.11	0.94±0.07	1.23±0.17
	15	0.5	1.9,1.5,0.1	3.40±1.63	0.70±0.10	0.78±0.25	0.75±0.16	1.21±0.15
	16	0.5	1,0.8,0.1	4.05±2.35	0.78±0.16	0.71±0.32	0.73±0.18	1.12±0.10
	17	0.8	2,1.2,0.6	4.25±2.19	0.91±0.14	0.75±0.33	0.80±0.21	1.12±0.13
	18	0.8	1.9,1.5,1.3	3.75±1.06	1.00±0.00	0.87±0.17	0.91±0.11	1.12±0.12
	19	0.8	1.9,1.5,0.1	2.90±1.74	0.71±0.12	0.87±0.23	0.82±0.12	1.10±0.11
	20	0.8	1,0.8,0.1	5.25±2.75	0.90±0.15	0.57±0.40	0.68±0.24	0.98±0.09

RFE-GB-MR, Recursive feature elimination and gradient boosting algorithm applied to martingale residual; RFE-ANN-MR, Recursive feature elimination and artificial neural network algorithm applied to martingale residual; RFE-RF-MR, Recursive feature elimination and random forest algorithm applied to martingale residual; RFE-SVM-MR, Recursive feature elimination and support vector machine algorithm applied to martingale residual

Table 2 (continued)

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
GAFS-GB-MR	13	0.5	2,1.2,0.6	6.95±1.53	1.00±0.00	0.34±0.25	0.56±0.17	1.23±0.08
	14	0.5	1.9,1.5,1.3	6.85±1.81	1.00±0.00	0.35±0.30	0.57±0.20	1.29±0.16
	15	0.5	1.9,1.5,0.1	5.85±1.92	0.90±0.15	0.47±0.29	0.61±0.18	1.21±0.14
	16	0.5	1,0.8,0.1	5.15±1.84	0.81±0.17	0.55±0.27	0.63±0.17	1.13±0.11
	17	0.8	2,1.2,0.6	6.05±1.98	0.93±0.13	0.45±0.33	0.61±0.23	1.11±0.12
	18	0.8	1.9,1.5,1.3	7.30±1.38	1.00±0.00	0.28±0.23	0.52±0.15	1.17±0.12
	19	0.8	1.9,1.5,0.1	6.15±1.69	0.86±0.16	0.40±0.23	0.56±0.14	1.10±0.10
	20	0.8	1,0.8,0.1	6.40±1.93	0.91±0.14	0.39±0.29	0.56±0.18	0.94±0.07
GAFS-ANN-MR	13	0.5	2,1.2,0.6	7.15±1.38	0.95±0.12	0.28±0.20	0.50±0.12	2.08±0.14
	14	0.5	1.9,1.5,1.3	6.60±1.27	1.00±0.00	0.40±0.21	0.60±0.14	2.18±0.15
	15	0.5	1.9,1.5,0.1	6.35±1.98	0.93±0.13	0.40±0.32	0.58±0.22	2.09±0.15
	16	0.5	1,0.8,0.1	6.25±1.44	0.85±0.20	0.38±0.21	0.53±0.15	1.57±0.14
	17	0.8	2,1.2,0.6	6.50±1.93	0.95±0.12	0.39±0.32	0.57±0.22	1.79±0.17
	18	0.8	1.9,1.5,1.3	7.10±1.51	1.00±0.00	0.31±0.25	0.54±0.16	1.86±0.13
	19	0.8	1.9,1.5,0.1	6.70±1.71	0.86±0.16	0.31±0.24	0.50±0.14	1.81±0.16
	20	0.8	1,0.8,0.1	6.80±1.60	0.85±0.20	0.29±0.19	0.47±0.10	1.35±0.11
GAFS-RF-MR	13	0.5	2,1.2,0.6	6.20±1.54	0.98±0.07	0.45±0.25	0.63±0.17	1.24±0.08
	14	0.5	1.9,1.5,1.3	5.85±1.38	1.00±0.00	0.52±0.23	0.68±0.15	1.29±0.15
	15	0.5	1.9,1.5,0.1	5.35±1.75	0.86±0.16	0.54±0.26	0.65±0.17	1.24±0.13
	16	0.5	1,0.8,0.1	5.85±1.49	0.88±0.16	0.46±0.22	0.60±0.15	1.12±0.11
	17	0.8	2,1.2,0.6	6.10±1.71	1.00±0.00	0.48±0.28	0.65±0.19	1.11±0.12
	18	0.8	1.9,1.5,1.3	6.40±1.50	1.00±0.00	0.43±0.25	0.62±0.16	1.16±0.11
	19	0.8	1.9,1.5,0.1	6.45±1.95	0.86±0.16	0.35±0.28	0.52±0.17	1.12±0.11
	20	0.8	1,0.8,0.1	6.50±1.31	0.91±0.14	0.37±0.22	0.55±0.17	0.95±0.09
GAFS-SVM-MR	13	0.5	2,1.2,0.6	6.05±1.35	1.00±0.00	0.49±0.22	0.66±0.15	1.23±0.09
	14	0.5	1.9,1.5,1.3	6.00±1.58	1.00±0.00	0.50±0.26	0.66±0.17	1.26±0.17
	15	0.5	1.9,1.5,0.1	5.15±1.30	0.90±0.15	0.59±0.20	0.69±0.14	1.23±0.13
	16	0.5	1,0.8,0.1	5.95±1.23	0.86±0.16	0.44±0.20	0.58±0.15	1.12±0.12
	17	0.8	2,1.2,0.6	6.30±1.34	0.96±0.10	0.43±0.21	0.61±0.15	1.13±0.13
	18	0.8	1.9,1.5,1.3	5.70±1.59	1.00±0.00	0.55±0.26	0.70±0.17	1.14±0.13
	19	0.8	1.9,1.5,0.1	5.85±1.72	0.83±0.17	0.44±0.26	0.57±0.17	1.13±0.13
	20	0.8	1,0.8,0.1	6.80±1.23	0.93±0.13	0.33±0.21	0.53±0.16	0.99±0.09

GAFS-GB-MR, Genetic algorithm feature selection and gradient boosting algorithm applied to martingale residual; GAFS-ANN-MR, Genetic algorithm feature selection and artificial neural network algorithm applied to martingale residual; GAFS-RF-MR, Genetic algorithm feature selection and random forest algorithm applied to martingale residual; GAFS-SVM-MR, Genetic algorithm feature selection and support vector machine algorithm applied to martingale residual

Table 3. Evaluation metrics for different combinations of machine learning and variable selection methods applied to adjust martingale residual with a 1-year follow-up

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-AMR	1	0	2,1.2,0.6	3.65±1.63	0.88±0.16	0.83±0.24	0.85±0.15	2.00±0.37
	2	0	1.9,1.5,1.3	4.35±1.84	1.00±0.00	0.77±0.30	0.85±0.20	2.13±0.29
	3	0	1.9,1.5,0.1	2.45±1.57	0.68±0.07	0.93±0.22	0.85±0.12	1.98±0.37
	4	0	1,0.8,0.1	3.30±1.65	0.73±0.17	0.81±0.21	0.78±0.11	1.47±0.39
	5	0.2	2,1.2,0.6	3.95±2.03	0.93±0.13	0.80±0.31	0.85±0.19	1.86±0.36
	6	0.2	1.9,1.5,1.3	3.50±1.10	1.00±0.00	0.91±0.18	0.94±0.12	2.02±0.30
	7	0.2	1.9,1.5,0.1	2.85±1.56	0.71±0.12	0.88±0.21	0.82±0.12	1.78±0.34
	8	0.2	1,0.8,0.1	3.75±2.65	0.73±0.13	0.74±0.38	0.73±0.21	1.51±0.55
	9	0.8	2,1.2,0.6	4.00±1.65	0.88±0.16	0.77±0.22	0.81±0.13	1.69±0.37
	10	0.8	1.9,1.5,1.3	4.30±1.97	0.98±0.07	0.77±0.32	0.84±0.21	1.80±0.35
	11	0.8	1.9,1.5,0.1	3.05±2.30	0.71±0.12	0.85±0.33	0.80±0.19	1.72±0.32
	12	0.8	1,0.8,0.1	3.15±1.49	0.71±0.12	0.83±0.22	0.79±0.14	1.48±0.49
RFE-ANN-AMR	1	0	2,1.2,0.6	4.40±1.69	0.96±0.10	0.75±0.27	0.82±0.18	2.58±0.40
	2	0	1.9,1.5,1.3	4.80±1.79	1.00±0.00	0.70±0.29	0.80±0.19	2.73±0.42
	3	0	1.9,1.5,0.1	3.65±1.72	0.70±0.10	0.74±0.27	0.72±0.18	2.67±0.49
	4	0	1,0.8,0.1	4.60±2.43	0.78±0.16	0.62±0.33	0.67±0.17	1.71±0.44
	5	0.2	2,1.2,0.6	4.50±1.67	0.95±0.12	0.72±0.24	0.80±0.15	2.43±0.44
	6	0.2	1.9,1.5,1.3	4.35±1.75	0.98±0.07	0.76±0.28	0.83±0.18	2.74±0.40
	7	0.2	1.9,1.5,0.1	4.60±2.25	0.81±0.17	0.64±0.31	0.70±0.18	2.33±0.38
	8	0.2	1,0.8,0.1	4.05±1.93	0.80±0.16	0.72±0.27	0.75±0.16	1.74±0.51
	9	0.8	2,1.2,0.6	4.65±2.15	0.93±0.13	0.69±0.33	0.77±0.21	2.06±0.41
	10	0.8	1.9,1.5,1.3	4.65±1.98	1.00±0.00	0.72±0.33	0.81±0.22	2.20±0.30
	11	0.8	1.9,1.5,0.1	3.50±1.76	0.78±0.16	0.80±0.23	0.80±0.13	2.15±0.30
	12	0.8	1,0.8,0.1	4.85±2.36	0.81±0.17	0.60±0.33	0.67±0.18	1.61±0.46
RFE-RF-AMR	1	0	2,1.2,0.6	3.75±1.99	0.88±0.16	0.81±0.29	0.83±0.17	2.02±0.38
	2	0	1.9,1.5,1.3	3.25±0.71	1.00±0.00	0.95±0.11	0.97±0.08	2.06±0.31
	3	0	1.9,1.5,0.1	2.30±0.57	0.68±0.07	0.95±0.07	0.86±0.04	2.01±0.44
	4	0	1,0.8,0.1	4.40±2.32	0.81±0.17	0.67±0.33	0.72±0.19	1.48±0.42
	5	0.2	2,1.2,0.6	3.15±1.53	0.83±0.17	0.89±0.20	0.87±0.12	1.91±0.40
	6	0.2	1.9,1.5,1.3	3.15±0.67	1.00±0.00	0.97±0.11	0.98±0.07	1.99±0.34
	7	0.2	1.9,1.5,0.1	2.15±0.48	0.66±0.00	0.97±0.08	0.87±0.05	1.76±0.36
	8	0.2	1,0.8,0.1	4.20±2.68	0.80±0.16	0.70±0.38	0.73±0.21	1.50±0.55
	9	0.8	2,1.2,0.6	3.25±1.61	0.85±0.17	0.88±0.22	0.87±0.13	1.63±0.40
	10	0.8	1.9,1.5,1.3	3.15±0.67	1.00±0.00	0.97±0.11	0.98±0.07	1.70±0.34
	11	0.8	1.9,1.5,0.1	2.65±1.22	0.70±0.10	0.90±0.15	0.83±0.07	1.72±0.32
	12	0.8	1,0.8,0.1	3.85±1.63	0.80±0.16	0.75±0.21	0.77±0.12	1.49±0.49

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, root-mean-squared error; RFE-GB-AMR, Recursive feature elimination and gradient boosting algorithm applied to adjust martingale residual; RFE-ANN-AMR, Recursive feature elimination and artificial neural network algorithm applied to adjust martingale residual; RFE-RF-AMR, Recursive feature elimination and random forest algorithm applied to adjust martingale residual

Table 3 (continued)

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-SVM-AMR	1	0	2,1.2,0.6	4.35±2.08	0.90±0.15	0.72±0.31	0.78±0.19	2.00±0.40
	2	0	1.9,1.5,1.3	4.20±1.39	1.00±0.00	0.80±0.23	0.86±0.15	2.02±0.33
	3	0	1.9,1.5,0.1	3.45±1.60	0.76±0.15	0.80±0.25	0.79±0.17	2.04±0.46
	4	0	1,0.8,0.1	4.60±2.70	0.80±0.16	0.63±0.37	0.68±0.19	1.57±0.50
	5	0.2	2,1.2,0.6	4.05±1.50	0.91±0.14	0.78±0.22	0.82±0.14	1.94±0.39
	6	0.2	1.9,1.5,1.3	4.40±1.31	1.00±0.00	0.76±0.21	0.84±0.14	2.01±0.36
	7	0.2	1.9,1.5,0.1	3.15±1.38	0.76±0.15	0.85±0.18	0.82±0.10	1.79±0.37
	8	0.2	1,0.8,0.1	4.55±2.11	0.78±0.16	0.63±0.31	0.68±0.20	1.63±0.60
	9	0.8	2,1.2,0.6	4.45±2.50	0.86±0.16	0.69±0.35	0.75±0.20	1.70±0.48
	10	0.8	1.9,1.5,1.3	4.20±1.28	1.00±0.00	0.80±0.21	0.86±0.14	1.68±0.32
	11	0.8	1.9,1.5,0.1	2.75±1.25	0.71±0.12	0.90±0.19	0.83±0.13	1.73±0.34
	12	0.8	1,0.8,0.1	4.40±2.30	0.80±0.16	0.66±0.32	0.71±0.19	1.53±0.50
GAFS-GB-AMR	1	0	2,1.2,0.6	6.65±2.10	0.95±0.12	0.36±0.32	0.56±0.20	2.00±0.35
	2	0	1.9,1.5,1.3	5.80±1.88	0.98±0.07	0.52±0.29	0.67±0.19	2.12±0.25
	3	0	1.9,1.5,0.1	5.90±1.33	0.83±0.17	0.43±0.20	0.56±0.14	2.05±0.39
	4	0	1,0.8,0.1	5.30±1.45	0.86±0.16	0.55±0.21	0.65±0.13	1.49±0.39
	5	0.2	2,1.2,0.6	6.55±1.73	0.96±0.10	0.39±0.27	0.58±0.18	1.88±0.35
	6	0.2	1.9,1.5,1.3	5.90±1.97	1.00±0.00	0.51±0.32	0.67±0.21	2.05±0.29
	7	0.2	1.9,1.5,0.1	5.60±1.63	0.81±0.17	0.47±0.26	0.58±0.18	1.78±0.34
	8	0.2	1,0.8,0.1	6.25±1.80	0.90±0.15	0.40±0.26	0.57±0.17	1.55±0.53
	9	0.8	2,1.2,0.6	5.70±1.86	0.98±0.07	0.54±0.31	0.68±0.22	1.66±0.38
	10	0.8	1.9,1.5,1.3	6.45±1.76	1.00±0.00	0.42±0.29	0.61±0.19	1.82±0.33
	11	0.8	1.9,1.5,0.1	6.35±2.03	0.86±0.16	0.37±0.29	0.53±0.18	1.74±0.32
	12	0.8	1,0.8,0.1	5.55±1.73	0.88±0.16	0.51±0.24	0.63±0.15	1.52±0.49
GAFS-ANN-AMR	1	0	2,1.2,0.6	6.00±1.55	0.86±0.19	0.43±0.26	0.57±0.20	2.59±0.40
	2	0	1.9,1.5,1.3	6.40±1.53	0.93±0.13	0.40±0.25	0.57±0.18	2.75±0.42
	3	0	1.9,1.5,0.1	6.40±1.66	0.86±0.22	0.36±0.23	0.53±0.15	2.67±0.49
	4	0	1,0.8,0.1	6.35±2.32	0.86±0.19	0.37±0.32	0.53±0.18	1.72±0.45
	5	0.2	2,1.2,0.6	6.55±1.50	0.91±0.14	0.36±0.23	0.55±0.15	2.43±0.43
	6	0.2	1.9,1.5,1.3	7.45±1.50	0.98±0.07	0.25±0.23	0.49±0.14	2.74±0.40
	7	0.2	1.9,1.5,0.1	6.40±1.60	0.91±0.14	0.39±0.23	0.56±0.15	2.34±0.38
	8	0.2	1,0.8,0.1	6.15±1.95	0.86±0.19	0.40±0.28	0.56±0.18	1.76±0.51
	9	0.8	2,1.2,0.6	6.65±1.66	0.90±0.15	0.34±0.23	0.52±0.13	2.07±0.43
	10	0.8	1.9,1.5,1.3	7.10±1.51	0.98±0.07	0.30±0.24	0.53±0.16	2.20±0.30
	11	0.8	1.9,1.5,0.1	7.05±1.43	0.88±0.19	0.26±0.19	0.47±0.13	2.16±0.29
	12	0.8	1,0.8,0.1	6.40±1.69	0.85±0.20	0.35±0.24	0.52±0.16	1.58±0.49

RFE-SVM-AMR, Recursive feature elimination and support vector machine algorithm applied to adjust martingale residual; GAFS-GB-AMR, Genetic algorithm feature selection and gradient boosting algorithm applied to adjust martingale residual; GAFS-ANN-AMR, Genetic algorithm feature selection and artificial neural network algorithm applied to adjust martingale residual

Table 3 (continued)

GAFS-RF-AMR	1	0	2,1.2,0.6	5.85±2.15	0.90±0.15	0.47±0.32	0.61±0.19	2.05±0.35
	2	0	1.9,1.5,1.3	6.05±1.14	0.98±0.07	0.48±0.17	0.65±0.10	2.16±0.34
	3	0	1.9,1.5,0.1	5.35±1.53	0.85±0.17	0.53±0.22	0.63±0.14	2.08±0.42
	4	0	1,0.8,0.1	6.05±1.31	0.91±0.14	0.45±0.21	0.60±0.15	1.49±0.43
	5	0.2	2,1.2,0.6	6.65±1.49	0.96±0.10	0.37±0.22	0.57±0.13	1.94±0.42
	6	0.2	1.9,1.5,1.3	5.90±1.29	1.00±0.00	0.51±0.21	0.67±0.14	2.13±0.36
	7	0.2	1.9,1.5,0.1	6.30±1.78	0.85±0.17	0.37±0.25	0.53±0.16	1.84±0.35
	8	0.2	1,0.8,0.1	6.30±1.59	0.88±0.16	0.39±0.24	0.55±0.16	1.52±0.55
	9	0.8	2,1.2,0.6	6.10±1.71	0.93±0.13	0.45±0.26	0.61±0.17	1.67±0.40
	10	0.8	1.9,1.5,1.3	5.90±1.51	0.98±0.07	0.50±0.23	0.66±0.15	1.80±0.34
	11	0.8	1.9,1.5,0.1	5.10±2.12	0.81±0.17	0.55±0.29	0.64±0.17	1.76±0.36
	12	0.8	1,0.8,0.1	5.65±1.87	0.83±0.17	0.47±0.27	0.59±0.17	1.49±0.49
GAFS-SVM-AMR	1	0	2,1.2,0.6	5.90±1.94	0.93±0.13	0.48±0.31	0.63±0.21	2.02±0.43
	2	0	1.9,1.5,1.3	5.70±1.34	1.00±0.00	0.55±0.22	0.70±0.14	2.04±0.37
	3	0	1.9,1.5,0.1	5.10±1.71	0.80±0.16	0.55±0.26	0.63±0.18	2.12±0.46
	4	0	1,0.8,0.1	6.20±1.85	0.86±0.16	0.40±0.26	0.55±0.16	1.58±0.50
	5	0.2	2,1.2,0.6	6.00±1.58	0.96±0.10	0.48±0.24	0.64±0.15	1.91±0.39
	6	0.2	1.9,1.5,1.3	6.15±1.59	1.00±0.00	0.47±0.26	0.65±0.17	2.04±0.36
	7	0.2	1.9,1.5,0.1	5.25±1.58	0.83±0.17	0.54±0.24	0.63±0.16	1.83±0.38
	8	0.2	1,0.8,0.1	6.60±1.63	0.90±0.15	0.35±0.27	0.53±0.19	1.64±0.59
	9	0.8	2,1.2,0.6	6.35±1.63	0.96±0.10	0.42±0.25	0.60±0.16	1.70±0.45
	10	0.8	1.9,1.5,1.3	6.10±1.44	1.00±0.00	0.48±0.24	0.65±0.16	1.73±0.32
	11	0.8	1.9,1.5,0.1	5.50±1.14	0.85±0.17	0.50±0.18	0.62±0.14	1.77±0.35
	12	0.8	1,0.8,0.1	6.35±1.81	0.93±0.13	0.40±0.28	0.58±0.18	1.52±0.50

GAFS-RF-AMR, Genetic algorithm feature selection and random forest algorithm applied to adjust martingale residual; GAFS-SVM-AMR, Genetic algorithm feature selection and support vector machine algorithm applied to adjust martingale residual

Table 4. Evaluation metrics for different combinations of machine learning and variable selection methods applied to adjust martingale residual with a 3-year follow-up

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-AMR	13	0.5	2,1.2,0.6	5.05±1.95	0.98±0.07	0.65±0.31	0.76±0.20	2.07±0.32
	14	0.5	1.9,1.5,1.3	3.85±1.22	1.00±0.00	0.85±0.20	0.90±0.13	1.99±0.22
	15	0.5	1.9,1.5,0.1	3.60±2.18	0.75±0.14	0.77±0.30	0.76±0.17	1.94±0.30
	16	0.5	1,0.8,0.1	3.05±1.76	0.71±0.12	0.85±0.24	0.80±0.13	1.69±0.36
	17	0.8	2,1.2,0.6	4.60±2.25	0.90±0.15	0.68±0.34	0.75±0.21	1.90±0.26
	18	0.8	1.9,1.5,1.3	3.50±0.82	1.00±0.00	0.91±0.13	0.94±0.09	1.88±0.25
	19	0.8	1.9,1.5,0.1	2.90±1.29	0.70±0.10	0.86±0.20	0.81±0.14	1.76±0.30
	20	0.8	1,0.8,0.1	3.75±2.19	0.78±0.16	0.76±0.32	0.77±0.19	1.64±0.34
RFE-ANN-AMR	13	0.5	2,1.2,0.6	4.45±1.79	0.98±0.07	0.75±0.28	0.82±0.18	2.76±0.36
	14	0.5	1.9,1.5,1.3	4.20±1.32	1.00±0.00	0.80±0.22	0.86±0.14	2.74±0.21
	15	0.5	1.9,1.5,0.1	3.15±1.87	0.75±0.14	0.85±0.26	0.81±0.15	2.66±0.35
	16	0.5	1,0.8,0.1	4.45±2.32	0.76±0.15	0.64±0.33	0.68±0.18	1.90±0.37
	17	0.8	2,1.2,0.6	5.10±2.26	0.93±0.13	0.61±0.34	0.72±0.21	2.40±0.30
	18	0.8	1.9,1.5,1.3	5.50±1.90	1.00±0.00	0.58±0.31	0.72±0.21	2.51±0.26
	19	0.8	1.9,1.5,0.1	3.60±1.95	0.73±0.13	0.76±0.29	0.75±0.18	2.27±0.33
	20	0.8	1,0.8,0.1	3.85±1.46	0.75±0.14	0.73±0.23	0.73±0.16	1.80±0.38
RFE-RF-AMR	13	0.5	2,1.2,0.6	3.10±1.07	0.93±0.13	0.95±0.15	0.94±0.10	2.07±0.36
	14	0.5	1.9,1.5,1.3	3.10±0.30	1.00±0.00	0.98±0.05	0.98±0.03	1.95±0.15
	15	0.5	1.9,1.5,0.1	2.30±0.92	0.68±0.07	0.95±0.11	0.86±0.05	1.93±0.30
	16	0.5	1,0.8,0.1	3.65±2.30	0.75±0.14	0.76±0.33	0.76±0.19	1.70±0.39
	17	0.8	2,1.2,0.6	3.70±1.65	0.88±0.16	0.82±0.22	0.84±0.13	1.92±0.26
	18	0.8	1.9,1.5,1.3	3.20±0.69	1.00±0.00	0.96±0.11	0.97±0.07	1.90±0.27
	19	0.8	1.9,1.5,0.1	2.60±1.60	0.70±0.10	0.91±0.22	0.84±0.12	1.69±0.25
	20	0.8	1,0.8,0.1	3.60±2.21	0.78±0.16	0.79±0.29	0.78±0.15	1.65±0.34
RFE-SVM-AMR	13	0.5	2,1.2,0.6	5.50±2.32	0.96±0.10	0.56±0.36	0.70±0.23	2.12±0.39
	14	0.5	1.9,1.5,1.3	4.75±2.14	1.00±0.00	0.70±0.35	0.80±0.23	1.93±0.16
	15	0.5	1.9,1.5,0.1	3.30±1.52	0.73±0.13	0.81±0.20	0.78±0.11	1.97±0.33
	16	0.5	1,0.8,0.1	4.35±2.45	0.80±0.16	0.67±0.36	0.71±0.22	1.73±0.40
	17	0.8	2,1.2,0.6	5.20±3.20	0.93±0.13	0.60±0.36	0.71±0.23	1.89±0.29
	18	0.8	1.9,1.5,1.3	4.60±2.13	1.00±0.00	0.73±0.35	0.82±0.23	1.88±0.23
	19	0.8	1.9,1.5,0.1	3.95±1.60	0.80±0.16	0.74±0.21	0.76±0.12	1.72±0.35
	20	0.8	1,0.8,0.1	5.05±2.64	0.83±0.17	0.57±0.39	0.66±0.24	1.65±0.40

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, root-mean-squared error; RFE-GB-AMR, Recursive feature elimination and gradient boosting algorithm applied to adjust martingale residual; RFE-ANN-AMR, Recursive feature elimination and artificial neural network algorithm applied to adjust martingale residual; RFE-RF-AMR, Recursive feature elimination and random forest algorithm applied to adjust martingale residual; RFE-SVM-AMR, Recursive feature elimination and support vector machine algorithm applied to adjust martingale residual

Table 4 (continued)

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
GAFS-GB-AMR	13	0.5	2,1.2,0.6	6.25±1.61	0.96±0.10	0.44±0.26	0.61±0.18	2.08±0.32
	14	0.5	1.9,1.5,1.3	6.10±1.80	1.00±0.00	0.48±0.30	0.65±0.20	2.03±0.22
	15	0.5	1.9,1.5,0.1	5.35±1.66	0.86±0.16	0.54±0.26	0.65±0.18	1.98±0.26
	16	0.5	1,0.8,0.1	5.35±1.53	0.85±0.17	0.53±0.20	0.63±0.12	1.71±0.37
	17	0.8	2,1.2,0.6	6.15±1.72	0.96±0.10	0.45±0.29	0.62±0.20	1.88±0.25
	18	0.8	1.9,1.5,1.3	6.50±1.57	1.00±0.00	0.41±0.26	0.61±0.17	1.91±0.26
	19	0.8	1.9,1.5,0.1	5.90±1.86	0.85±0.17	0.44±0.27	0.57±0.17	1.80±0.27
	20	0.8	1,0.8,0.1	6.00±1.77	0.91±0.14	0.45±0.27	0.61±0.18	1.66±0.35
GAFS-ANN-AMR	13	0.5	2,1.2,0.6	6.35±1.87	0.85±0.20	0.36±0.26	0.52±0.16	2.78±0.36
	14	0.5	1.9,1.5,1.3	6.75±1.55	1.00±0.00	0.37±0.25	0.58±0.17	2.74±0.21
	15	0.5	1.9,1.5,0.1	6.80±1.73	0.90±0.15	0.31±0.26	0.51±0.17	2.66±0.35
	16	0.5	1,0.8,0.1	5.30±1.78	0.85±0.25	0.54±0.24	0.64±0.17	1.93±0.39
	17	0.8	2,1.2,0.6	6.20±1.76	0.93±0.13	0.43±0.26	0.60±0.16	2.40±0.31
	18	0.8	1.9,1.5,1.3	7.25±1.37	0.98±0.07	0.28±0.23	0.51±0.16	2.52±0.26
	19	0.8	1.9,1.5,0.1	6.05±1.46	0.95±0.12	0.46±0.23	0.62±0.16	2.27±0.33
	20	0.8	1,0.8,0.1	6.35±2.20	0.86±0.19	0.37±0.31	0.53±0.19	1.81±0.39
GAFS-RF-AMR	13	0.5	2,1.2,0.6	6.05±1.95	0.93±0.13	0.45±0.29	0.61±0.17	2.17±0.32
	14	0.5	1.9,1.5,1.3	6.05±1.63	1.00±0.00	0.49±0.27	0.66±0.18	2.07±0.18
	15	0.5	1.9,1.5,0.1	5.50±1.53	0.86±0.16	0.51±0.22	0.63±0.15	2.02±0.29
	16	0.5	1,0.8,0.1	5.60±1.23	0.90±0.15	0.51±0.19	0.64±0.14	1.72±0.38
	17	0.8	2,1.2,0.6	5.90±1.68	0.96±0.10	0.50±0.27	0.65±0.17	1.93±0.29
	18	0.8	1.9,1.5,1.3	6.40±1.35	1.00±0.00	0.43±0.22	0.62±0.15	1.99±0.25
	19	0.8	1.9,1.5,0.1	5.45±1.66	0.90±0.15	0.54±0.22	0.66±0.12	1.76±0.27
	20	0.8	1,0.8,0.1	5.50±1.76	0.86±0.16	0.51±0.26	0.63±0.17	1.64±0.34
GAFS-SVM-AMR	13	0.5	2,1.2,0.6	6.70±1.94	1.00±0.00	0.38±0.32	0.58±0.21	2.13±0.34
	14	0.5	1.9,1.5,1.3	6.10±1.37	1.00±0.00	0.48±0.22	0.65±0.15	1.90±0.18
	15	0.5	1.9,1.5,0.1	5.70±1.30	0.83±0.17	0.46±0.20	0.58±0.15	2.03±0.32
	16	0.5	1,0.8,0.1	6.40±1.72	0.93±0.13	0.40±0.26	0.57±0.16	1.73±0.38
	17	0.8	2,1.2,0.6	6.20±1.32	0.98±0.07	0.45±0.23	0.63±0.16	1.88±0.33
	18	0.8	1.9,1.5,1.3	6.55±1.76	1.00±0.00	0.40±0.29	0.60±0.19	1.90±0.24
	19	0.8	1.9,1.5,0.1	6.10±1.80	0.88±0.16	0.42±0.28	0.57±0.19	1.76±0.34
	20	0.8	1,0.8,0.1	6.10±1.61	0.90±0.15	0.43±0.25	0.58±0.17	1.65±0.39

GAFS-GB-AMR, Genetic algorithm feature selection and gradient boosting algorithm applied to adjust martingale residual; GAFS-ANN-AMR, Genetic algorithm feature selection and artificial neural network algorithm applied to adjust martingale residual; GAFS-RF-AMR, Genetic algorithm feature selection and random forest algorithm applied to adjust martingale residual; GAFS-SVM-AMR, Genetic algorithm feature selection and support vector machine algorithm applied to adjust martingale residual

Table 5. Evaluation metrics for different combinations of machine learning and variable selection methods applied to deviance residual with a 1-year follow-up

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-DR	1	0	2,1.2,0.6	3.80±1.50	0.96±0.10	0.85±0.23	0.88±0.15	0.91±0.05
	2	0	1.9,1.5,1.3	4.45±1.95	1.00±0.00	0.75±0.32	0.83±0.21	0.96±0.09
	3	0	1.9,1.5,0.1	3.15±2.03	0.75±0.14	0.85±0.27	0.81±0.14	0.91±0.08
	4	0	1,0.8,0.1	3.45±2.13	0.75±0.14	0.80±0.30	0.78±0.17	0.87±0.07
	5	0.2	2,1.2,0.6	3.95±1.84	0.96±0.10	0.82±0.29	0.87±0.19	0.90±0.07
	6	0.2	1.9,1.5,1.3	4.25±1.65	1.00±0.00	0.79±0.27	0.86±0.18	0.95±0.07
	7	0.2	1.9,1.5,0.1	2.80±1.23	0.75±0.14	0.90±0.16	0.85±0.10	0.88±0.09
	8	0.2	1,0.8,0.1	3.60±2.28	0.78±0.16	0.79±0.33	0.78±0.20	0.87±0.08
	9	0.8	2,1.2,0.6	4.15±1.42	1.00±0.00	0.80±0.23	0.87±0.15	0.86±0.07
	10	0.8	1.9,1.5,1.3	3.75±1.51	1.00±0.00	0.87±0.25	0.91±0.16	0.89±0.08
	11	0.8	1.9,1.5,0.1	2.70±1.21	0.71±0.12	0.90±0.19	0.84±0.13	0.86±0.09
	12	0.8	1,0.8,0.1	3.15±1.59	0.75±0.14	0.85±0.23	0.81±0.14	0.84±0.07
RFE-ANN-DR	1	0	2,1.2,0.6	5.35±2.39	0.95±0.12	0.58±0.36	0.70±0.23	1.68±0.06
	2	0	1.9,1.5,1.3	3.95±1.31	1.00±0.00	0.84±0.21	0.89±0.14	1.73±0.06
	3	0	1.9,1.5,0.1	3.30±1.78	0.70±0.10	0.80±0.26	0.76±0.15	1.67±0.07
	4	0	1,0.8,0.1	4.00±2.20	0.75±0.14	0.70±0.31	0.72±0.18	1.37±0.03
	5	0.2	2,1.2,0.6	4.60±2.25	0.95±0.12	0.70±0.35	0.78±0.22	1.61±0.05
	6	0.2	1.9,1.5,1.3	4.80±1.90	1.00±0.00	0.70±0.31	0.80±0.21	1.68±0.06
	7	0.2	1.9,1.5,0.1	3.55±1.90	0.78±0.16	0.80±0.26	0.79±0.15	1.60±0.05
	8	0.2	1,0.8,0.1	4.45±2.23	0.78±0.19	0.65±0.31	0.69±0.17	1.33±0.04
	9	0.8	2,1.2,0.6	4.30±1.75	0.93±0.13	0.75±0.26	0.81±0.17	1.41±0.08
	10	0.8	1.9,1.5,1.3	4.65±1.42	1.00±0.00	0.72±0.23	0.81±0.15	1.46±0.07
	11	0.8	1.9,1.5,0.1	3.40±2.11	0.71±0.12	0.79±0.31	0.76±0.19	1.41±0.07
	12	0.8	1,0.8,0.1	4.30±2.31	0.76±0.15	0.66±0.33	0.70±0.19	1.17±0.05
RFE-RF-DR	1	0	2,1.2,0.6	3.20±1.00	0.95±0.12	0.94±0.15	0.94±0.11	0.93±0.08
	2	0	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.95±0.08
	3	0	1.9,1.5,0.1	2.10±0.30	0.66±0.00	0.98±0.05	0.87±0.03	0.91±0.07
	4	0	1,0.8,0.1	4.95±2.76	0.78±0.16	0.56±0.40	0.63±0.23	0.90±0.07
	5	0.2	2,1.2,0.6	3.00±0.79	0.95±0.12	0.97±0.11	0.96±0.08	0.93±0.06
	6	0.2	1.9,1.5,1.3	3.20±0.69	1.00±0.00	0.96±0.11	0.97±0.07	0.95±0.07
	7	0.2	1.9,1.5,0.1	2.15±0.36	0.68±0.07	0.98±0.05	0.88±0.04	0.89±0.08
	8	0.2	1,0.8,0.1	3.95±2.79	0.76±0.15	0.72±0.40	0.73±0.23	0.89±0.08
	9	0.8	2,1.2,0.6	3.45±1.73	0.91±0.14	0.88±0.25	0.89±0.16	0.87±0.08
	10	0.8	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.89±0.07
	11	0.8	1.9,1.5,0.1	2.25±0.91	0.68±0.07	0.96±0.11	0.87±0.05	0.86±0.08
	12	0.8	1,0.8,0.1	4.10±2.31	0.78±0.16	0.70±0.32	0.73±0.18	0.86±0.06

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, root-mean-squared error; RFE-GB-DR, Recursive feature elimination and gradient boosting algorithm applied to deviance residual; RFE-ANN-DR, Recursive feature elimination and artificial neural network algorithm applied to deviance residual; RFE-RF-DR, Recursive feature elimination and random forest algorithm applied to deviance residual;

Table 5 (continued)

Method	S	C	Non zero coef-ficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-SVM-DR	1	0	2,1.2,0.6	4.80±1.96	0.98±0.07	0.69±0.31	0.78±0.20	0.88±0.07
	2	0	1.9,1.5,1.3	3.75±1.25	1.00±0.00	0.87±0.20	0.91±0.13	0.92±0.09
	3	0	1.9,1.5,0.1	3.15±1.72	0.75±0.14	0.85±0.23	0.81±0.13	0.93±0.10
	4	0	1,0.8,0.1	4.75±2.78	0.80±0.16	0.60±0.40	0.67±0.23	0.92±0.06
	5	0.2	2,1.2,0.6	4.25±1.86	0.96±0.10	0.77±0.29	0.83±0.18	0.93±0.09
	6	0.2	1.9,1.5,1.3	3.45±0.60	1.00±0.00	0.92±0.10	0.95±0.06	0.89±0.07
	7	0.2	1.9,1.5,0.1	3.00±1.21	0.71±0.12	0.85±0.16	0.81±0.09	0.90±0.09
	8	0.2	1,0.8,0.1	5.15±2.88	0.85±0.17	0.56±0.41	0.66±0.23	0.92±0.10
	9	0.8	2,1.2,0.6	4.15±1.69	0.93±0.13	0.77±0.24	0.82±0.15	0.89±0.10
	10	0.8	1.9,1.5,1.3	3.40±0.82	1.00±0.00	0.93±0.13	0.95±0.09	0.85±0.09
	11	0.8	1.9,1.5,0.1	3.40±2.01	0.73±0.13	0.80±0.28	0.77±0.16	0.88±0.11
	12	0.8	1,0.8,0.1	4.75±2.89	0.80±0.16	0.60±0.41	0.67±0.23	0.87±0.07
GAFS-GB-DR	1	0	2,1.2,0.6	6.50±1.90	0.95±0.12	0.39±0.27	0.57±0.15	0.92±0.05
	2	0	1.9,1.5,1.3	6.30±1.83	1.00±0.00	0.45±0.30	0.63±0.20	0.95±0.08
	3	0	1.9,1.5,0.1	6.10±2.26	0.86±0.16	0.41±0.33	0.56±0.20	0.91±0.07
	4	0	1,0.8,0.1	6.30±1.89	0.85±0.17	0.37±0.27	0.53±0.17	0.88±0.06
	5	0.2	2,1.2,0.6	5.85±1.92	0.95±0.12	0.50±0.29	0.65±0.18	0.93±0.07
	6	0.2	1.9,1.5,1.3	6.45±1.57	1.00±0.00	0.42±0.26	0.61±0.17	0.97±0.07
	7	0.2	1.9,1.5,0.1	5.25±1.51	0.85±0.17	0.55±0.22	0.65±0.14	0.90±0.09
	8	0.2	1,0.8,0.1	5.65±1.95	0.86±0.16	0.49±0.27	0.61±0.16	0.88±0.07
	9	0.8	2,1.2,0.6	6.30±1.94	0.98±0.07	0.44±0.32	0.62±0.22	0.87±0.09
	10	0.8	1.9,1.5,1.3	6.75±1.44	1.00±0.00	0.37±0.24	0.58±0.16	0.91±0.09
	11	0.8	1.9,1.5,0.1	5.40±1.31	0.83±0.17	0.51±0.18	0.62±0.12	0.87±0.08
	12	0.8	1,0.8,0.1	5.80±1.98	0.83±0.17	0.45±0.28	0.57±0.17	0.85±0.07
GAFS-ANN-DR	1	0	2,1.2,0.6	6.85±1.78	0.96±0.10	0.34±0.27	0.55±0.17	1.68±0.06
	2	0	1.9,1.5,1.3	6.40±1.42	1.00±0.00	0.43±0.23	0.62±0.15	1.75±0.07
	3	0	1.9,1.5,0.1	6.10±1.83	0.90±0.15	0.43±0.28	0.58±0.19	1.67±0.07
	4	0	1,0.8,0.1	5.90±1.86	0.85±0.22	0.44±0.24	0.57±0.15	1.37±0.03
	5	0.2	2,1.2,0.6	6.60±1.60	0.98±0.07	0.39±0.27	0.58±0.18	1.61±0.05
	6	0.2	1.9,1.5,1.3	6.80±1.67	1.00±0.00	0.36±0.27	0.57±0.18	1.68±0.06
	7	0.2	1.9,1.5,0.1	6.50±1.39	0.91±0.14	0.37±0.21	0.55±0.14	1.61±0.05
	8	0.2	1,0.8,0.1	6.20±1.76	0.83±0.20	0.38±0.23	0.53±0.14	1.33±0.05
	9	0.8	2,1.2,0.6	6.55±1.66	0.93±0.13	0.37±0.28	0.56±0.20	1.42±0.08
	10	0.8	1.9,1.5,1.3	6.75±1.80	1.00±0.00	0.37±0.30	0.58±0.20	1.46±0.07
	11	0.8	1.9,1.5,0.1	6.55±1.63	0.91±0.14	0.36±0.25	0.55±0.16	1.41±0.07
	12	0.8	1,0.8,0.1	5.60±1.66	0.83±0.20	0.48±0.23	0.60±0.15	1.17±0.05

RFE-SVM-DR, Recursive feature elimination and support vector machine algorithm applied to deviance residual; GAFS-GB-DR, Genetic algorithm feature selection and gradient boosting algorithm applied to deviance residual; GAFS-ANN-DR, Genetic algorithm feature selection and artificial neural network algorithm applied to deviance residual

Table 5 (continued)

GAFS-RF-DR	1	0	2,1.2,0.6	5.75±1.55	0.96±0.10	0.52±0.21	0.67±0.12	0.95±0.06
	2	0	1.9,1.5,1.3	5.80±0.83	1.00±0.00	0.53±0.13	0.68±0.09	1.01±0.08
	3	0	1.9,1.5,0.1	5.15±1.53	0.86±0.16	0.57±0.22	0.67±0.15	0.94±0.08
	4	0	1,0.8,0.1	6.25±1.61	0.88±0.16	0.40±0.24	0.56±0.16	0.90±0.06
	5	0.2	2,1.2,0.6	6.15±1.30	0.96±0.10	0.45±0.20	0.62±0.13	0.96±0.05
	6	0.2	1.9,1.5,1.3	6.35±1.92	1.00±0.00	0.44±0.32	0.62±0.21	1.01±0.08
	7	0.2	1.9,1.5,0.1	6.00±2.05	0.86±0.16	0.43±0.29	0.57±0.18	0.93±0.07
	8	0.2	1,0.8,0.1	5.70±1.41	0.85±0.17	0.47±0.23	0.60±0.17	0.89±0.09
	9	0.8	2,1.2,0.6	6.05±1.87	0.96±0.10	0.47±0.28	0.63±0.17	0.89±0.07
	10	0.8	1.9,1.5,1.3	5.75±1.48	1.00±0.00	0.54±0.24	0.69±0.16	0.94±0.05
	11	0.8	1.9,1.5,0.1	5.60±1.78	0.85±0.17	0.49±0.26	0.61±0.17	0.89±0.08
	12	0.8	1,0.8,0.1	5.35±1.38	0.85±0.17	0.53±0.19	0.63±0.13	0.85±0.06
GAFS-SVM-DR	1	0	2,1.2,0.6	5.45±1.35	1.00±0.00	0.59±0.22	0.72±0.15	0.89±0.06
	2	0	1.9,1.5,1.3	6.00±1.89	1.00±0.00	0.50±0.31	0.66±0.21	0.93±0.08
	3	0	1.9,1.5,0.1	5.25±1.55	0.80±0.16	0.52±0.24	0.61±0.17	0.93±0.08
	4	0	1,0.8,0.1	6.30±1.59	0.83±0.17	0.36±0.24	0.52±0.16	0.92±0.06
	5	0.2	2,1.2,0.6	6.50±1.90	0.96±0.10	0.40±0.28	0.58±0.17	0.94±0.08
	6	0.2	1.9,1.5,1.3	6.00±1.33	1.00±0.00	0.50±0.22	0.66±0.14	0.94±0.07
	7	0.2	1.9,1.5,0.1	5.05±1.46	0.81±0.17	0.56±0.20	0.65±0.13	0.91±0.07
	8	0.2	1,0.8,0.1	6.20±1.47	0.86±0.16	0.40±0.27	0.55±0.21	0.93±0.08
	9	0.8	2,1.2,0.6	6.20±1.43	1.00±0.00	0.46±0.23	0.64±0.15	0.90±0.10
	10	0.8	1.9,1.5,1.3	6.40±1.53	1.00±0.00	0.43±0.25	0.62±0.17	0.88±0.07
	11	0.8	1.9,1.5,0.1	5.35±1.38	0.88±0.16	0.55±0.22	0.66±0.16	0.90±0.10
	12	0.8	1,0.8,0.1	6.30±1.65	0.88±0.16	0.39±0.24	0.55±0.15	0.87±0.06

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; GAFS-RF-DR, Genetic algorithm feature selection and random forest algorithm applied to deviance residual; GAFS-SVM-DR, Genetic algorithm feature selection and support vector machine algorithm applied to deviance residual

Table 6. Evaluation metrics for different combinations of machine learning and variable selection methods applied to deviance residual with a 3-year follow-up

Method	S	C	Nonzerocoef-ficients	No. V	Sensitivity	Specificity	TA	RMSE
RFE-GB-DR	13	0.5	2,1.2,0.6	4.25±1.58	1.00±0.00	0.79±0.26	0.86±0.17	1.00±0.06
	14	0.5	1.9,1.5,1.3	3.30±0.57	1.00±0.00	0.95±0.09	0.96±0.06	1.03±0.11
	15	0.5	1.9,1.5,0.1	3.60±2.32	0.75±0.14	0.77±0.33	0.76±0.20	0.97±0.09
	16	0.5	1,0.8,0.1	2.90±1.58	0.71±0.12	0.87±0.24	0.82±0.15	1.04±0.09
	17	0.8	2,1.2,0.6	4.95±2.30	0.96±0.10	0.65±0.36	0.76±0.23	0.99±0.08
	18	0.8	1.9,1.5,1.3	3.75±1.48	1.00±0.00	0.87±0.24	0.91±0.16	1.00±0.08
	19	0.8	1.9,1.5,0.1	2.90±1.91	0.71±0.12	0.87±0.28	0.82±0.17	0.98±0.11
	20	0.8	1,0.8,0.1	4.35±3.06	0.78±0.16	0.66±0.44	0.70±0.25	1.03±0.07
RFE-ANN-DR	13	0.5	2,1.2,0.6	4.35±1.38	1.00±0.00	0.77±0.23	0.85±0.15	1.73±0.06
	14	0.5	1.9,1.5,1.3	4.00±1.52	1.00±0.00	0.83±0.25	0.88±0.16	1.77±0.06
	15	0.5	1.9,1.5,0.1	3.45±1.66	0.75±0.14	0.80±0.23	0.78±0.14	1.72±0.05
	16	0.5	1,0.8,0.1	3.85±2.13	0.76±0.15	0.74±0.30	0.75±0.17	1.50±0.06
	17	0.8	2,1.2,0.6	4.15±1.95	0.85±0.17	0.73±0.26	0.77±0.15	1.64±0.06
	18	0.8	1.9,1.5,1.3	4.40±1.93	1.00±0.00	0.76±0.32	0.84±0.21	1.67±0.04
	19	0.8	1.9,1.5,0.1	2.90±1.25	0.71±0.12	0.87±0.16	0.82±0.09	1.62±0.06
	20	0.8	1,0.8,0.1	4.00±1.83	0.73±0.17	0.70±0.25	0.71±0.15	1.44±0.06
RFE-RF-DR	13	0.5	2,1.2,0.6	3.05±0.75	0.96±0.10	0.97±0.11	0.97±0.07	1.02±0.07
	14	0.5	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.02±0.10
	15	0.5	1.9,1.5,0.1	2.05±0.22	0.66±0.00	0.99±0.03	0.88±0.02	0.99±0.09
	16	0.5	1,0.8,0.1	3.60±2.28	0.76±0.15	0.78±0.30	0.77±0.15	1.08±0.08
	17	0.8	2,1.2,0.6	3.40±1.23	0.95±0.12	0.90±0.18	0.92±0.12	1.03±0.08
	18	0.8	1.9,1.5,1.3	3.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.09
	19	0.8	1.9,1.5,0.1	2.05±0.22	0.66±0.00	0.99±0.03	0.88±0.02	0.98±0.09
	20	0.8	1,0.8,0.1	3.95±2.37	0.73±0.13	0.70±0.34	0.71±0.20	1.05±0.08
RFE-SVM-DR	13	0.5	2,1.2,0.6	4.50±1.39	0.98±0.07	0.74±0.21	0.82±0.14	1.01±0.07
	14	0.5	1.9,1.5,1.3	3.75±1.25	1.00±0.00	0.87±0.20	0.91±0.13	0.98±0.10
	15	0.5	1.9,1.5,0.1	2.50±0.68	0.70±0.10	0.93±0.11	0.85±0.08	0.99±0.10
	16	0.5	1,0.8,0.1	4.45±2.21	0.80±0.16	0.65±0.31	0.70±0.18	1.08±0.09
	17	0.8	2,1.2,0.6	4.35±1.84	0.96±0.10	0.75±0.28	0.82±0.18	0.99±0.08
	18	0.8	1.9,1.5,1.3	3.60±1.18	1.00±0.00	0.90±0.19	0.93±0.13	0.98±0.07
	19	0.8	1.9,1.5,0.1	3.05±1.57	0.75±0.14	0.86±0.23	0.82±0.15	0.99±0.10
	20	0.8	1,0.8,0.1	5.60±2.74	0.85±0.17	0.49±0.39	0.61±0.23	1.05±0.08

Values are presented as mean±SD

S, Scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; RFE-GB-DR, Recursive feature elimination and gradient boosting algorithm applied to deviance residual; RFE-ANN-DR, Recursive feature elimination and artificial neural network algorithm applied to deviance residual; RFE-RF-DR, Recursive feature elimination and random forest algorithm applied to deviance residual; RFE-SVM-DR, Recursive feature elimination and support vector machine algorithm applied to deviance residual

Table 6 (continued)

Method	S	C	Non zero coefficients	No. V	Sensitivity	Specificity	TA	RMSE
GAFS-GB-DR	13	0.5	2,1.2,0.6	6.20±1.60	1.00±0.00	0.46±0.26	0.64±0.17	1.01±0.07
	14	0.5	1.9,1.5,1.3	6.05±1.66	1.00±0.00	0.49±0.27	0.66±0.18	1.03±0.11
	15	0.5	1.9,1.5,0.1	5.35±1.92	0.85±0.17	0.53±0.28	0.63±0.17	0.99±0.08
	16	0.5	1,0.8,0.1	5.70±1.65	0.85±0.17	0.47±0.27	0.60±0.19	1.07±0.10
	17	0.8	2,1.2,0.6	6.75±1.37	1.00±0.00	0.37±0.22	0.58±0.15	0.99±0.07
	18	0.8	1.9,1.5,1.3	5.90±1.77	1.00±0.00	0.51±0.29	0.67±0.19	1.00±0.09
	19	0.8	1.9,1.5,0.1	6.00±1.37	0.83±0.17	0.41±0.19	0.55±0.13	0.99±0.10
	20	0.8	1,0.8,0.1	5.85±1.89	0.90±0.15	0.47±0.29	0.61±0.19	1.02±0.07
GAFS-ANN-DR	13	0.5	2,1.2,0.6	6.50±1.79	1.00±0.00	0.41±0.29	0.61±0.19	1.73±0.06
	14	0.5	1.9,1.5,1.3	6.45±1.14	1.00±0.00	0.42±0.19	0.61±0.12	1.77±0.06
	15	0.5	1.9,1.5,0.1	6.00±1.83	0.93±0.13	0.46±0.26	0.62±0.15	1.73±0.05
	16	0.5	1,0.8,0.1	6.75±1.97	0.90±0.15	0.32±0.29	0.51±0.19	1.51±0.06
	17	0.8	2,1.2,0.6	6.75±1.74	0.96±0.10	0.35±0.26	0.56±0.15	1.64±0.07
	18	0.8	1.9,1.5,1.3	6.05±1.82	1.00±0.00	0.49±0.30	0.66±0.20	1.68±0.04
	19	0.8	1.9,1.5,0.1	5.95±1.87	0.91±0.14	0.46±0.29	0.61±0.19	1.64±0.06
	20	0.8	1,0.8,0.1	6.45±1.35	0.90±0.15	0.37±0.22	0.55±0.17	1.43±0.06
GAFS-RF-DR	13	0.5	2,1.2,0.6	6.20±1.57	0.96±0.10	0.45±0.24	0.62±0.15	1.06±0.07
	14	0.5	1.9,1.5,1.3	5.95±1.23	1.00±0.00	0.50±0.20	0.67±0.13	1.09±0.09
	15	0.5	1.9,1.5,0.1	5.70±2.05	0.83±0.17	0.46±0.29	0.58±0.18	1.05±0.09
	16	0.5	1,0.8,0.1	6.20±1.10	0.90±0.15	0.41±0.15	0.57±0.11	1.08±0.08
	17	0.8	2,1.2,0.6	6.35±1.46	0.98±0.07	0.43±0.23	0.61±0.15	1.03±0.08
	18	0.8	1.9,1.5,1.3	5.55±1.73	1.00±0.00	0.57±0.28	0.71±0.19	1.06±0.08
	19	0.8	1.9,1.5,0.1	5.40±1.56	0.88±0.16	0.54±0.24	0.65±0.17	1.02±0.09
	20	0.8	1,0.8,0.1	6.15±1.72	0.86±0.16	0.40±0.29	0.56±0.21	1.04±0.07
GAFS-SVM-DR	13	0.5	2,1.2,0.6	6.05±1.35	1.00±0.00	0.49±0.22	0.66±0.15	1.01±0.07
	14	0.5	1.9,1.5,1.3	5.70±1.49	1.00±0.00	0.55±0.24	0.70±0.16	1.00±0.09
	15	0.5	1.9,1.5,0.1	5.25±1.48	0.91±0.14	0.58±0.22	0.69±0.15	1.02±0.09
	16	0.5	1,0.8,0.1	6.20±1.88	0.93±0.13	0.43±0.27	0.60±0.17	1.10±0.09
	17	0.8	2,1.2,0.6	6.20±1.60	1.00±0.00	0.46±0.26	0.64±0.17	1.00±0.07
	18	0.8	1.9,1.5,1.3	6.45±1.63	1.00±0.00	0.42±0.27	0.61±0.18	1.02±0.09
	19	0.8	1.9,1.5,0.1	5.60±1.72	0.86±0.16	0.50±0.25	0.62±0.16	1.02±0.10
	20	0.8	1,0.8,0.1	6.75±1.44	0.95±0.12	0.35±0.23	0.55±0.16	1.05±0.07

Values are presented as mean±SD

S, scenario; C, Censoring rate; No. V, Number of selected variables; TA, Total accuracy; RMSE, Root-mean-squared error; GAFS-GB-DR, Genetic algorithm feature selection and gradient boosting algorithm applied to deviance residual; GAFS-ANN-DR, Genetic algorithm feature selection and artificial neural network algorithm applied to deviance residual; GAFS-RF-DR, Genetic algorithm feature selection and random forest algorithm applied to deviance residual; GAFS-SVM-DR, Genetic algorithm feature selection and support vector machine algorithm applied to deviance residual

Application

Data Description

We reviewed the records of all patients admitted to Sina Hospital (Farshchian) in Hamadan, Iran, with a diagnosis of schizophrenia based on ICD-10 (International Classification of Diseases, version 10) from March 2011 to March 2019. A total of 413 patients met the inclusion criteria (patients without a change in disease during each hospitalization and without other psychiatric disorders simultaneously). Approximately 73% of the patients were male, and their ages ranged from 17 to 77 years old with the age at onset of illness between 5.5 and 69.5 years. Descriptive statistics of the demographic and clinical characteristics of schizophrenia patients are presented in Tables 7 and 8. For more detailed information, please refer to the previous study on this data.[13]

Table 7. Description of demographic characteristics of patients with schizophrenia (n=413)

Variable	Variable levels	Frequency (%)
Gender	Male	300 (72.60)
	Female	113 (27.40)
Age (year)	<25	55 (13.32)
36.16±11.18	25-34	154 (37.29)
	35-44	111 (26.87)
	≥45	93 (22.52)
Birth season	Spring	130 (31.47)
	Summer	145 (35.11)
	Autumn	67 (16.22)
	Winter	71 (17.20)
Education status	Illiterate	36 (8.72)
	Under diploma	237 (57.38)
	Diploma	102 (24.70)
	Academic	38 (9.20)
Marital status	Married	130 (31.48)
	Separated/Divorced/Widowed	50 (12.11)
	Single	233 (56.41)
Number of children	0	282 (68.28)
	1-2	74 (17.92)
	≥3	57 (13.80)
Employment status	Employed	74 (17.92)
	Housewife	45 (10.89)
	Unemployed/Disabled/Retired	294 (71.19)
Residence status	Urban	279 (67.55)
	Rural	134 (32.45)
The township of residence	Hamadan	163 (39.47)
	Other Hamadan province townships	197 (47.70)
	Out of Hamadan province	53 (12.83)
Living status	With parents	229 (55.45)

Table 7 (continued)

	With spouse	115 (27.84)
	With siblings/children/other people or alone	69 (16.71)
Having homogeneous sibling	Yes	369 (89.35)
	No	44 (10.65)
Number of siblings	0-1	43 (10.41)
	2-3	98 (23.73)
	≥4	272 (65.86)
Having a history of emotional distress	Yes	286 (69.25)
	No	127 (30.75)
Having a history of arrest or prison	Yes	43 (10.41)
	No	370 (89.59)
Having a history of substance abuse	Yes	145 (35.11)
	No	268 (64.89)
Having a history of smoking	Yes	211 (51.09)
	No	202 (48.91)

Table 8. Description of clinical characteristics of patients with schizophrenia (n=413)

Variable	Variable levels	Frequency (%)
Age at onset of illness (year)	<20	129 (31.24)
26.44±10.48	20-29	166 (40.19)
	30-39	74 (17.92)
	≥40	44 (10.65)
Duration of illness (year)	<1	42 (10.17)
9.72±8.32	1-9	213 (51.57)
	10-19	92 (22.28)
	≥20	66 (15.98)
Number of previous psychiatric hospitalizations	0	219 (53.03)
	1	79 (19.13)
	2	45 (10.90)
	3	35 (8.47)
	4	17 (4.11)
	≥5	18 (4.36)
Having a family history of psychiatric illness	Yes	181 (43.83)
	No	232 (56.17)
Having a history of medical diseases	Yes	177 (42.90)
	No	236 (57.10)
Having a history of non-adherence to antipsychotic drugs	Yes	303 (73.37)
	No	110 (26.63)
Having a history of the suicide attempt	Yes	81 (19.61)
	No	332 (80.39)

Results

Due to the high relapse rate in schizophrenia, early identification of critical factors associated with relapse is essential for timely treatment and reversal of disease progression. In this study, we focused on the occurrence of rehospitalization in patients with schizophrenia, which is a recurrent event. The aim of this study was to utilize heuristic approaches for selecting predictor variables in the field of recurrent events, with an application to schizophrenia disorder. For this purpose, we used the most performed combination of machine learning and variable selection methods applied to risk indices, for schizophrenia data. The result presented in Figure 1 showed that selected variables related to the risk of rehospitalization in schizophrenia patients were the number of children, age, marital status, and history of substance abuse.

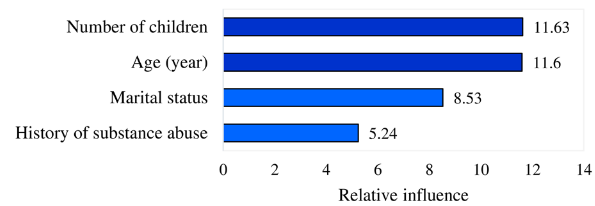

Figure 1. Variable importance plot for RFE-RF-DR (recursive feature elimination method and random forest algorithm applied to deviance residual) model

Figure 2 represents partial dependence plots for selected predictor variables. Based on the figure, having no children, age under 40 and above 50, being separated, divorced, or widowed, and having a positive history of substance abuse are related to a higher predicted risk of experiencing more events (rehospitalization).

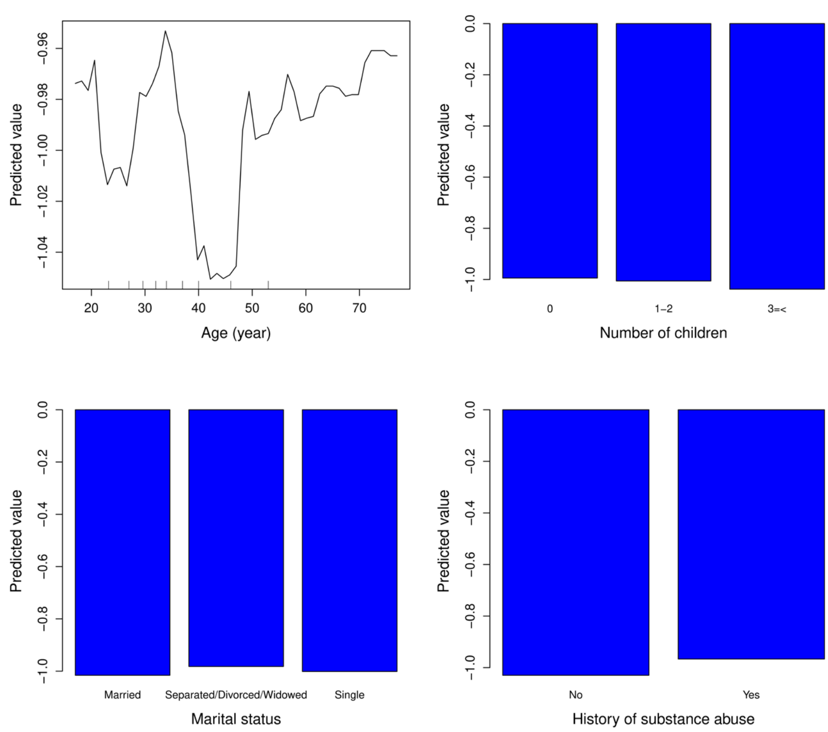

Figure 2. Partial dependence plots for predictor variables selected by RFE-RF-DR (recursive feature elimination method and random forest algorithm applied to deviance residual) model; A higher predicted value represents a higher chance of experiencing more events during a certain period of time

Figure 3 also shows the interaction between the three important variables of the number of children, age, and marital status in predicting the risk of experiencing more events. According to the results, the lower predicted risks of experiencing more events are related to patients who were separated, divorced, or widowed with three or more children. In addition, the relationship between age and risk of rehospitalization was not consistent across all patient subgroups. For instance, in those patients who were single (with no children), the risk increases as age increases (the left-top panel).

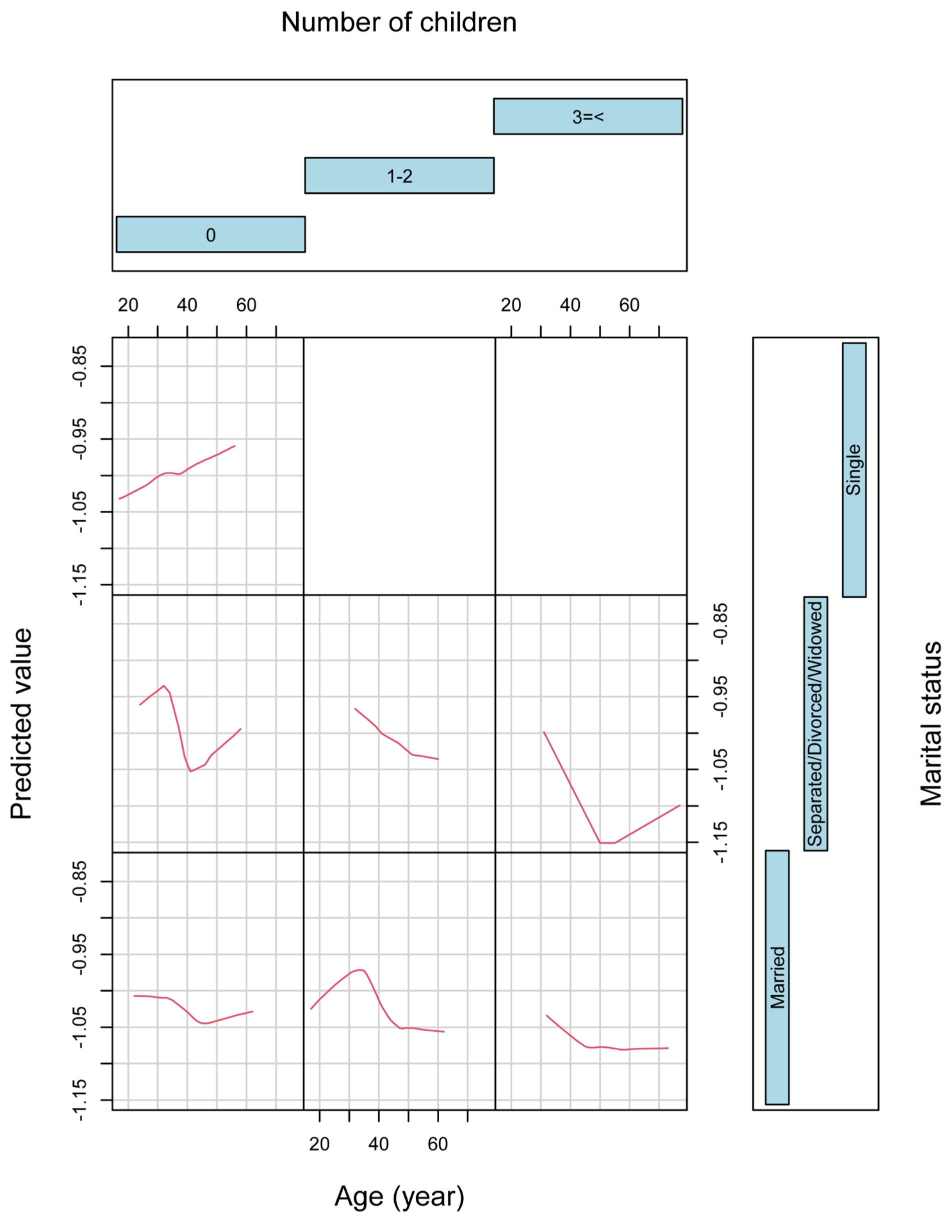

Figure 3. Coplot for the interaction of the three topmost important predictor variables of the RFE-RF-DR (recursive feature elimination method and random forest algorithm applied to deviance residual) model; A higher predicted value represents a higher chance of experiencing more events during a certain period of time.

Discussion

In this study, we used a two-step algorithm for recurrent event datasets to select predictor variables. Firstly, we calculated risk indices as single summary statistics for each subject and then applied four different machine learning methods (GB, ANN, RF, and SVM) with two variable selection methods (RFE and GAFS) to three risk indices (MR, AMR, and DR). According to the simulation study, RFE-RF applied to DR had better performance compared to other algorithms (however, the use of the two-step algorithm for DR is restricted to situations where there are no covariates that vary with time). On the other hand, regardless of the risk index type, the combination of various machine learning methods and the GAFS technique (which is very time-consuming compared to RFE) only led to a good sensitivity performance. Duan and Fu's simulation study showed that the GB method applied to DR and AMR results in less loss in ranking the importance of variables compared to other methods.[6] Recently, an extension of Duan and Fu's algorithm for Multi-State data has been used to compare three different machine learning approaches (GB, RF, and ANN) to rank the factors related to disease progression based on their importance, and finally, the best model obtained from the simulation study (GB method applied to the MR) was fitted on the COVID-19 data.[14]

In the current study, the RFE-RF-DR selected the variables of the number of children, age, marital status, and history of substance abuse. Although some studies have suggested that a social support network, including family and children, can reduce the number of relapse events and hospitalizations for individuals with schizophrenia,[15, 16] there is no clear evidence in the literature of a relationship between the number of children a person has and their likelihood of experiencing a relapse. The risk of relapse within three years is found to be higher in individuals diagnosed with schizophrenia and also in psychiatric patients at younger ages.[17] Similarly, age younger than 35 years is related to early relapse in patients with schizophrenia.[18] It can be attributed to the fact that younger patients tend to discontinue their medication.[19] Additionally, being married has been shown to improve quality of life and reduce suicidal thoughts in middle-aged and older individuals with schizophrenia or schizoaffective disorder with depressive symptoms.[20] On the other hand, males who were unmarried, whether they were never married, separated, divorced, or widowed, had a greater probability of developing schizophrenia than those who were married.[21] Moreover, substances use in people with psychosis is generally associated with poorer outcomes, including increased psychotic symptoms and reduced treatment compliance. Specifically, severe cannabis abuse has been identified as a stressor that can trigger relapse in patients with schizophrenia.[22]

Conclusion

Our findings revealed that the proposed machine learning-based model is a promising technique for selecting predictor variables associated with a recurrent outcome when analyzing multivariate time-to-event data with recurrent events. It is suggested to use the methods used in this study and other machines for analyzing other recurrent event datasets and other multivariate survival data. Also, extending these methods to the context of spatially correlated survival data is of great interest.

Statements and Declarations

Conflicts of Interest

The authors declare that they have no conflicts of interest associated with the material of this paper.

Ethics Approval and Consent to Participate

The information for this study was obtained from the medical records of discharged patients, which were not available for obtaining informed consent. All procedures were conducted according to applicable guidelines and regulations. This study was granted a waiver of informed consent and was approved by the Ethical Committee of Hamadan University of Medical Sciences (Ethical code: IR.UMSHA.REC.1398.296).

Funding

The research leading to these results received funding from Hamadan University of Medical Sciences under Grant No. 9804182853.

Availability of data and materials

The data are available upon reasonable requests from the corresponding author.

Authors' Contributions

MA: Data curation, Investigation, Software, and Writing - original draft; LT: Methodology, Supervision, Project Administration, and Writing - review & editing; ShP: Investigation, and Writing - original draft; BA: Software.

Acknowledgments

We would like to appreciate the Vice-chancellor of Education of the Hamadan University of Medical Science for technical support. We also express our gratitude to the staff of Sina (Farshchian) Educational and Medical Center in Hamadan for providing appropriate facilities for collecting patients' information.

Abbreviations

GB: Gradient Boosting
ANN: Artificial Neural Network
RF: Random Forest
SVM: Support Vector Machine
MR: Martingale Residual
AMR: Adjust Martingale Residual
D: Deviance
DR: Deviance Residual
RFE: Recursive Feature Elimination
GAFS: Genetic Algorithm Feature Selection
TA: Total Accuracy
RMSE: Root-mean-squared error

References

1. Lee ET, Wang J. Statistical methods for survival data analysis: John Wiley & Sons; 2003.

2. Cook RJ, Lawless JF. The statistical analysis of recurrent events: Springer; 2007.

3. Emsley R, Chiliza B, Asmal L, Harvey BH. The nature of relapse in schizophrenia. BMC psychiatry. 2013;13(1):1-8.

4. Shashaani S, Vahdat K. Improved feature selection with simulation optimization.

Optim Eng. 2022:1-41.

5. Malley JD, Malley KG, Pajevic S. Statistical learning for biomedical data: Cambridge University Press; 2011.

6. Duan R, Fu H. Estimate variable importance for recurrent event outcomes with an application to identify hypoglycemia risk factors. Stat Med. 2015;34(19):2743-54.

7. Therneau TM, Grambsch PM, Fleming TR. Martingale-based residuals for survival models. Biometrika. 1990;77(1):147-60.

8. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat. 1982;10(4):1100-20.

9. Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol. 1972;34(2):187-202.

10. Box-Steffensmeier JM, Jones BS. Event history modeling: A guide for social scientists: Cambridge University Press; 2004.

11. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46:389-422.

12. Kuhn M, Johnson K. Applied predictive modeling: Springer; 2013.

13. Arayeshgari M, Roshanaei G, Ghaleiha A, Poorolajal J, Tapak L. Investigating factors associated with the number of rehospitalizations among patients with schizophrenia disorder using penalized count regression models. BMC Medical Res Methodol. 2022;22(1):1-13.

14. Alafchi B, Tapak L, Doosti H, Chesneau C, Roshanaei G. A two-step algorithm to estimate variable importance for multi-state data: an application to COVID-19. Comput Model Eng Sci. 2023;135(3):2047-64.

15. Pharoah F, Mari JJ, Rathbone J, Wong W. Family intervention for schizophrenia. Cochrane Database Syst Rev. 2010(12).

16. Pitschel-Walz G, Leucht S, Bäuml J, Kissling W, Engel RR. The effect of family interventions on relapse and rehospitalization in schizophrenia—a meta-analysis. Schizophr Bull. 2001;27(1):73-92.

17. Hui CL, Tang JY, Leung C-M, Wong GH, Chang W-C, Chan SK, et al. A 3-year retrospective cohort study of predictors of relapse in first-episode psychosis in Hong Kong. Aust N Z J Psychiatry. 2013;47(8):746-53.

18. Gündogmus I, Aydin MB, Öz S, Tasçi AB, Uzun Ö. Clinical and demographic factors associated with early relapse in patients with schizophrenia: a naturalistic observation study. Int Clin Psychopharmacol. 2021;36(6):288-95.

19. Hui CL, Chen EY, KC Y, CW L, Chiu CP. Anti-psychotics adherence among out-patients with schizophrenia in Hong Kong. Keio J Med. 2006;55(1):9-14.

20. Nyer M, Kasckow J, Fellows I, Lawrence EC, Golshan S, Solorzano E, et al. The relationship of marital status and clinical characteristics in middle-aged and older patients

with schizophrenia and depressive symptoms. Ann Clin Psychiatry. 2010;22(3):172-9.

21. Kebede D, Alem A, Shibre T, Negash A, Deyassa N, Beyero T. The sociodemographic correlates of schizophrenia in Butajira, rural Ethiopia. Schizophr Res. 2004;69(2-3):133-41.

22. Winklbaur B, Ebner N, Sachs G, Thau K, Fischer G. Substance abuse in patients with schizophrenia. Dialogues Clin Neurosci. 2006;8(1):37-43.