World Cancer Research Journal

WCRJ 2023; 10: e2634
DOI: 10.32113/wcrj_20237_2634

RANDOM SURVIVAL FOREST IN DETERMINATION OF IMPORTANT RISK FACTORS ON OVERALL SURVIVAL AND DISEASE-FREE SURVIVAL IN GASTRIC CANCER PATIENTS

M. SAFARI[1], M. RAHIMI[2], J. FARADMAL[3], B. GHADERI[4], M.R ANARI[5], G. ROSHANAEI[3]

[1]Department of Biostatistics and Epidemiology, Arak University of Medical Sciences, Arak, Iran
[2]Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran
[3]Department of Biostatistics, School of Public Health, Modeling of Noncommunicable Diseases Research center, Hamadan University of Medical Sciences, Hamadan, Iran
[4]Department of Medicine, School of Medicine, Kurdistan University of Medical Sciences, Kurdistan, Iran
[5]Department of Statistics, Shahid Beheshti University, Tehran, Iran

Corresponding Author
Ghodratollah Roshanaei, PhD; e-mail: gh.roshanaei@umsha.ac.ir

ABSTRACT – *Objective:* Although the incidence of stomach cancer is decreasing in the world, its incidence is still high in Iran. Despite different treatments for cancer, disease recurrence, and death may occur in some patients. Various factors affect survival and recurrence after treatment. This study aims to identify factors affecting overall survival (OS) and disease-free survival (DFS) in patients with gastric cancer (GC) using a random survival forest (RSF).

Patients and Methods: In this retrospective study, 553 patients with GC, diagnosed between 2010 and 2018 in Kurdistan province in the west of Iran, were assessed. Important factors of OS and DFS were identified using the COX model and RSF. Analysis of data was implemented by R free software version 3.5.3.

Results: The mean (Standard Deviation(SD)) age of patients was 66.99 (13.3) years. The median of OS and DFS was 18 and 37.5 months, respectively. Using RSF, the important affected factors on OS were tumor grade, stage, age, recurrence, surgery, and metastasis, respectively. Also according to the RSF model, stage, tumor grade, radiotherapy, tumor site, surgery, and age were the important risk factors for DFS. Based on the prediction error criterion, the random survival forest performed well in predicting disease-free survival. meanwhile, both RSF and Cox models had the same performance in predicting overall survival.

Conclusions: Due to the relationship between tumor grade, disease stage and age, the random survival forest identified these variables as important variables in predicting both outcomes, although the Cox model was not able to detect these factors, which indicates better performance of RSF.

KEYWORDS: Gastric cancer, Cox model, Random survival forest, Overall survival, Disease free survival.

INTRODUCTION

Gastric cancer (GC) is the third cause of cancer-related death worldwide, which is also higher among Asians[1,2]. Gastric carcinoma is a fatal disease with low overall survival in the world, and new cases of this disease mostly occur in Asian and South American countries[1-4]. The overall survival of patients is low as the disease is diagnosed in advanced stages, which is associated with metastasis[2,3]. However, diagnosis in the early stages of the disease significantly improves the survival of these patients[4-6].

GC was the fifth cause of cancer and the fourth cause of death in the world in 2020 and one of the most prevalent and deadly cancers on the globe[5,7]. The incidence of GC is high in Asia. A 5-year survival rate of 55-66% is reported for this disease, and the main cause of death after curative surgery of GC is recurrence because most patients experience this outcome. Although the recurrence rate is very low in the early stages of cancer after curative resection (CR), advanced GC cases show a high rate of recurrence after CR[8,9].

Recurrence is one of the key factors affecting the survival of GC patients, and post-CR recurrence of GC usually has destructive effects on survival. Therefore, recurrence patterns should be identified after the CR of GC by determining recurrence timing to provide information about the postoperative follow-up to find recurrence in time[10,11]. The identification of factors affecting recurrence and death allows the classification of patients based on risk prognosis to better manage treatment protocols, which will improve survival and reduce recurrence rates[12].

In many studies on survival, risk factors are identified using the Cox proportional hazards (CPH) model based on the time until the occurrence of the event[13]. The CPH model is the widely used semi-parametric model for modeling factors affecting survival and recurrence. Nonetheless, the presence of limited assumptions, such as the proportionality of hazards, the linear relationships between variables with the hazard, and the limitation of the number of variables in the model, makes this model inefficient in some applications of survival analysis[14]. In the presence of the mentioned limitations, the non-parametric RSF method is a powerful technique for risk prediction in right-censored data, which can be a suitable alternative to the semi-parametric CPH model. The main feature of this method is its proper performance in measuring the importance of each variable in predicting the time to the event. The RSF is a non-parametric method that considers no specific assumptions and is more efficient than the classical methods of survival analysis, particularly when there are many predictor variables with collinearity or the covariates have nonlinear and complex interactions[15,16].

Since the identification of patients' OS and DFS patterns and the affecting factors can help doctors to determine the appropriate treatment to improve survival, this study aims to identify the important variables affecting OS and DFS in GC patients applying the RSF model.

PATIENTS AND METHODS

Patients

This retrospective cohort study was conducted on 553 GC patients referring to Tohid Hospital in Kurdistan province, during 2012-2018. The collected data, including demographic, clinical, and pathological variables, specifically age at the diagnosis, gender, tumor grade, tumor site, disease stage, surgery, chemotherapy, radiotherapy, local recurrence, distant metastasis, the number of chemotherapy courses, history of smoking, and family history of cancer, were extracted from patients' records. The patients' survival status was monitored through periodic visits and telephone calls. The overall survival was defined from diagnosis to death or censoring in months. DFS of patients with surgery was calculated from the time of surgery to the occurrence of local recurrence or metastasis.

Methods

The CPH model and the RSF model (a non-parametric method) were applied in this study. Three spiliting rules of log-rank, log-rank score, and random were used in this research[17,18].
The models were compared by the Integrated Brier Score (IBS) index in which values close to zero indicate better performance of the model. The efficiency and comparison of the models were examined using the prediction error index, which ranges between 0 and 1, with a value of zero meaning accurate prediction or better efficiency[19,20].

Statistical Analysis

Analyses were performed using "random-ForestSRC" and "Survival", a freely available package from the Comprehensive R software (CRAN) in version 3.5.3. A *p*-value < 0.05 was defined as statistically significant.

RESULTS

In this study, 412 (74.5%) out of 553 patients were males, with a mean (SD) age at the diagnosis of 64.1 (13.2) years in the range of 19-94 years. Entirely, 375 (67.8%) patients were dead by the end of the study. Table 1 represents the demographic and clinical characteristics of the patients. The mean and median follow-up period of the patients were 28.6 and 18 months, respectively.

Table 1. Demographic and clinical characteristics of gastric cancer patients.

Variable	Subgroup	N (%)	Median OS	*p*-value
Sex	**Female**	141 (25.5)	23	0.06
	Male	412 (74.5)	18	
Age (year)	**≤ 55**	120 (21.7)	40	<0.001
	56-70	233 (42.1)	20	
	> 70	200 (36.2)	13	
Number of chemo cycle	**1-5**	211(42.4)	16	0.015
	6-10	229(46)	18	
	>11	58(11.6)	32	
Tumor grade	**Well**	70(12.7)	49	<0.001
	Moderate	90(16.3)	16	
	Poor	108(19.5)	10	
	Unknown	51.5	20	
Stage	**II**	55(9.9)	61	<0.001
	III	97(17.5)	17	
	IV	188(34)	12	
	Missing	213(38.5)	22	
Surgery	**No**	360 (65.1)	16	<0.001
	Yes	193 (34.9)	27	
Radiotherapy	**No**	367(66.4)	17	0.03
	Yes	186(33.6)	25	
Chemotherapy	**No**	55(9.9)	18	0.76
	Yes	498(90.1)	19	
Site of Tumor	**Antrum**	120(21.7)	24	0.39
	Body	57(10.3)	17	
	Cardia	277(50.1)	17	
	Fundus	53(9.6)	20	
	Unknown	46(8.3)	24	
Distance metastasis	**No**	367(68)	60	< 0.001
	Yes	177(32)	9	
Local recurrence	**No**	471(85.2)	-	0.004
	Yes	82(14.8)	18	
Smoking	**No**	286(51.7)	35	0.85
	yes	267(48.3)	34	
Family history of cancer	**Yes**	67(12.1)	30	0.34
	No	486(87.9)	35	

The mean and median survival periods of all patients were 45.6 and 43 months, respectively, and the median survival periods in men and women were 43 and 41 months, respectively. OS rates at 1, 3, and 5 years were 86%, 62%, and 31%, respectively. The important factors affecting OS were identified by the RSF method with all three splitting rules (Table 2). The OS of gastric cancer patients is shown in Figure 1.

Table 2. Evaluation indicators of the Cox and RSF models.

	OS	
Model	**Error rate**	**IBS [0,time=71]**
Cox	27.2	0.109
RSF (log.rank.score)	**27.1**	**0.110**
RSF (random)	29.3	0.115
RSF (log.rank)	27.8	0.113

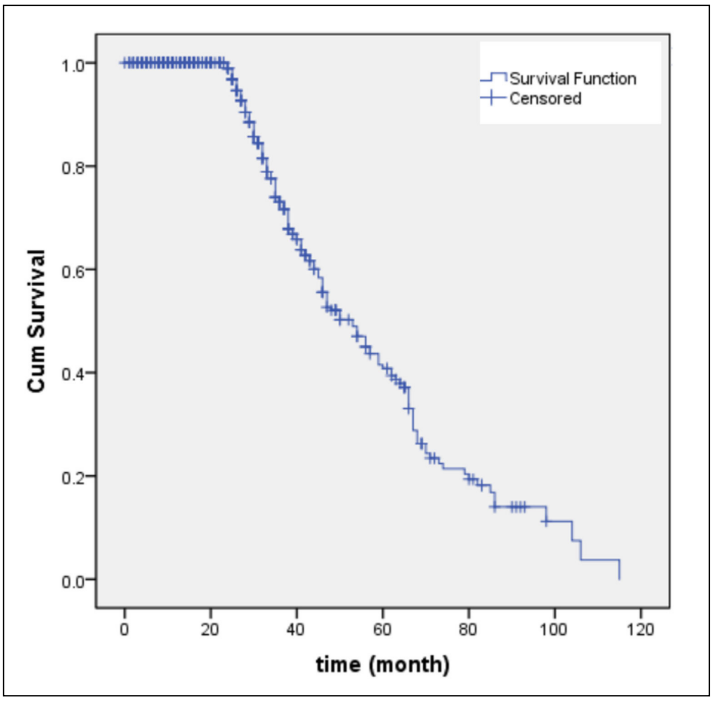

Figure 1. OS of gastric cancer patients.

Table 2 displays the evaluation indices of the goodness of fit of the model. Based on the IBS index and the prediction error rate, the log-rank score splitting method was chosen as the appropriate model. Figure 2 depicts the goodness of fit results of the RSF model with the log-rank score splitting rule to identify the key OS-affecting variables. As shown in Figure 2, tumor grade, disease stage, age at the diagnosis, local recurrence, surgery, and distant metastasis are the important OS-affecting variables. The 5-year survival for the important variables identified using the RSF model is depicted in Figure 3. The 5-year survival probabilities are adjusted for the other variables. The results show that the predicted 5-year survival decreases with increasing age, and the probability of 5-year survival decreases with increasing the disease grade and stage. Also Figure 3 illustrates the estimated 5-year survival probabilities for the levels of the other important variables identified based on the RSF method.

Figure 4 compares the estimated prediction error for CPH and RSF models with different splitting rules. The lowest prediction error was obtained for the RSF model with the log-rank score-splitting rule. The effect of factors on patients' OS was determined using the multivariate Cox model. The results (Table 3) of this model revealed that the age at the diagnosis, local recurrence, disease stage, and surgery influenced the survival of patients.

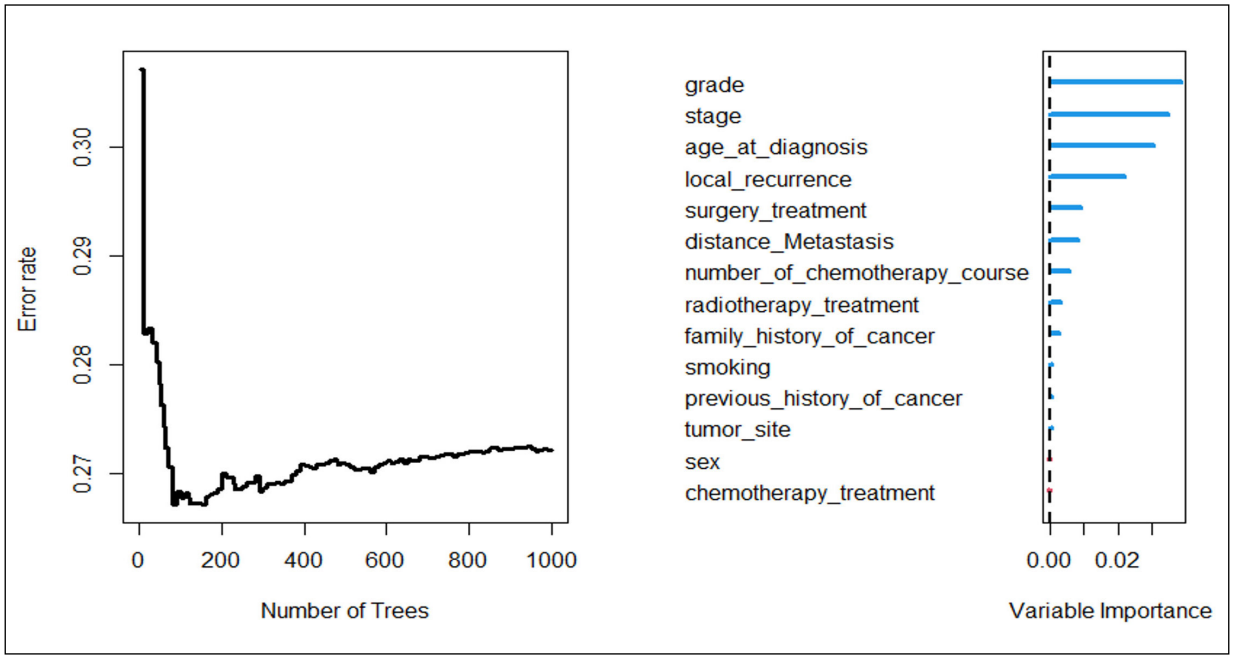

Figure 2. Out-of-Bag variable importance and Error rate of RSF for Log-Rank score Splitting Rule.

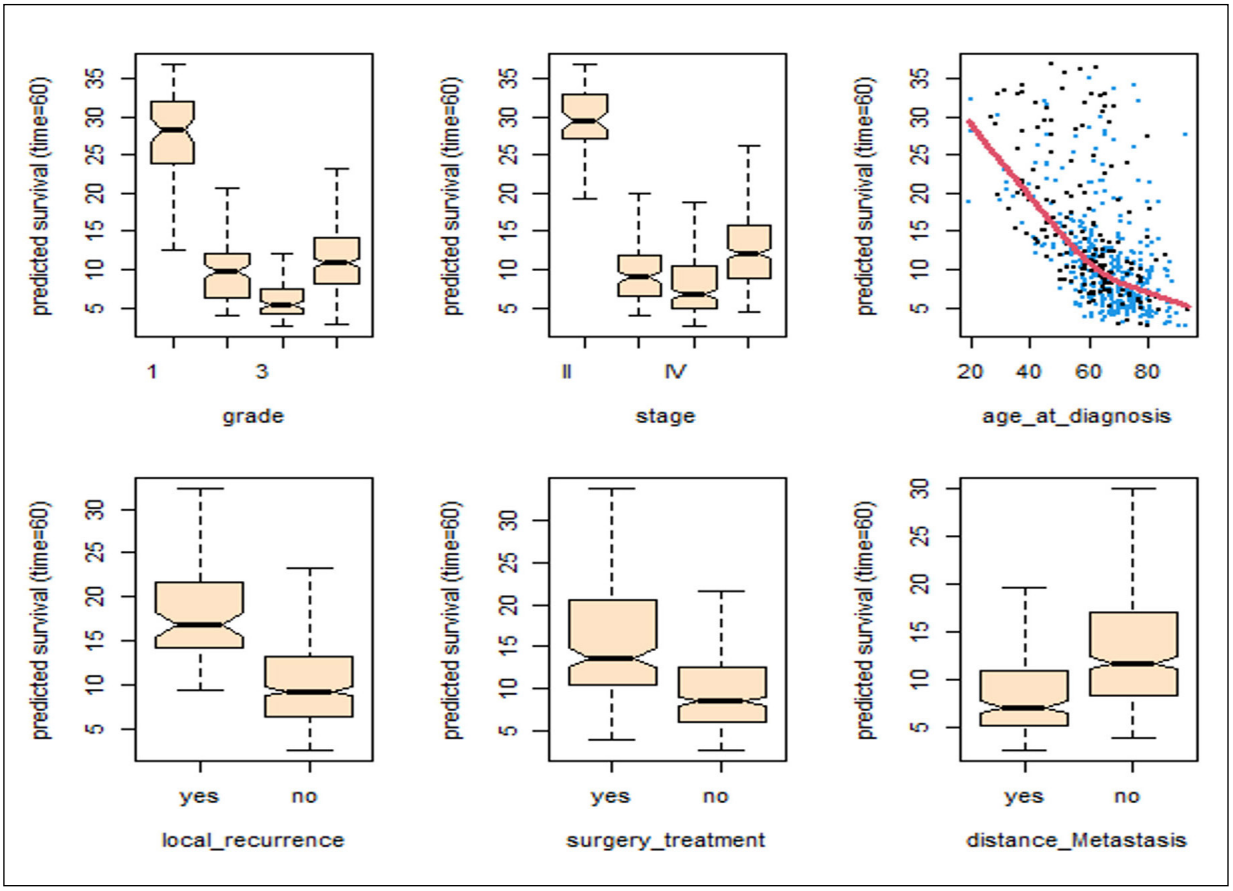

Figure 3. Partial 5-year predicted survival for six most influential variables on survival in gastric cancer data. Values on the vertical axis represent the predicted survival probability for a given predictor, after adjusting for all other predictors.

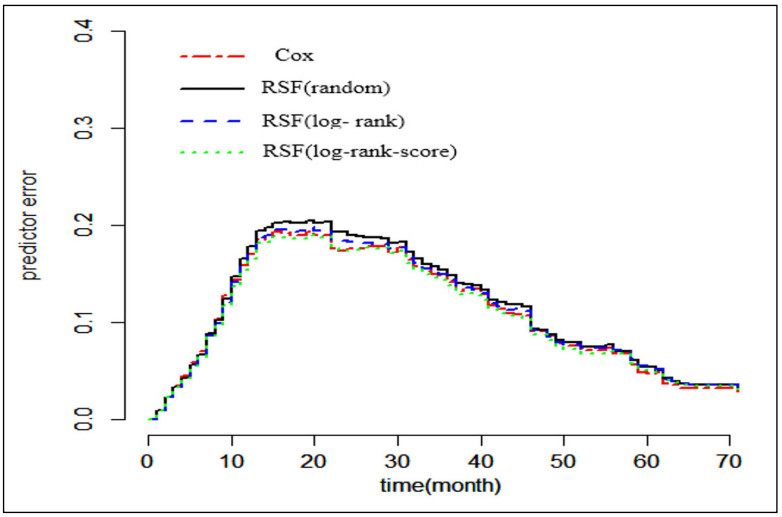

Figure 4. Prediction Error Curves for Cox model (Red), RSF with random splitting rule (black), RSF with log-rank score splitting rule (Green), and RSF with log-rank splitting rule (blue).

Table 3. Multivariable Cox Regression model of Prognostic Factors on OS.

Variable	level	HR	95% CI for HR	
			lower	upper
Age_At_Diagnosis		**1.03**	**1.02**	**1.04**
Tumor Grade	Well	1		
	Moderate	1.73	0.88	3.39
	Poor	2.28	0.72	4.24
Family_History_of_Cancer	No			
	Yes	1.14	0.81	1.61
Smoking	No	1		
	Yes	0.90	0.72	1.13
Local_Recurrence	No	1		
	Yes	**1.84**	**1.33**	**2.55**
Distance_Metastasis	No			
	Yes	1.41	0.97	2.05
Number_of_Chemotherapy_Course	1-5	1		
	6-10	0.85	0.59	1.22
	>11	0.59	0.37	1.12
Tumor Site	Antrum	1		
	Body	1.17	0.79	1.74
	Cardia	0.97	0.73	1.28
	Fundus	0.95	0.63	1.44
Previous_History_of_Cancer	No	1		
	Yes	0.78	0.46	1.32
Sex	Male	1		
	Female	0.9	0.7	1.2
Stage	II			
	III	**2.21**	**1.14**	**4.69**
	IV	**3.86**	**1.83**	**8.14**
Surgery_Treatment	Yes			
	No	**1.52**	**1.04**	**2.22**
Chemotherapy_Treatment	Yes	1		
	No	0.59	0.37	1.04
Radiotherapy_Treatment	Yes	1		
	No	0.87	0.60	1.28

Disease-free survival (DFS)

In this study, metastasis was observed in 88 (45.6%) out of 193 patients who underwent surgery (distant metastasis in 38 patients, local recurrence in 34 patients, and both local and distant recurrence in 16 patients). The most common site of metastasis was liver in 59.1% of metastatic cancer patients. Other sites of metastasis in patients and over survival of them are presented in Table 4. The mean (SD) age of the operated patients was 61.5 (14.1) years. Table 5 shows the characteristics of operated patients.

Table 4. Metastatic sites of patients with gastric cancer.

Sites of Metastases	Frequency	Percent	OS (month)
Liver	52	59.1	15
Lung	9	10.2	17
Bone	4	4.5	23
Intestine	5	5.7	24
Bladder	1	1.1	-
Liver and Intestine	5	5.7	14
Liver and Lung	4	4.6	15
Liver, Lung and Intestine	2	2.3	-
Bladder and Intestine	5	5.7	13
Lung and Intestine	1	1.1	-
Total	88	100	22

Figure 5 shows the probability of DFS in operated GC patients. The mean and median DFS values were 31 and 37.5 months, respectively, and the median DFS values in men and women were 30 and 48 months, respectively. DFS rates at 1, 3, and 5 years were 74.5%, 45.5%, and 13.5%, respectively. The important risk factors for DFS were determined using the RSF method. Table 6 shows the goodness of fit indices of the model. According to the IBS index and the prediction error rate, the RSF model with the log-rank splitting method was selected as the appropriate model. Figure 6 illustrates the error rate and variable importance of the RSF model with the log-rank splitting rule to identify the important variables in the prediction of the DFS. As shown in Figure 6, disease stage, tumor grade, radiotherapy, tumor site, surgery, and age at the diagnosis are the major DFS-affecting variables. Figure 7 displays the prediction error values for CPH and RSF models with different splitting rules. The 5-year survival rates for the important variables identified using the RSF model are presented in Figure 8. The adjusted 5-year DFS probabilities in the presence of other variables reveal that the predicted 5-year survival rates decrease with increasing age. Also, the increased grade and stage of the disease reduce the 5-year DFS probabilities. The 5-year DFS probabilities for the levels of the other identified important variables are shown in Figure 8. Table 7 presents the results obtained for the effects of factors on DFS determined using the CPH model. The results of the CPH model revealed that the DFS was significantly affected by the age at the diagnosis, chemotherapy, tumor stage, and tumor site. The risk of disease recurrence increases with increasing the age of the diagnosis and the disease stage. The risk of recurrence decreased in patients who received chemotherapy, also patients with a tumor site in the upper part of the stomach showed an increased risk of recurrence. The proportional hazard (PH) assumption for OS wasn't satisfied (p<0.001) while the PH assumption for DFS was held.

DISCUSSION

In this study, the important factors affecting OS and DFS in GC patients were determined using the RSF and COX models. The results indicated that the RSF model performed better than the CPH model in determining the important variables, and the RSF model is advantageous as it does not require limited assumptions. The major OS predictors in the RSF method were determined according to tumor grade, disease stage, age at the diagnosis, local recurrence, surgery, and distant metastasis. Disease stage, tumor grade, radiotherapy, tumor site, surgery, and age at the diagnosis were the main predictors of DFS.

Table 5. Demographic and clinical characteristics of gastric cancer patients.

Variable	Subgroup	N (%)	Median DFS	*p*-value
Sex	Female	47 (24.4)	30	0.46
	Male	146 (75.6)	48	
Age	≤ 55	56 (29)	31	0.06
	56-70	74 (39.4)	48	
	> 70	61 (31.6)	29	
Number of chemo course	1-5	75 (46)	23	0.26
	6-10	64 (39.3)	36	
	>11	24 (14.7)	25	
Tumor grade	Well	36 (18.7)	48	0.001
	Moderate	33 (17.1)	36	
	Poor	28 (14.5)	10	
	Unknown	96 (49.7)	36	
Stage	II	26 (13.5)	51	<0.001
	III	33 (17.1)	47	
	IV	61 (31.6)	16	
	Unknown	73 (37.8)	48	
Radiotherapy	No	32 (16.6)	--	0.08
	Yes	161 (83.4)	31	
Chemotherapy	No	30 (15.5)	--	0.02
	Yes	163 (84.5)	30	
Site of Tumor	Antrum	51 (26.4)	47	0.27
	Body	11 (5.7)	17	
	Cardia	100 (51.8)	29	
	Fundus	14 (7.3)	27	
	Unknown	17 (8.8)	40	
Distance metastasis alone	No	155 (80.3)	51	< 0.001
	Yes	38 (19.7)	12	
Local recurrence alone	No	159 (82.4)	56	< 0.001
	Yes	34 (17.6)	19	
Smoking	No	109 (56.5)	36	0.15
	Yes	84 (43.5)	31	
Family history of cancer	Yes	26 (13.5)	27	0.34
	No	167 (86.5)	36	

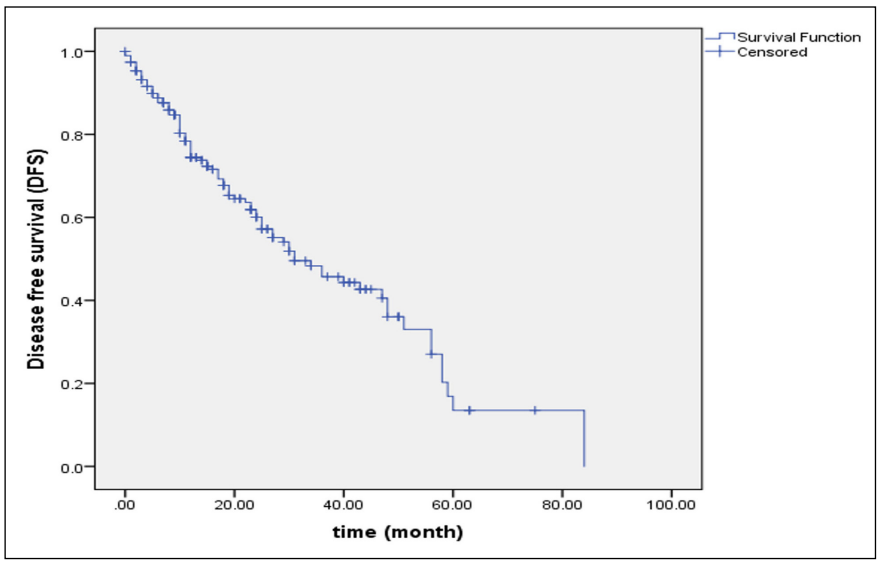

Figure 5. DFS of gastric cancer patients undergoing surgery

Table 6. Goodness of fit indices of the models.

Model	Error rate	IBS[0,time=63]
Cox	22.7	0.242
RSF(log.rank.score)	21.9	0.219
RSF(random)	22.6	0.242
RSF(log.rank)	**21.8**	**0.212**

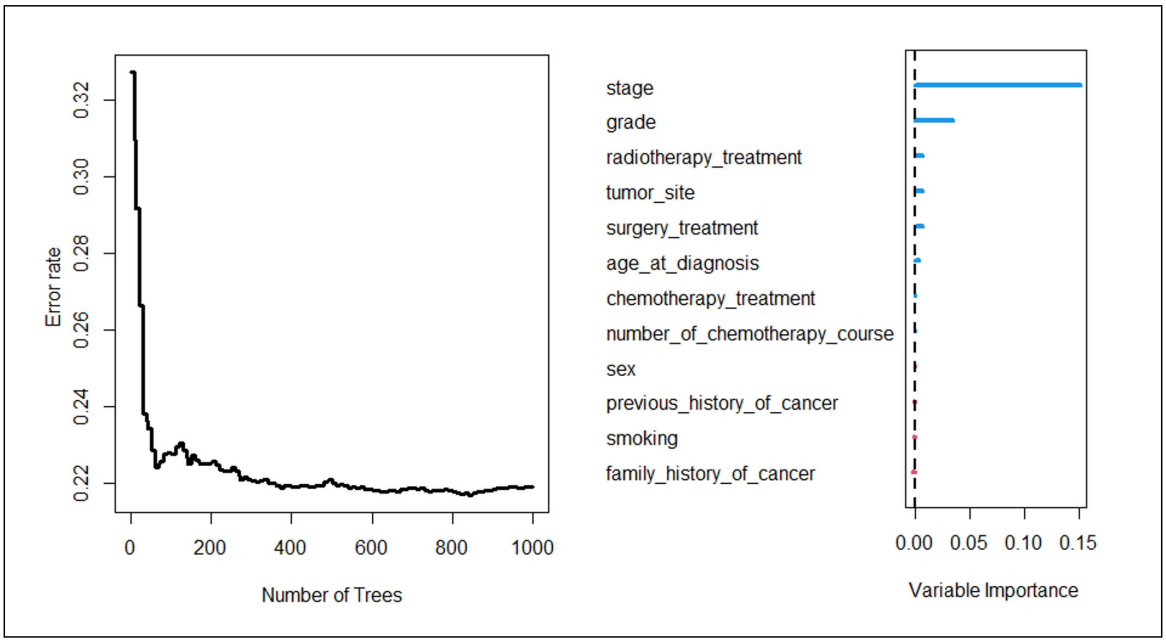

Figure 6. Out-of-Bag variable importance and Error rate of RSF for Log-Rank Splitting Rule.

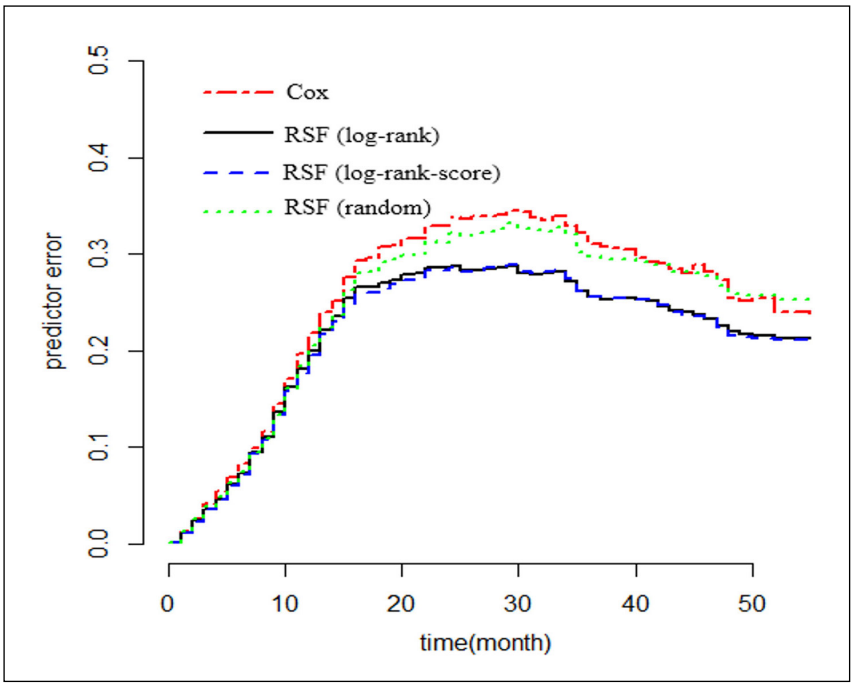

Figure 7. Prediction Error Curves for Cox model (Red), RSF with log-rank splitting rule (Black), RSF with random splitting rule (Green), and RSF with log-rank score splitting rule (Blue).

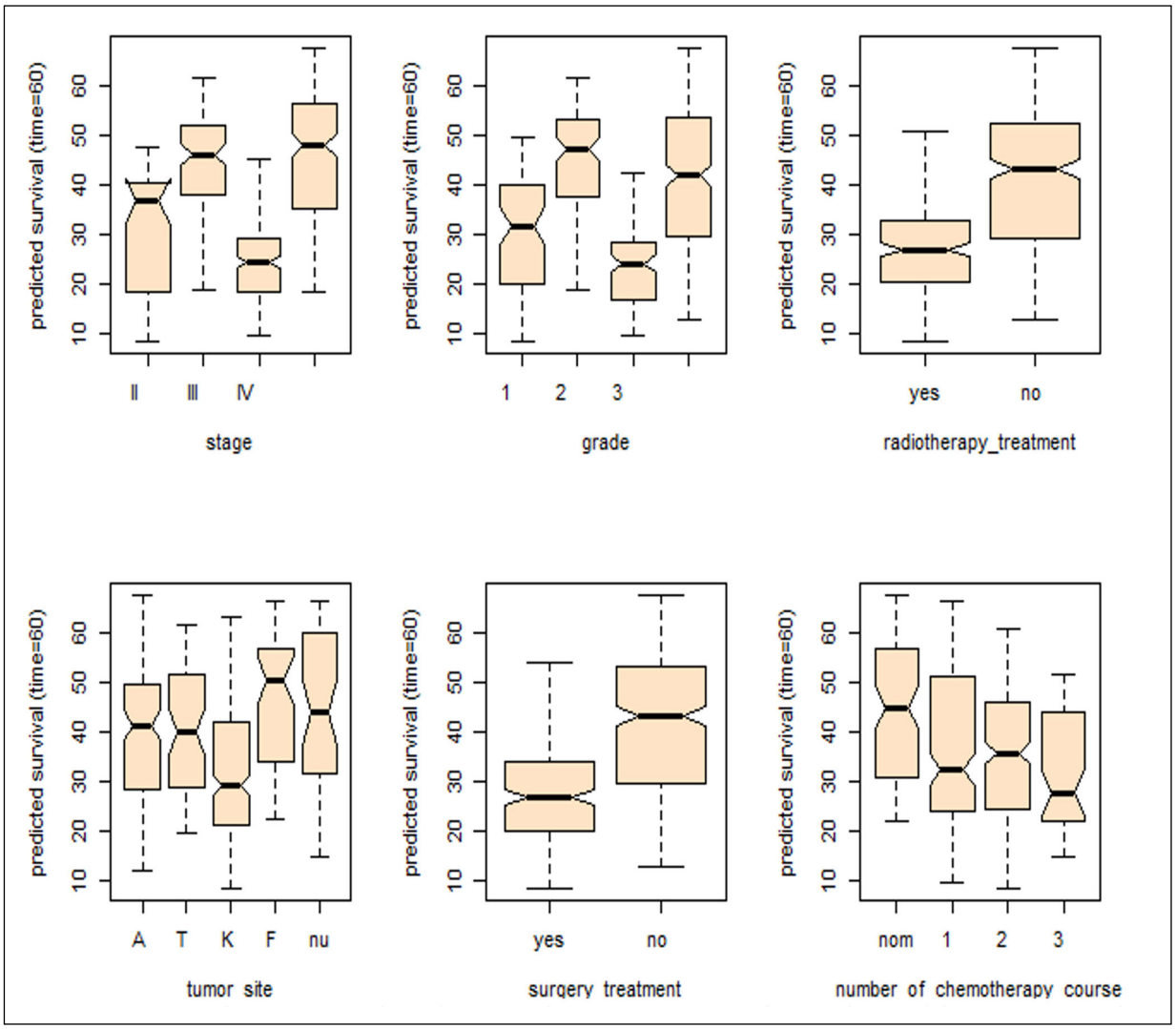

Figure 8. Partial 5-year predicted survival for six most influential variables on survival in colorectal cancer data. Values on the vertical axis represent the predicted survival probability for a given predictor, after adjusting for all other predictors.

Also in the present study, the result of the CPH model showed that OS was influenced by the variables of age at the diagnosis, local recurrence, disease stage, and surgery. Furthermore, age at the diagnosis, tumor site, chemotherapy, and disease stage were among the main predictors of DFS.

In a study on GC patients by Toyokawa et al[21], age at the diagnosis and chemotherapy were factors affecting OS and DFS in patients who were in stage I of the disease, and tumor size and chemotherapy were factors affecting OS and DFS in the patients at stage II of the disease[21].

Yaprak et al[22] reported median OS and DFS times of 51 and 35 months, respectively, in GC patients without metastasis who were in stages 1-3 of the disease. In their study, tumor grade and the disease stage significantly affected survival rates, and survival probabilities of 85%, 55%, and 45% were respectively obtained for one, three, and five years, and the DFS probabilities for one, three, and five years were 72%, 49%, and 38%, respectively[22].

In a study on GC patients with metastases, Safari et al[23] identified the type of surgery, metastasis site, chemotherapy, age, tumor grade, and surgery, the number of involved lymphomas, gender, and radiotherapy as the major OS-affecting variables. Adham et al[24] introduced age, tumor size, and metastasis as the key OS-affecting variables based on the RSF model.

In a study on GC patients after CR, Zhu et al[25] estimated survival rates of 92.5%, 65.3%, and 46.8% for one, three, and five years, respectively. In their study, OS was influenced by the variables of age, disease stage, and tumor site, and DFS was significantly affected by the disease stage[25]. Itaimi et al[26] estimated a three-year OS rate of 58% in GC patients. In their study, OS was significantly affected by local recurrence

Table 7. Multivariable Cox Regression of Prognostic Factors on DFS.

Variable	level	HR	95% CI for HR	
			lower	upper
Age_At_Diagnosis		**1.03**	**1.02**	**1.04**
Grade	Well	1		
	Moderate	0.77	0.32	1.86
	Poor	1.02	0.51	2.07
Family_History_of_Cancer	No	1		
	Yes	0.86	0.57	1.31
Smoking	No			
	Yes	0.75	0.55	1.01
Number of Chemo Course	1-5	1		
	6-10	1.55	0.90	2.66
	>11	1.07	0.63	1.80
Tumor_Site	Antrum	1		
	Body	**2.25**	**1.28**	**3.93**
	Cardia	1.25	0.84	1.87
	fundus	0.76	0.38	1.51
Previous_History_of_Cancer	Yes	1		
	No	1.55	0.76	3.18
Sex female	Male	1		
	Female	0.77	0.53	1.12
Stage	II	1		
	III	**1.65**	**0.63**	**4.30**
	IV	**6.86**	**3.13**	**15.05**
Surgery_Treatment	Yes	1		
	No	0.64	0.34	1.22
Chemotherapy_Treatment	Yes	1		
	No	**0.40**	**0.19**	**0.86**
Radiotherapy_Treatment	Yes	1		
	No	1.42	0.74	2.70

and tumor stage, and the number of involved lymph nodes was one of the factors affecting DFS. Han et al[27] conducted a study on men with GC in stages 3-4 of the disease and observed that smoking was the only factor affecting OS and DFS. In most of the reviewed studies, age at the diagnosis, tumor grade, and stage, surgery, and chemotherapy were among the variables affecting OS, while DFS was influenced by age, disease stage, and tumor site[22-27]. The difference between the previous studies in determining the influential variables can be attributed to various characteristics of examined patients and variables in such studies. In particular, the effect of some genes alongside demographic and clinical characteristics was investigated in some studies. In most studies based on the CPH model, the non-significance of this variable can be the reason for the collinearity and correlation between the variables. As such, variables (e.g., the age of diagnosis), which are associated with the variables of disease stage and tumor size, may not be recognized as significant variables. This seems reasonable in the CPH model, but all three variables are identified as important in the RSF model regardless of the correlation between the variables.

In the present study, the RSF model performed better than the CPH model for identifying the variables affecting DFS. However, the performance of CPH and RSF models was almost the same in evaluating the variables affecting OS, although this has not been confirmed in some studies on survival[23-25]. In this study, the RSF model with the log-rank score division rule had the best performance in determining the key variables affecting OS, and the coordination indices of this model and the CPH model were respectively obtained at 73.1% and 72.9% in this study. With a coordination index of 70.3%, this model was also selected as an appropriate model by Adham et al[24]. Ingrisch et al[28] obtained coordination indices of 65.7% and

65.2% for the RSF model with the log-rank division rule and the CPH model, respectively[28]. In other studies conducted on other types of diseases, the RSF model can better identify influential variables in these conditions unlike the CPH model because survival-affecting variables may have collinearity or complex relationships in such situations. Similarly, the better performance of the RSF model than the CPH model in survival prediction was confirmed in studies on patients with cardiac arrhythmia by Miao et al[29], kidney transplant patients by Roshanaei et al[30], colorectal cancer by Myte et al[31], acute liver failure by Zhang et al[32], time to recurrence in ovarian cancer patients by Deldar et al[33], and head and neck patients by Datma et al[34]. Therefore, the RSF model works at least the same as the CPH model in identifying survival-affecting variables without limiting assumptions. Thus, the results of this model in identifying important variables influencing survival can help doctors in diagnostic and preventive assessments.

In the present study, the analysis of three division methods in the RSF model revealed better performance of log-rank score and log-rank division rules, which corresponds to most previous studies[23,29,30,34,35]. Regarding the benefits of this study, the main strength might be that GC patients were monitored in the long-term, and this enabled us to assess affected risk factor on interesting outcomes precisely. Moreover, the applied method is strongly recommended when the predictor variables are correlated or there is a nonlinear relationship among the independent variables. Finally, the suggested method requires no limiting assumptions for the analysis, which is another compelling benefit to use RSF rather than the conventional methods of survival analysis.

The first strength of the current study is the long-term monitoring of GC patients. Secondly, this method determines the effect of variables influencing the response prediction in order of importance. This method also works well if the variables of interest are correlated or there is a nonlinear structure and even interactions between the variables. Moreover, it requires no limiting assumptions for the analysis unlike the conventional methods of survival analysis.

As for limitation of the research, it is important to keep in mind that the study was conducted with a single-center data. It is obvious that the results will be more accurate if more samples in multiple centers and more auxiliary variables would be available. Lastly, this is a retrospective study in which some data were not fully recorded for some cases.

CONCLUSIONS

RSF complements the Cox model by providing the relative importance of model covariates, though the Cox model gives a clinically understandable result on effect of each covariate on survival. Compared with Cox models, the RSF model can effectively predict the survival of patients with better performance.

Ethics approval:
The study was approved by the Ethics Committee of the Hamadan University of Medical Sciences (Ethics Committee approval code IR.UMSHA.REC.1397.103; Project number 97030110).

Informed consent:
In this study, informed consent was not necessary because of the use of anonymized patients' records.

Availability of data and materials:
Data are available on reasonable request from the corresponding author.

Conflict of Interests:
The authors have no conflicts of interest.

Funding:
No specific funding was obtained for this study.

Author contributions:
MS, MR, and GH designed the study. MR and BG collected the data. MS, GR, and JF analyzed and interpreted the data. MS and MRA drafted the manuscript. GR, JF and MS provided administrative, technical, or material support. All authors contributed to the article and approved the submitted version.

ORCID ID:
Malihe Safari: https://orcid.org/0000-0002-8068-5921
Mohamadreza Anari: https://orcid.org/0009-0001-8930-9551
Ghodratollah Roshanaei: https://orcid.org/0000-0002-3547-9125
Mastore Rahimi: https://orcid.org/0000-0002-9742-7293
Bayazid Ghaderi: https://orcid.org/0000-0001-9174-829X
Javad Faradmal: https://orcid.org/0000-0001-5514-3584

REFERENCES

1. Prashanth R, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. Przeglad Gastroenterol 2019; 14: 26-38.
2. Toni G, Panarese I, Di Francia R, Franco R. Molecular Classification of Gastric Cancer. WCRJ 2020; 7: e1472.
3. Mazidimoradi A, Baghernezhad Hesary F, Gerayllo S, Banakar N, Allahqoli L, Salehiniya H. Global distribution of incidence, mortality, and burden of stomach cancers and its relationship with the sociodemographic index. WCRJ 2023; 10: e2519.
4. Yoshida N, Doyama H, Yano T, Horimatsu T, Uedo N, Yamamoto Y, Kakushima N, Kanzaki H, Hori S, Yao K, Oda I, Katada C, Yokoi C, Ohata K, Yoshimura K, Ishikawa H, Muto M. Early gastric cancer detection in high-risk patients: a multicentre randomised controlled trial on the effect of second-generation narrow band imaging. Gut 2021; 70: 67-75.
5. Wu J, Wu XD, Gao Y, Gao Y. Correlation between preoperative systemic immune-inflammatory indexes and the prognosis of gastric cancer patients. Eur Rev Med Pharmacol Sci 2023; 27: 5706-5720.
6. Katai H, Ishikawa T, Akazawa K, Isobe Y, Miyashiro I, Oda I, Tsujitani S, Ono H, Tanabe S, Fukagawa T, Nunobe S, Kakeji Y, Nashimoto A; Registration Committee of the Japanese Gastric Cancer Association. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001-2007). Gastric Cancer 2018; 21: 144-154.
7. .7Cuzzuol BR, Vieira ES, Araújo GRL, Apolonio JS, de Carvalho LS, da Silva Junior RT, Bittencourt de Brito B, Freire de Melo F. Gastric Cancer: A Brief Review, from Risk Factors to Treatment. Arch Gastroenterol Res 2020; 1: 34-39.
8. .8 Markar SR, Karthikesalingam A, Jackson D, Hanna GB. Long-term survival after gastrectomy for cancer in randomized, controlled oncological trials: comparison between West and East. Ann Surg Oncol 2013; 20: 2328-2338.
9. .9Lai JF, Xu WN, Noh SH, Lu WQ. Effect of World Health Organization (WHO) Histological Classification on Predicting Lymph Node Metastasis and Recurrence in Early Gastric Cancer. Med Sci Monit 2016; 22: 3147-3153.
10. .10 Spolverato G, Ejaz A, Kim Y, Squires MH, Poultsides GA, Fields RC, Schmidt C, Weber SM, Votanopoulos K, Maithel SK, Pawlik TM. Rates and patterns of recurrence after curative internt resection for gastric cancer: a United States multi-institutional analysis. J Am Coll Surg 2014; 219: 664-75.
11. BY Zhu, SQ Yuan, RC Nie, SM Li, LR Yang, JL Duan, YB Chen, XS Zhang, Prognostic Factors and Recurrence Patterns in T4 Gastric Cancer Patients after Curative Resection. J Cancer 2019; 10: 1181-1188.
12. Itaimi A, Baraket O, Triki W, Ayed K, Bouchoucha S. Prognostic factors affecting survival and recurrence in gastric carcinoma. Cancer Rep Rev 2018; 2: 1-4.
13. Kleinbaum D, Klein M. Survival Analysis: A Self-Learning Text. Third ed. New York: Springer; 2012.
14. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High Dimensional Variable Selection for Survival Data. J Am Stat Ass 2010; 105: 205-17.
15. Breiman L. Random forests. Machine Learning 2001; 45: 5-32.
16. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The Annals of Applied Statistics 2008; 2: 841–60.
17. Wang H, Shen L, Geng J, Wu Y, Xiao H, Zhang F, Si H. Prognostic value of cancer antigen -125 for lung adenocarcinoma patients with brain metastasis: A random survival forest prognostic model. Sci Rep 2018; 8: 5670.
18. Nasejje JB, Mwambi H. Application of random survival forests in understanding the determinants of under-five child mortality in Uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. BMC Res Notes 2017; 10: 459.
19. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biometr J 2006; 48: 1029–1040.
20. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statis Med 2004; 23: 2109–2123.
21. Toyokawa T, Ohira M, Sakurai K, Kubo N, Tanaka H, Muguruma K, Hirakawa K. The Role of Adjuvant Chemotherapy for Patients with Stage IB Gastric Cancer. Anticancer Res 2015; 35: 4091-4097.
22. Yaprak G, Tataroglu D, Dogan B, Pekyurek M. Prognostic factors for survival in patients with gastric cancer: Singlecentre experience. North Clin Istanb 2020; 7: 146–152.
23. Safari M, Abbasi M, Gohari Ensaf F, Berangi Z, Roshanaei G. Identification of Factors Affecting Metastatic Gastric Cancer Patients' Survival Using the Random Survival Forest and Comparison with Cox Regression Model. irje 2020; 15: 343-351.
24. Adham D, Abbasgholizadeh N, Abazari M. Prognostic Factors for Survival in Patients with Gastric Cancer using a Random Survival Forest. Asian Pac J Cancer Prev 2017; 18(1):129-134.
25. Zhu BY, Yuan SQ, Nie RC, Li SM, Yang LR, Duan JL, Chen YB, Zhang XS. Prognostic Factors and Recurrence Patterns in T4 Gastric Cancer Patients after Curative Resection. J Cancer 2019; 10: 1181-1188.
26. Itaimi A, Baraket O, Triki W, Ayed K, Bouchoucha S, Prognostic factors affecting survival and recurrence in gastric carcinoma carcinoma. Cancer Rep Rev 2018; 2: 1-4.
27. Han MA, Kim YW, Choi IJ, Oh MG, Kim CG, Lee JY, Cho SJ, Eom BW, Yoon HM, Ryu KW. Association of smoking history with cancer recurrence and survival in stage III-IV male gastric cancer patients. Cancer Epidemiol Biomarkers Prev 2013; 22:1805-1812.

28. Ingrisch M, Schöppe F, Paprottka K, Fabritius M, Strobl FF, De Toni EN, Ilhan H, Todica A, Michl M, Paprottka PM.: "Prediction of 90Y Radioembolization Outcome from Pretherapeutic Factors with Random Survival Forests". Journal of Nuclear Medicine 2018; 59: 769-773.
29. Miao F, Cai YP, Zhang YX, Li Y, Zhang YT. Risk prediction of one-year mortality in patients with cardiac arrhythmias using random survival forest. Comput Math Methods Med 2015; 2015: 303250.
30. Roshanaei G, Omidi T, Faradmal J, Safari M, Poorolajal J. Determining affected factors on survival of kidney transplant in living donor patients using a random survival forest. Koomesh 1397; 20: 517-523.
31. Myte R. Covariate selection for colorectal cancer survival data: A Comparison case study between random survival forests and the cox proportional-hazards model. Umeå: Umeå University; 2013.
32. Zhang, Zhi-Qiao; He, Gang; Luo, Zhao-Wen; Cheng, Can-Chang; Wang, Peng1; Li, Jing1; Zhu, Ming-Gu; Ming, Lang1; He, Ting-Shan1; Ouyang, Yan-Ling1; Huang, Yi-Yan1; Wu, Xing-Liu; Ye, Yi-Nong. Individual mortality risk predictive system of patients with acute-on-chronic liver failure based on a random survival forest model, Chinese Medical Journal 2021; 134: 1701-1708.
33. Deldar M, Anbiaee R, Sayehmiri K. Predicting Epithelial Ovarian Cancer first recurrence with Random Survival Forest: Comparison Parametric, Semi-Parametric, and Random Survival Forest Methods. JBE 2021; 6: 267-274.
34. Datema FR, Moya A, Krause P, Bäck T, Willmes L, Langeveld T, Baatenburg de Jong RJ, Blom HM. Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. Head Neck 2012; 34: 50-8.
35. Roshanaei G, Safari M, Faradmal J, Abbasi M, Khazaei S. Factors affecting the survival of patients with colorectal cancer using random survival forest. J Gastrointest Cancer 2022; 53: 64-71.