

Predictive model for survival in patients with gastric cancer

Ladan Goshayeshi^{1,2}, Benyamin Hoseini³, Zahra Yousefli⁴, Alireza Khooei⁵, Kobra Etminani⁶, Abbas Esmailzadeh^{1,7}, Amin Golabpour^{6,8}

¹ Department of Gastroenterology and Hepatology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

² Surgical Oncology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

³ Department of Health Informatics Technology, Neyshabur University of Medical Sciences, Neyshabur, Iran

⁴ Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁵ Department of Pathology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁶ Department of Biomedical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁷ Gastroenterology and Hepatology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

⁸ School of Medicine, Shahrood University of Medical Sciences, Shahrood, Iran

Type of article: Original

Abstract

Background and aim: Gastric cancer is one of the most prevalent cancers in the world. Characterized by poor prognosis, it is a frequent cause of cancer in Iran. The aim of the study was to design a predictive model of survival time for patients suffering from gastric cancer.

Methods: This was a historical cohort conducted between 2011 and 2016. Study population were 277 patients suffering from gastric cancer. Data were gathered from the Iranian Cancer Registry and the laboratory of Emam Reza Hospital in Mashhad, Iran. Patients or their relatives underwent interviews where it was needed. Missing values were imputed by data mining techniques. Fifteen factors were analyzed. Survival was addressed as a dependent variable. Then, the predictive model was designed by combining both genetic algorithm and logistic regression. Matlab 2014 software was used to combine them.

Results: Of the 277 patients, only survival of 80 patients was available whose data were used for designing the predictive model. Mean \pm SD of missing values for each patient was 4.43 ± 1.41 combined predictive model achieved 72.57% accuracy. Sex, birth year, age at diagnosis time, age at diagnosis time of patients' family, family history of gastric cancer, and family history of other gastrointestinal cancers were six parameters associated with patient survival.

Conclusion: The study revealed that imputing missing values by data mining techniques have a good accuracy. And it also revealed six parameters extracted by genetic algorithm effect on the survival of patients with gastric cancer. Our combined predictive model, with a good accuracy, is appropriate to forecast the survival of patients suffering from Gastric cancer. So, we suggest policy makers and specialists to apply it for prediction of patients' survival.

Keywords: Gastric Cancer, Survival, Missing Value, Predictive Model, Mashhad

1. Introduction

Gastric cancer is known as one of the common causes of cancer mortality worldwide (1), as of the 74,067 new cases of cancer in Iran in 2009, 6,886 were detected as having gastric cancer (2). The five most common cancers (excluding skin cancer) are stomach, esophagus, colon-rectum, bladder and leukemia in males, and in females are breast, esophagus, stomach, colon-rectum and cervix uteri (3). Despite a decline that has occurred in incidences of gastric

Corresponding author:

Amin Golabpour, Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. Tel: +989166537914, Fax: +989166537914, Email: GolabpourA911@mums.ac.ir

Received: April 26, 2017, Accepted: July 20, 2017, Published: December 2017

iThenticate screening: June 23, 2017, English editing: August 05, 2017, Quality control: November 12, 2017

This article has been reviewed / commented by three experts

© 2017 The Authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

cancer in Iran during recent years, according to results of the present study, Iran is placed in the group of countries with a moderate risk of gastric cancer (4). High level clustering of the cases was seen in northern, northwestern, western, and northeastern areas for esophagus, gastric, and colorectal cancers (5). According to the latest reports of the Iranian Ministry of Health and Medical Education, the five leading gastrointestinal causes of death in order of frequency were: gastric cancer, hepatobiliary cancer, liver cirrhosis, esophageal cancer, and colorectal cancer. Gastric adenocarcinoma has been introduced as the most fatal cancer in Iran (6, 7). Patients are often diagnosed with advanced disease and five-year survival rates are poor, usually less than 30% (8). A few studies have already investigated the survival pattern of stomach cancer in different regions of Iran and prognostic factors are being ascertained (9-12). The likelihood of death was higher in men and likelihood of tumors was higher in corpus and cardia patients more than 70 years old (13). In Yazd, 295 gastric cancer patients had a 5-year survival rate of 8% (14). In another study from the national cancer registry file, where 3,439 cases of stomach cancer between 2001 and 2005 were studied, age, tumor size and advance pathologic stage was related to survival. But sex, distant metastasis, histology type, tumor grade and lymph node metastasis were not prognostic factors (15). The statistics of gastric cancer should be updated periodically to identify trends in incidence, prevalence, and mortality, all of which have important implications for health policy planning (16). To the best of our knowledge, despite high prevalence of gastric cancer in the northeastern area, there is no study on survival of gastric cancer and prognostic factors in this area. So, to investigate the survival and the factors influencing the survival time of gastric cancer patients, which are the main objectives of this study (14), we aimed to design a model for predicting the survival rate of patients with stomach cancer regarding to histopathology type, anatomical site, and family history between 2010 and 2015 in Mashhad, Iran.

2. Material and Methods

The data of this research was from the type of survival studies. This data were collected through a historical cohort conducted between 2011 and 2016. Study population were 277 patients suffering from gastric cancer. The available pathological sheets of these patients have been collected from the Iranian Cancer Registry from 2011 to 2014 as well as the laboratory of Imam Reza Hospital from 2015 to 2016 in Mashhad, Iran. Patients' medical records have been assessed and in cases of their satisfaction, the patients have been interviewed.

Table 1. The features of the independent variables of gastric cancer

No	Variable name	Scale	Description
1	Sex	Nominal	Male: 61, Female:19
2	Birth year	Interval	The year patients were burned. Range between 1926 to 1967
3	Education	Ordinal	Included the following 2 values: 1) illiterate, 2) under diploma
4	Race	Ordinal	Included the following 3 values: 1) Persian, 2) Turkish, 3) Kurdish
5	PMH	Ordinal	Included the following 6 values: 1) hypertension (HTN), 2) coronary artery disease (CAD), 3) diabetes mellitus (DM), 4) DM +HTN, 5) DM+HTN+CAD, 6) HTN+CAD
6	Age at Diagnosis	Interval	Ranged between 46 to 87
7	Family History of Gastric Cancer	Ordinal	Family history of gastric cancer: 1) first degree relative (FDR), 2) second degree relatives (SDR)
8	Age at dx of Family GC	Interval	Family 's age at the diagnosis time included value between 45 to 82
9	History of other GI cancer	Ordinal	History of other GI cancer, 1) First Degree Relative, 2) Second Degree Relatives
10	Type of other GI Cancer	Ordinal	Included following 4 values: 1) small intestine, 2) liver, 3) esophagus, 4) large intestine
11	Hx of extra GI Cancer	Ordinal	Familial history of extragastric cancer. Included following 2 values: 1) First Degree Relative, 2) Second Degree Relatives
12	Treatment	Ordinal	Included following 3 values: 1) Surgery, 2) Surgery + Chemo + radio, 3) Chemo
13	Cause of Death	Ordinal	Included following 3 values: 1) Cancer, 2) MI, 3) PTE
14	Pathology	Ordinal	Included following 7 values: 1) adenocarcinoma, 2) inflammatory tumor, 3) mucinous adenocarcinoma, 4) neuroendocrine carcinoma, 5) signet ring cell carcinoma, 6) GIST tumor, 7) undifferentiated carcinoma
15	Addiction	Nominal	17 patients were addicted and 63 were not.
16	Survival	Nominal	The survival for 33 patients was 1 year and for 47 patients was 2 years

In cases of deceased patients, their relatives were interviewed if they were satisfied to do so. Survival prediction modeling for the patients requires the information of the deceased patients, especially their chances of survival after their cancer diagnosis. Among 277 patients, 197 were either alive at the time of this research or their survival was not clear. Therefore, the information of 80 patients was applied for the purpose of modeling. This information included 15 independent variables and one dependent variable among which 3, 10 and 3 variables were nominal, ordinal and interval respectively. The details of these variables have been described in Table 1.

2.1. Data preprocessing

Of the 277 patients, only survival of 80 patients was available whose data were used for conducting a predictive model. We removed 118 patients because they were alive when research was in progress, removed 10 patients due to having outlier data and removed 10 patients due to unavailability of survival data.

The steps of preprocessing have been depicted in Figure 1.

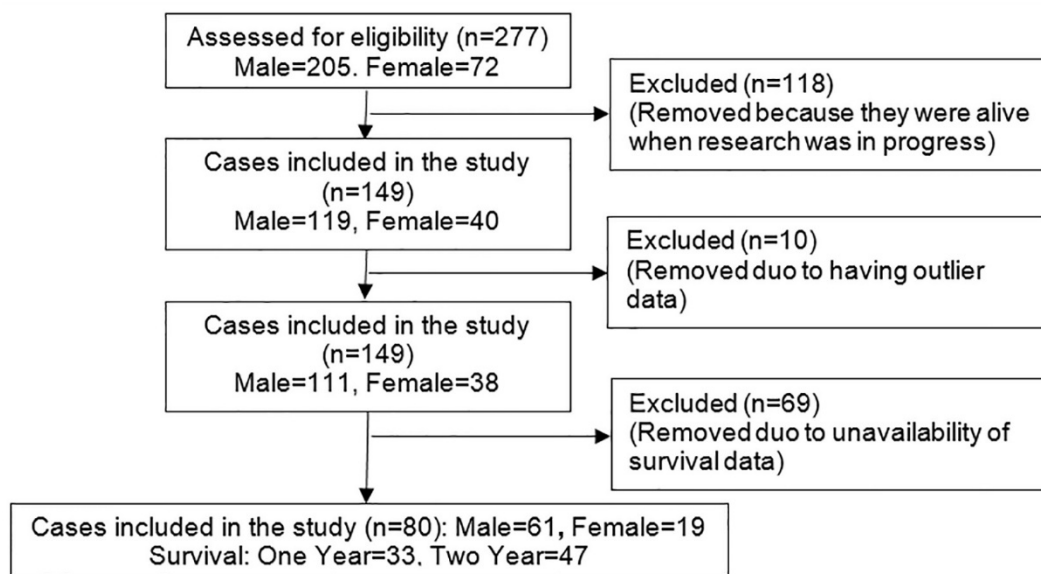


Figure 1. The steps of preprocessing for the recommended model

2.1.1. The missing values

There were 354 fields (29.5%) from 80 records of the patients applied for modeling, which reveals the high amount of missing values in the study. The number and percentage of the missing values for each independent variable have been given in Table 2; from 15 independent variables, 8 variables were missing. The variable "Hx. of extra. GI. Cancer" had the highest percentage of missing data with 88.75%. The mean and standard deviation of the missing variables for each patient were 4.43 ± 1.41 . MICE algorithm (6), with some modifications which will be mentioned later, was used to impute missing values. The form of the algorithm to find the missing values is in a way in which at first, the variables must be prioritized based on their missing values. The variables with the lowest missing values come first and the ones with the highest come thereafter. Using data mining technique, a data classification method was generated in which the independent variables were variables without missing data, and dependent variables were the ones with the least number of missing variables. Then, with the use of data mining algorithm, the missing values of the dependent variable were imputed and the dependent variables were added to the previous collection of the independent variables. The next variable with the least number of missing values was chosen as a dependent variable and again, the previous process was repeated for it. This process continued until no variable with missing value existed (17). The above process was repeated several times until no change occurred in the data. By the end of this process, all the empty fields were imputed and no missing value existed.

2.1.2. Recognition of the out-of-range data

In order to cross out the out-of-range data for each feature, algorithm IQR was used (7). The number of the records with the out-of-range data is 10, which all have been crossed out.

Table 2. The percentage of missing values in independent variables

Variable	Missing		Valid (n)
	n	Percent	
1 Hx of extra GI Cancer	71	88.75%	9
2 Type of other GI Cancer	64	80%	16
3 Hx of other GI Cancer	64	80%	16
4 Age at Dx of Family GC	58	72.5%	22
5 FH of gastric cancer	57	71.25%	23
6 PMH	35	43.75%	45
7 Age at Diagnosis	4	5%	76
8 Birth Year	1	1.25%	79

3. The recommended model

In this research, Logistic Regression Algorithm, Genetics Algorithm and MICE Algorithm were applied. Logistic regression is a predictive analysis which is used to describe the data and to explain the relationship between one dependent binary variable and one or more independent, variable, ordinal, interval, or ratio-level independent variables. A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions. Multivariate imputation by chained equations (MICE) has emerged as a principled method of dealing with missing data. Despite properties that make MICE particularly useful for large imputation procedures, advances in data mining now make it accessible to many researchers. The issue was dealt with in three forms. In the first form, logistic regression was used for all 15 independent variables. In the second form, the missing values were imputed by MICE algorithm. Then, the dimensions were reduced by genetic algorithm. In the third form, the missing values were imputed by the optimized algorithm and then the dimensions were reduced by genetic algorithm.

3.1. Solve problem by regression algorithm

In this stage, all data with their features has been considered. The data was divided in two sections: test and train. The division technique will be described in the assessment section. The regression algorithm was applied for train data and a model was designed. To assess the recommended model, test data was not used (model assessment technique is described in the assessment section).

3.2. Solve the problem by MICE algorithm and genetic algorithm

In this technique, at first, the missing values were imputed and then the dimensions of the data were reduced by MICE algorithm and genetic algorithm respectively. After the dimensions' reduction, the data was divided into two sections of train and test. Like the previous stage, logistic regression algorithm was applied for train data and then the model was assessed by test data. In Figure 2, the way to reduce dimensions by genetic algorithm and application of the regression algorithm has been demonstrated. First, the missing values were imputed by MICE algorithm and then, all 80 records were applied for genetic algorithm. Genetic algorithm chose a subset of records which has been assessed by regression algorithm. This process was repeated until the best subset was chosen as output, and effective parameters for survival rate were selected. Then data has been divided into two sections of test and train, regression algorithm was applied for the train data and a model was designed. The designed model was assessed by test data.

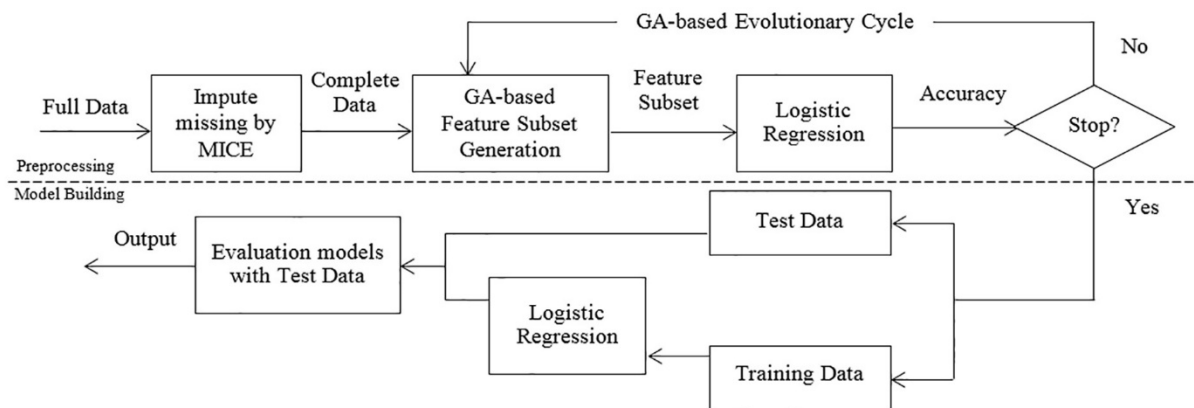


Figure 2. Model design by logistic algorithm and genetic algorithm

3.3. Solve the problem by MICE optimized algorithm and genetic algorithm

This model looks like section 4-2 and the only difference is that MICE optimized algorithm has been applied instead of MICE algorithm. The process of this optimization on MICE algorithm has been described in section 2-1-1. The solution has been given in Figure 3.

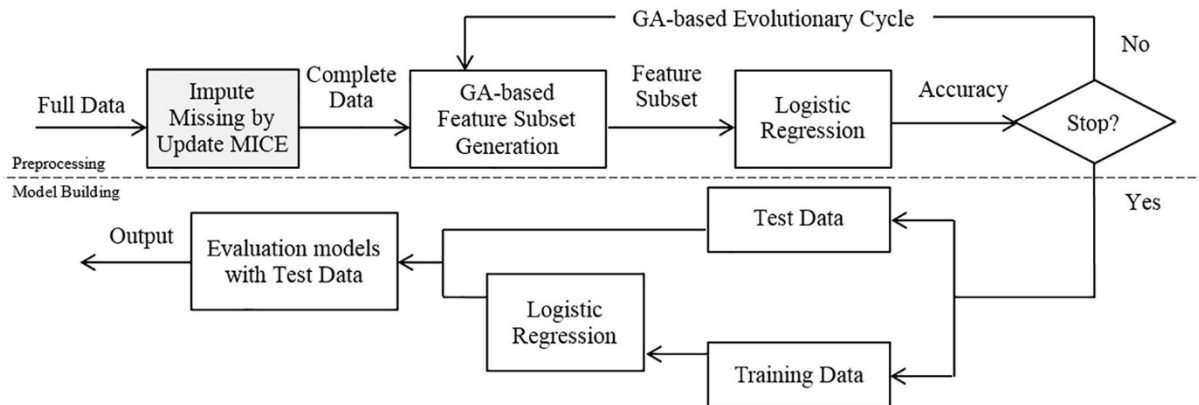


Figure 3. Model design by MICE optimized algorithm, genetic algorithm and logistic algorithm

4. Results

To assess the algorithms, three features of Sensitivity, Specificity and Accuracy have been used. In order to conduct the research, data has been divided into two test and train sections. 10-fold technique was used to divide data. Then, regression algorithm was run 10,000 times and algorithm accuracy was measured 1,000 times and its mean was chosen as the algorithm's accuracy. In the model assessment, the problem was solved in three situations. In the first one, regression algorithm was used for 15 independent variables and in the second, the independent variables were reduced by genetic algorithm and then the missing values were imputed through MICE algorithm. In the third case, the independent variables were reduced using genetic algorithm and then the missing data was imputed by MICE optimized algorithm.

4.1. Application of logistic algorithm for all variables

In this stage, logistic regression was applied for all variables. The algorithm was run for 10,000 times and then Sensitivity, Specificity and Accuracy algorithms were run on both test and test data.

4.2. Application of logistic regression along with dimensions' reduction with the help of genetic algorithm and MICE algorithm

First, the missing values were imputed by MICE algorithm and then the number of independent variables were reduced to nine by the use of genetic algorithm. Education, race, PMH, age at diagnosis, family history of gastric cancer, age at Dx of family GC, Hx of other GI cancer, type of other GI Cancer, Hx. of extra GI cancer were the nine parameters associated with patient survival. After this reduction, logistic regression algorithm was run on these nine variables and repeated 10,000 times. Then the algorithm of Sensitivity, Specificity and Accuracy were run for training, as well as test data.

4.3. Application of logistic regression with dimensions' reduction with the help of genetic algorithm and optimized MICE algorithm

As explained in 2-1-1, the missing values were imputed by data mining algorithms, and independent variables were reduced to six: Sex, birth year, age at diagnosis time, age at diagnosis time of patients' family, family history of gastric cancer, and family history of other gastrointestinal cancers. After the reduction of the variable, logistic regression algorithm was run on these six variables for 10,000 times. Then, Sensitivity, Specificity and Accuracy algorithm was run on test data. The logistic regression coefficients are described in Table 3. If the regression coefficient is positive, the rise of that variable causes the rise of patient's survival rate, and if it is negative, its rise causes the fall of the patient's survival rate.

4.4. A comparison among the three techniques for problem solution

In this stage, the three different techniques for problem solution were compared by considering the three assessment parameters. The details of this comparison are demonstrated in Figure 4. As seen in Figure 4, the optimized MICE technique has worked better on all three parameters of Sensitivity, Specificity and Accuracy.

Table 3. The coefficients of the recommended model of logistic regression

Beta ₀	Sex		Birth Year		Age at Diagnosis		Age of diagnosis time at patients' family		Family history of gastric cancer		Type of other GI Cancer	
β_6	e^{β_5}	β_5	e^{β_4}	β_4	e^{β_3}	β_3	e^{β_2}	β_2	e^{β_1}	β_1	β_0	e^{β_6}
-0.25	0.95	-0.05	0.46	-0.76	0.67	-0.40	1.21	0.19	1.29	0.25	-0.25	0.78

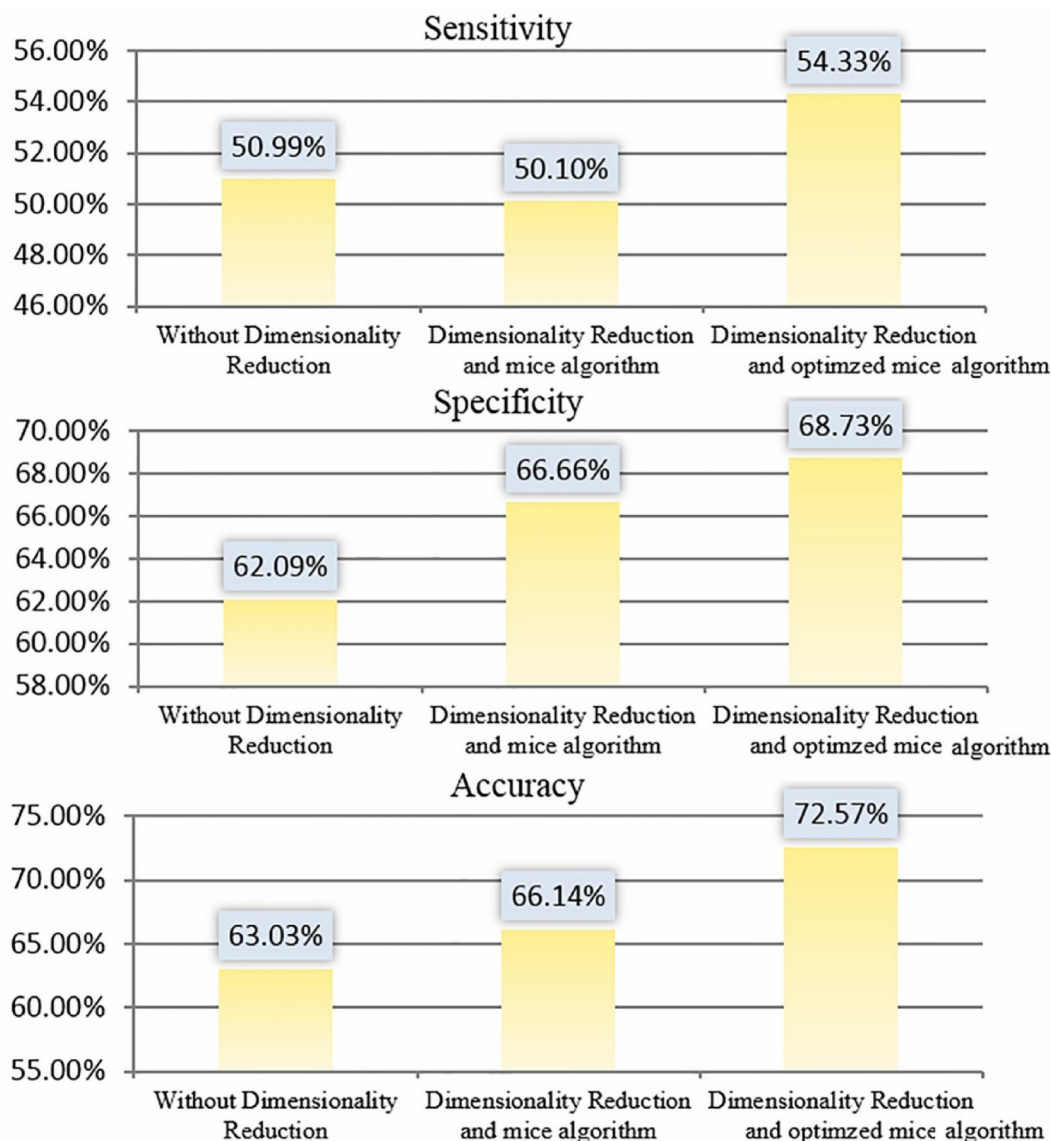


Figure 4. A comparison among the three assessment parameters with the help of three techniques for problem solution

5. Discussion

In this research, a model was designed to predict the chances of survival in gastric cancer patients. From 15 independent variables, 6 variables were chosen for the purpose of modeling. Two dependent variables studied the

survival rate of the patients in two forms of one and two years. According to the logistic regression coefficients, each variable was described. The survival rate falls and if it is one (male), the chances of survival falls. Based on the amount of e^{β_1} according to Table 3, the possibility of survival fall is 29% more in women than in men. In two articles of Yang et al. in 2011 (19) and Zeraati and Amiri in 2016 (20), it was demonstrated that gender directly influences the survival rate in gastric cancer. The 2nd, 3rd and 4th variables were patient's age, the age of diagnosis and age of diagnosis time at patients' family, respectively. As the age rises, the period of patient's survival falls and vice versa. The two studies of De Angelis et al. in 2014 (21) and Ferro et al. in 2014 (22) reveal that as the age of patient as well as the age of diagnosis rise, the survival period falls. Due to a negative coefficient for the variable "Family history of gastric cancer", if there has been any case of gastric cancer in the family, the possibility of survival decreases and in the case of gastric cancer in relatives, the possibility of survival rises. Based on the amount of e^{β_1} according to Table 3, the chance of survival in patients with gastric cancer history in the family is 5% less than patients with gastric cancer history in relatives. In the paper of Rugge et al in 2015 (23), the influence of cancer history in family on survival possibility has been revealed. Due to a negative coefficient for the variable "Other type of cancer", if this variable is small intestine or liver, survival rate falls and if it is esophagus or large intestine, survival rate rises. The most noticeable limitation in this study is the low number of records used for modeling. Therefore, it is highly recommended that this study be repeated in future with more records to find a model with higher level of accuracy, although the designed model with these available records was good enough in accuracy, which can be among the advantages of this study. As seen in Figure 4, with dimensions' reduction algorithm as well as MICE optimized algorithm, the accuracy of the model increases from 63.03% to 72.57% which shows the appropriateness of this recommended technique in modeling the prediction of gastric cancer patients' survival rate.

6. Conclusions

In this study, a prediction model is presented for the survival time of patients with gastritis. In this model, six variables of gender, year of birth, age at diagnosis, age of diagnosis in the family of patients, family history of gastric cancer, and family history of other gastrointestinal cancers affect survival time. Having a patient's survival time can be a great help in determining the treatment method for doctors. It is suggested that the proposed model be evaluated for other examples. A further study on the methods for extracting variables that affect gastric cancer can be appropriate for future research.

Acknowledgments:

This work was a part of the first author's PhD dissertation, supported by a grant [grant #931034] from Mashhad University of Medical Sciences Research Council. The funder was involved in the preparation and publication process of the manuscript.

Conflict of Interest:

There is no conflict of interest to be declared.

Authors' contributions:

All authors contributed to this project and article equally. All authors read and approved the final manuscript.

References:

- 1) Ghadimi MR, Mahmoodi M, Mohammad K, Zeraati H, Hosseini M, Sheikh Fathollahi M. Comparison of Survival analysis of Gastrointestinal Cancer patients using parametric and Cox models. *sjsph*. 2010; 8(2): 1–14.
- 2) Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev*. 2014; 23(5): 700–13. doi: 10.1158/1055-9965.EPI-13-1057. PMID: 24618998, PMCID: PMC4019373.
- 3) Baeradeh N, Lotfi M, Fallahzadeh H, Kargar S, Salman Roghani H. Survival rate of patients with stomach cancer and its effective factors in Yazd Province. *J Community Health Res*. 2015; 3:278-87.
- 4) Malekzadeh R, Derakhshan MH, Malekzadeh Z. Gastric cancer in Iran: epidemiology and risk factors. *Arch Iran Med*. 2009; 12(6): 576–83. PMID: 19877751.
- 5) Soleimani A, Hassanzadeh J, Motlagh AG, Tabatabaee H, Partovipour E, Keshavarzi S, et al. Spatial analysis of common gastrointestinal tract cancers in counties of Iran. *Asian Pac J Cancer Prev*. 2015; 16: 4025–9. doi: 10.7314/APJCP.2015.16.9.4025. PMID: 25987080.
- 6) Naghavi M. Iranian annual of national death registration report. Tehran Minist Health Med Educ. 2005.

- 7) Akbari ME, Abachizadeh K, Tabatabaei SM, Ghanbari Motlagh A, Majd Jabari Z, Khaiamzadeh M, et al. Cancer in Iran. 1st ed. Qom: Darolfekr Publications; 2008.
- 8) Tehran Iran: Cancer Office, Center for Disease Control, Deputy for Health, Ministry of Health and Medical Education; 2002. Iranian Annual of Cancer Registration 2001-2002.
- 9) Sadighi S, Raafat J, Mohagheghi MA, Meemary F. Gastric carcinoma: 5 year experience of a single institute. *Asian Pac J Cancer Prev.* 2005; 6(2): 195–6. PMID: 16101332.
- 10) Zeraati H, Mahmoudi M, Kazemnejad A, Mohammed K. Postoperative life expectancy in gastric cancer patients and its associated factors. *Saudi Med J.* 2005; 26(8): 1203–7. PMID: 16127513.
- 11) Samadi F, Babaei M, Yazdanbod A, Fallah M, Nouraei M, Nasrollahzadeh D, et al. Survival rate of gastric and esophageal cancers in Ardabil province, North-West of Iran. *Arch Iran Med.* 2007; 10(1): 32–7. doi: 07101/AIM.009. PMID: 17198451.
- 12) Eskandar H, Hossein SS, Rahim M, Jalal H, Mehrdad A, Rajabi T. Clinical profile of gastric cancer in Khuzestan, southwest of Iran. *World J Gastroenterol.* 2006; 12(30): 4832–5. PMID: 16937464, PMCID: PMC4087616.
- 13) Movahedi M, Afsharfard A, Moradi A, Moadeli AN, Khoshnevis J, Fattahi F, et al. Survival rate of gastric cancer in Iran. *J Res Med Sci.* 2009; 14(6): 367–73. PMID: 21772910, PMCID: PMC3129080.
- 14) Akhondi-Meybodi M, Ghane M, Akhondi-Meybodi S, Dashti G. Five-year Survival Rate for Gastric Cancer in Yazd Province, Central Iran, from 2001 to 2008. *Middle East J Dig Dis.* 2017; 9(1): 39. doi: 10.15171/mejdd.2016.50. PMID: 28316765, PMCID: PMC5308133.
- 15) Baghestani AR, Hajizadeh E, Fatemi SR. Bayesian analysis for survival of patients with gastric cancer in Iran. *Asian Pac J Cancer Prev.* 2009; 10(5): 823–6. PMID: 20104972.
- 16) Ganji A, Malekzadeh F, Safavi M, Nassri-Moghaddam S, Nourie M, Merat S, et al. Digestive and liver disease statistics in Iran. *Middle East J Dig Dis.* 2009; 1(2): 56–62.
- 17) Golabpour A, Etmnani K, Doosti H, Heidarian Miri H, Ghanbari R. Providing an imputation algorithm for missing values of longitudinal data using Cuckoo Search algorithm: A case study on cervical dystonia. *Electron Physician.* 2017; 9(6): 4648-54. doi: 10.19082/4648.
- 18) Witten IH, Frank E, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques*, Third Edition. Burlington, MA: Morgan Kaufmann; 2011.
- 19) Yang D, Hendifar A, Lenz C, Togawa K, Lenz F, Lurje G, et al. Survival of metastatic gastric cancer: Significance of age, sex and race/ethnicity. *J Gastrointest Oncol.* 2011; 2(2): 77–84. doi: 10.3978/j.issn.2078-6891.2010.025. PMID: 22811834, PMCID: PMC3397601.
- 20) Zeraati H, Amiri Z. Estimating postoperative survival of gastric cancer patients and factors affecting it in Iran: Based on a TNM-7 Staging System. *Acta Med Iran.* 2016; 54(2): 114–8. PMID: 26997598.
- 21) De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: results of EURO CARE--5-a population-based study. *Lancet Oncol.* 2014; 15(1): 23–34. doi: 10.1016/S1470-2045(13)70546-1. PMID: 24314615.
- 22) Ferro A, Peleteiro B, Malvezzi M, Bosetti C, Bertuccio P, Levi F, et al. Worldwide trends in gastric cancer mortality (1980–2011), with predictions to 2015, and incidence by subtype. *Eur J Cancer.* 2014; 50(7): 1330–44. doi: 10.1016/j.ejca.2014.01.029. PMID: 24650579.
- 23) Ruge M, Fassan M, Graham DY. Epidemiology of gastric cancer. In: *Gastric Cancer*. Springer. 2015: 23–34. doi: 10.1007/978-3-319-15826-6_2.