



Estimation Recurrence Free Survival of the Epithelial Ovarian Cancer Using Classification and Regression Tree

Maryam Deldar^{1,2,3}, Kouros Sayehmiri^{4*}, Robab Anbiaee⁵ and Anahita Jalilian⁶

¹Department of Biostatistics and Epidemiology, School of Health, Kerman University of Medical Sciences, Kerman, Iran

²Bam University of Medical Sciences, Bam, Iran

³Ilam University of Medical Sciences, Ilam, Iran

⁴Department of Biostatistics, Psychosocial Injuries Research Center, School of Health, Ilam University of Medical Sciences, Ilam, Iran

⁵Imam Hossein Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁶School of Medicine, Ilam University of Medical Sciences, Ilam, Iran

*Corresponding author: Department of Biostatistics, Psychosocial Injuries Research Center, School of Health, Ilam University of Medical Sciences, Ilam, Iran. Email: sayehmiri@razi.tums.ac.ir

Received 2022 December 19; Revised 2023 June 18; Accepted 2023 June 20.

Abstract

Background: Epithelial ovarian cancer is one of the leading causes of death from gynecological cancers in the Western world. One of the important objectives of medical research is to determine predictors of an event. Regarding the interaction of risk factors, regression methods are unsuitable when the number of factors is high.

Objectives: Regarding frequency predictors of recurrence-free survival in epithelial ovarian cancer, our aim in this article is to determine predictors and time to first recurrence using a classification and regression tree model.

Methods: This retrospective analysis used medical and chemotherapy records of 141 patients with epithelial ovarian cancer between 2007 and 2018. They were referred to Imam Hossein Hospital in Tehran. Data were analyzed using classification and regression trees in Rver3.4.3.

Results: The regression tree results showed that the worst recurrence-free survival in metastatic patients was in grade II patients (15.03 ± 11 months), but in patients without metastases were in patients with CA125 tumor marker above 207 that used 3-week chemotherapy courses (14.53 ± 6.4 months). The classification tree also showed that the most probability of the first recurrence in metastatic patients was in patients with adjuvant chemotherapy (0.81), and patients without metastases were among those with stages 2, 3, and 4 with the maximum platelet count above 305,000 and less than 35 years old (0.75).

Conclusions: The classification and regression tree models, without any assumptions, can estimate the probability of recurrence in different subgroups. These models can be used in deciding due to the ease of interpretation by physicians and paramedics.

Keywords: Classification and Regression Tree, Ovarian Cancer, Recurrence Free Survival, Predictor

1. Background

Cancer is one of the three leading causes of death worldwide (1). Sexual and reproductive health is an important component of life quality, which can affect patient satisfaction (2). Ovarian cancer (OC) is one of the most lethal malignancies of the female reproductive system (3). Epithelial ovarian cancer is one of the leading causes of death from gynecological cancers in the Western world (4), and only about 40% of women with ovarian cancer survive 5 years after diagnosis (5). In Iran, ovarian cancer is the eighth most common cancer, with an age-standard incidence of 3.9/100,000 (6). Although the five-year survival rate has increased in the last decade, the

low survival rate causes recurrence (7). Ovarian cancer often does not have a specific symptom in its early stages, so it is spread at diagnosis (8).

Methods based on tree models, unlike classical methods, require fewer assumptions and include a wider range of data, so in the last two decades, these methods have become more popular than classical models. In particular, these models fit well with high data volumes, and the classical models have no problems with missing data. They also do not require the strict preconditions of common regression models, such as the normal distribution and homogeneity of variances (9). A decision tree is one of the nonparametric methods for classifying data that, according to the nature of the dependent

variable, is divided into two categories, a classification tree for classification variables and a regression tree for continuous variables (10). Classification and regression tree with the growth of the largest tree creates a sequence of trees and prunes it so that only the root node remains. Then it uses cross-validation to estimate the incorrect classification cost of each subtree and selects the tree with the lowest estimated cost (11). The classification and regression tree (CART) considers the values of the predictors sequentially, meaning the variables are arranged according to their importance (12).

In a study conducted in China in 2018 to determine the effect of age on the survival of head and neck cancers, the classification and regression tree showed that the relative importance of age was 3.21% for oral cancer, 8.32% for oropharyngeal cancer, 2.56% for hypopharyngeal cancer and 16.51% for laryngeal cancer (13). In another study using the decision tree to determine lung cancer predictors in 2010, the best predictor by using CART was exposure to known lung carcinogens, the second predictor was 8.6 years or higher latency, and the third predictor was the smoking history of fewer than 11.25 packages in a year (14). In another study conducted in China in 2012 on non-small-cell lung cancer patients treated with gefitinib after chemotherapy, a classification and regression tree formed four subgroups, and the median PFS in the four subgroups ranged from 12 to 42 months (15).

2. Objectives

Many studies have investigated the factors affecting the survival or recurrence of ovarian cancer. However, so far, no study using the Classification and Regression Tree (CART) has investigated the recurrence rate and factors affecting the recurrence of epithelial ovarian cancer. We investigate the factors affecting the first recurrence of epithelial ovarian cancer over an 11-year period using the CART tree.

3. Methods

3.1. Participants

We considered a set of variables that were predictors of the first recurrence of epithelial ovarian cancer. This retrospective study reviewed medical and chemotherapy records of patients with ovarian cancer who had been referred to the oncology and radiotherapy department of Imam Hossein Hospital in Tehran between 2007 and 2018. Therefore, we did not need to determine the sample size, and the sampling was a census of the patients referred to Imam Hossein Hospital in Tehran, which we can consider

as a sample of the entire population of epithelial ovarian cancer. The data were extracted from the patient's files in the hospital's archives.

All patients underwent the first stage of treatment, including the first stage of chemotherapy, after the initial cancer diagnosis. Recurrence of tumor masses in any part of the body had been detected by imaging, with or without biopsy, but it had usually been detected by the tumor marker CA125, and chemotherapy had restarted after recurrence. So the recurrence time for patients who had experienced recurrence was between the end of the first chemotherapy period and the restart of later chemotherapy. For patients who had been censored, the time of censoring was recurrence time, and for patients who had not experienced recurrence to the end of follow-up, time to the end of follow-up was considered time to censor. The first recurrence by the end of 2018 was considered a failure.

Potential prognostic variables accessed in this study included clinical variables such as age, body mass index, presence of ascites at diagnosis, tumor-related variables including tumor stage, tumor grade, metastatic tumor, tumor histology, tumor size, CA125 tumor marker at diagnosis, and hematologic variables such as white blood cells, platelets, hemoglobin, type of chemotherapy (adjuvant or neoadjuvant) and chemotherapy courses.

3.2. Statistical Analysis

To determine the effect of the variables under study on the first recurrence of epithelial ovarian cancer, we defined time to first recurrence as a dependent variable. To calculate the median time to the first recurrence, the time to the first recurrence was calculated for all patients, and then the median was calculated from these times. Because the time until the first recurrence is a continuous variable, the regression tree was used to determine the importance of the variables in the time until the first recurrence. In the regression and classification tree, the dependent variable is placed in the tree leaves, so the time to the first recurrence is reported in the leaves, and we can calculate the shortest time to the first recurrence and the path from the root to these leaves. This path includes patients with specific clinical characteristics.

In the classification tree, due to the fact that the dependent variable must be a classification variable, it was considered to have the first recurrence and censoring as a classification variable. In leaves is reported probability the first recurrence and based on it having recurrence or censoring have belonged them. In this work, instead of having or not having recurrence was reported the probability of recurrence in the leaves. The leaves can contain different values of the probability of recurrence,

Table 1. Descriptive of Continuous Variables Identified as Risk Factors for Epithelial Ovarian Cancer

Continuous Variables	Range	Mean	Median
Age, y	23 - 82	52.7	52
BMI, kg/m ²	12.66 - 39.45	27.15	27
Baseline CA125	8 - 4900	619	200
Minimum platelet count	76000 - 410000	152950	135000
Maximum platelet count	178000 - 819000	381000	353000
Mean platelet count	126000 - 516000	251190	232750
Mean white blood cells	3065 - 10627	5192	4965
Mean hemoglobin	8.36 - 13.23	10.87	11
Minimum hemoglobin	7.3 - 12.8	9.66	9.6
Size of the primary tumor, cm	1 - 26	10.52	9.75

based on which the highest probability of recurrence and its path from the root to the leaf can be reported that include a certain class of patients. Statistical analyses were performed using R (Ver 3.4.3). $P \leq 0.05$ was considered as statistical significance.

4. Results

4.1. Descriptive Results

In this retrospective study, 141 eligible patients were included in our follow-up from 2007 to 2018; 58 (41%) patients had a first recurrence in our follow-up, and the rest either did not experience the first recurrence or were censored during this period (59%). So the median time to the first recurrence in these patients was 17 (0.5 - 0.57) months, and the median age of the patients was 52 (23 - 82) years. Disease-free survival was 0.82 for one year, 0.55 for two years, 0.44 for three years, 0.42 for four years, 0.39 for five years, and 0.36 for ten years. [Table 1](#) and [Table 2](#) show the properties of the selected continuous and discrete variables, respectively.

4.2. Survival Analysis Using the CART Method

4.2.1. Regression Tree

Terminal nodes in the regression tree were time to the first recurrence of epithelial ovarian cancer. The root of the regression tree was based on metastatic tumors so that the shortest time to first recurrence in patients with metastasis was among patients with grade 2 tumor (15.03 ± 11 months), also in patients without metastasis, the shortest time to the first recurrence was among patients who had a CA125 tumor marker at diagnosis above 207 and used 3-week chemotherapy courses (14.53 ± 6.4 months). The number of individuals and the mean time to the first recurrence is specified in all nodes of the regression tree, including the final nodes, as shown in [Figure 1](#).

Table 2. Descriptive of Categorical Variables Identified as Risk Factors for Epithelial Ovarian Cancer

Categorical Variables	No. (%)
Metastatic tumor	
Yes	52 (36.9)
No	80 (56.7)
Tumor grade at diagnosis	
Grade I	24 (17)
Grade II	32 (22.7)
Grade III	37 (26.2)
FIGO stage at diagnosis	
Stage I	34 (24.1)
Stage II	16 (11.3)
Stage III	44 (31.2)
Stage IV	14 (9.9)
Baseline ascites	
Presence of ascites	58 (41.1)
No presence of ascites	78 (55.3)
Chemotherapy course	
Three weeks	64 (45.4)
One week	37 (26.2)
Adjuvant chemotherapy	
Adjuvant	64 (45.4)
Neoadjuvant	19 (13.5)
Tumor histology	
Papillary serous	85 (60.3)
Others (endometrioid, clear cell, mucinous)	30 (21.2)

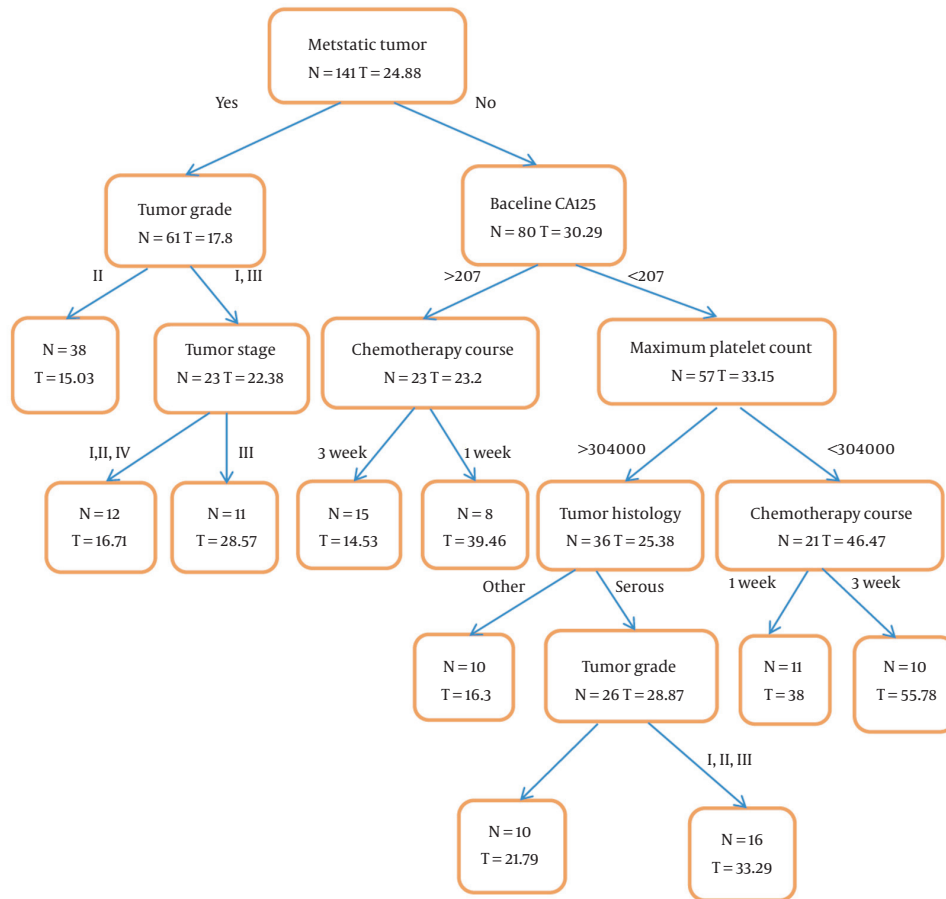


Figure 1. Regression tree for the first recurrence of epithelial ovarian cancer. Nodes display the time to the first recurrence of patients with epithelial ovarian cancer. T= mean time to first recurrence (month)

4.2.2. Classification Tree

Having or not having the first recurrence reported in terminal nodes of the classification tree during our study, our response was two-state. The root of the classification tree is based on the metastatic tumor. In this tree, the number of individuals and the probability of the first recurrence in each node has been determined, especially in the final nodes was reported the probability of the first recurrence, so the probability of the first recurrence in each specific class of patients through the classification tree was reported (in this tree don't report time to recurrence). The results are shown in Figure 2. Descriptive statistics were reported in the nodes and leaves of the tree. So the risk of the first recurrence is highest among patients without metastasis who received adjuvant chemotherapy (0.81). Also, the highest probability of the first recurrence in patients without metastases was among patients with stages 2,3, and 4 with a maximum platelet count above

305,000 and less than 35 years old (0.75).

The variables' importance in the first recurrence of epithelial ovarian cancer using the classification and regression tree model is shown in Table 3. According to the regression tree results, chemotherapy courses with 17% significance and metastatic tumor with 14% significance were in the first and second category, respectively, and according to the results of the classification tree, tumor stage with 21% significance and metastatic tumor with 15% importance are in the first and second category.

5. Discussion

In our study using the regression tree, the shortest time to the first recurrence in metastatic patients was among patients with grade 2 tumor (15.03 months), but in patients without metastasis, the shortest time to the first recurrence was among patients that had the

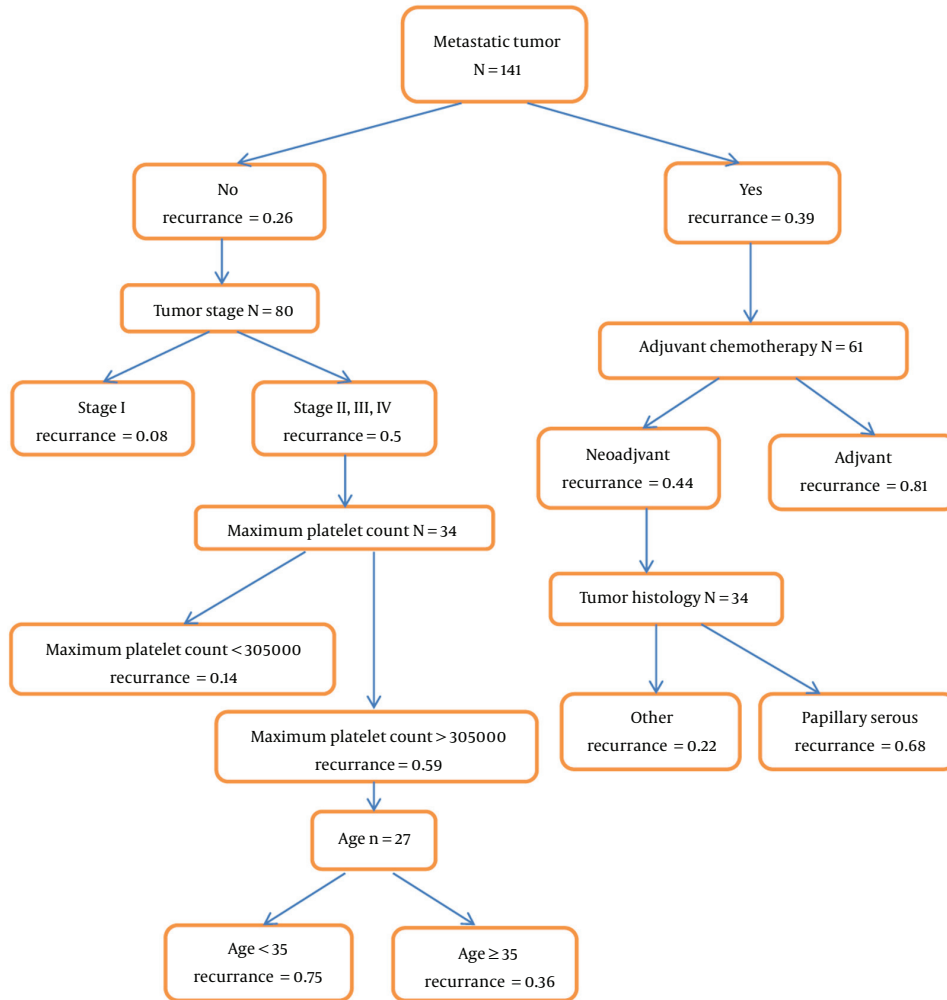


Figure 2. Classification tree for the first recurrence of epithelial ovarian cancer. Nodes display the percentage of the first recurrence of patients with epithelial ovarian cancer.

tumor marker CA125 above 207 at diagnosis and used three-week chemotherapy courses (14.53 months). Also, using the classification tree, the risk of first recurrence in metastatic patients that received adjuvant chemotherapy is the highest possible. But patients without metastases at stages 2, 3, and 4, with a maximum platelet count above 305,000, and those under 35 years old had a higher risk of recurrence. In both the regression and classification trees, the metastatic tumor was identified as an important risk factor for the first recurrence of epithelial ovarian cancer.

In our study, the median time to the first recurrence of epithelial ovarian cancer among patients referred to Imam Hossein Hospital in Tehran was 17 (0.5 - 127) months, and the median age of patients in our study was 52 (23 - 82) years old. In the study of Komura et al. in 2019, this

median age in Japan was less than 59 years (16). In our study, tumor stage was the most important risk factor for epithelial ovarian cancer identified by the classification tree. In the Clarke et al. study in 2019 that examined long-term survival predictors of patients with grade III and IV serous ovarian cancer, the lower stage of the disease was significantly associated with long-term survival (17), which is consistent with our study. In our study in people without metastases, a tumor marker above 207 shortens the time to recurrence; a study in 2018 showed that higher levels of the CA125 marker tumor increase the probability of abdominal recurrence in high-grade serous ovarian cancer patients (18).

Various studies have investigated the factors affecting the diseases using classification and regression trees, but

Table 3. Variables Importance by the CART Model

Regression Tree		Classification Tree	
Variables	Importance, %	Variables	Importance, %
Chemotherapy course	17	FIGO stage at diagnosis	21
Metastatic tumor	14	Metastatic tumor	15
Tumor grade at diagnosis	10	Tumor histology	14
Baseline CA125	9	Tumor grade at diagnosis	12
Maximum platelet count	9	Adjuvant chemotherapy	10
FIGO stage at diagnosis	9	Age	7
Age	8	Maximum platelet count	6
Mean platelet count	6	Mean platelet count	5
Adjuvant chemotherapy	6	Chemotherapy course	5
Tumor histology	6	Baseline ascites	3
Minimum hemoglobin	3	Minimum hemoglobin	1
Mean hemoglobin	2		
Baseline ascites	1		

this method has not been used in ovarian cancer, so we used in this section other studies that apply classification and regression trees to other diseases, especially cancers. Based on a study by Saki Malehi et al. that used a decision tree to evaluate prognostic variables in the survival rate classification of patients with colorectal cancer. The decision tree model showed that disease stage at diagnosis, patient age at the time of diagnosis, tumor morphology, and disease severity are important prognostic factors in the survival of patients with colorectal cancer (19).

Navarro Silvera et al.'s study in 2014 that investigated diet and lifestyle as risk factors in patients with gastric and esophageal cancer showed that the frequency of symptoms of gastroesophageal reflux disease was reported to be the most important risk factor. For esophageal cell cancer, smoking was the most important risk factor (20). In 2019 Greene et al. used classification and regression trees to predict cervical cancer screening; this model identified subgroups with the probability of receiving screening and several new variables that may underlie the use of SMW in cervical cancer screening (21).

5.1. Limitations

One disadvantage of classification and regression trees is that the CART tree divisions are binary, and when the number of variable levels is greater than two, the results can be confusing.

5.2. Conclusions

When the number of predictor variables is high, due to the interaction effects of the variables, regression

methods are not very suitable. Classification and regression tree models without the need for none specific assumptions can predict the recurrence probability of different subgroups. These models do not require special knowledge due to ease of interpretation and can be easily used by physicians and paramedics.

Footnotes

Authors' Contribution: Study concept and design: M. D., and K. S.; Analysis and interpretation of data: M. D., and K. S.; Drafting of the manuscript: M. D.; Critical revision of the manuscript for important intellectual content: M. D., K. S., R. A., and A. J.; Statistical analysis: M. D.

Conflict of Interests: The authors declare no conflict of interest.

Data Reproducibility: The dataset presented in the study is available on request from the corresponding author during submission or after publication. The data are not publicly available due to privacy.

Ethical Approval: This study is approved under the ethical approval code of "IR.MEDILAM.REC.1398.071" at Ilam University of Medical Sciences.

Funding/Support: This study was not supported by any grant.

References

- Zarbaksh S, Tabrizi Amjad M, Yousefi B, Aldaghi M, Sameni H. Histopathological and Follicular Atresia Assessment of Rat's

- Ovarian Tissue Following Experimental Chronic Spinal Cord Injury. *Middle East J Rehabil Health*. 2017;**In Press**(In Press). <https://doi.org/10.5812/mejrh.14303>.
2. Mirmohammadkhani M, Ghahremanfard F, Tayyebi K, Ghadamgahi HB. Early Direct Costs of Diagnostic and Therapeutic Services for Patients with Cancer: A Descriptive Study in Semnan, Iran, 2011 - 2014. *Middle East J Rehabil Health*. 2019;**In Press**(In Press). <https://doi.org/10.5812/mejrh.55457>.
 3. Huang L, Zhang J, Deng Y, Wang H, Zhao P, Zhao G, et al. Niclosamide (NA) overcomes cisplatin resistance in human ovarian cancer. *Genes Dis*. 2023;**10**(4):1687-701. [PubMed ID: 37397523]. [PubMed Central ID: PMC10311098]. <https://doi.org/10.1016/j.gendis.2022.12.005>.
 4. Marchetti C, Kristeleit R, McCormack M, Mould T, Olaitan A, Widschwendter M, et al. Outcome of patients with advanced ovarian cancer who do not undergo debulking surgery: A single institution retrospective review. *Gynecol Oncol*. 2017;**144**(1):57-60. [PubMed ID: 27825669]. <https://doi.org/10.1016/j.ygyno.2016.11.001>.
 5. Protani MM, Nagle CM, Webb PM. Obesity and ovarian cancer survival: a systematic review and meta-analysis. *Cancer Prev Res (Phila)*. 2012;**5**(7):901-10. [PubMed ID: 22609763]. <https://doi.org/10.1158/1940-6207.CAPR-12-0048>.
 6. Akbari A, Azizmohammad Looha M, Moradi A, Esmaeli Akbari M. Ovarian Cancer in Iran: National Based Study. *Iran J Public Health*. 2023. <https://doi.org/10.18502/ijph.v52i4.12453>.
 7. Palomar Munoz A, Cordero Garcia JM, Talavera Rubio MDP, Garcia Vicente AM, Pena Pardo FJ, Jimenez Londono GA, et al. Value of [18F]FDG-PET/CT and CA125, serum levels and kinetic parameters, in early detection of ovarian cancer recurrence: Influence of histological subtypes and tumor stages. *Medicine (Baltimore)*. 2018;**97**(17). e0098. [PubMed ID: 29702969]. [PubMed Central ID: PMC5944512]. <https://doi.org/10.1097/MD.00000000000010098>.
 8. Gupta D, Lis CG. Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *J Ovarian Res*. 2009;**2**:13. [PubMed ID: 19818123]. [PubMed Central ID: PMC2764643]. <https://doi.org/10.1186/1757-2215-2-13>.
 9. Sheykholeslami AS, Behnampour N, Mohammadpour RA, Abdollahi F. [Application of Survival Tree Model in Determining Affecting Factors in Breastfeeding Duration]. *Iran J Health Sci*. 2021;**9**(2):9-17. Persian. <https://doi.org/10.18502/jhs.v9i2.6567>.
 10. Behnampour N, Hajizadeh E, Semnani S, Zayeri F. The introduction and application of classification tree model for determination of risk factor for esophageal cancer in Golestan province. *Jorjani Biomed J*. 2013;**1**(2):38-46.
 11. Loh W. *Classification and Regression Tree Methods*. Wiley StatsRef: Statistics Reference Online; 2014.
 12. Chester R, Khondoker M, Shepstone L, Lewis JS, Jerosch-Herold C. Self-efficacy and risk of persistent shoulder pain: results of a Classification and Regression Tree (CART) analysis. *Br J Sports Med*. 2019;**53**(13):825-34. [PubMed ID: 30626599]. <https://doi.org/10.1136/bjsports-2018-099450>.
 13. Yang CC, Su YC, Lin YW, Huang CI, Lee CC. Differential impact of age on survival in head and neck cancer according to classic Cox regression and decision tree analysis. *Clin Otolaryngol*. 2019;**44**(3):244-53. [PubMed ID: 30578588]. <https://doi.org/10.1111/coa.13274>.
 14. Kim TW, Koh DH, Park CY. Decision tree of occupational lung cancer using classification and regression analysis. *Saf Health Work*. 2010;**1**(2):140-8. [PubMed ID: 22953174]. [PubMed Central ID: PMC3430888]. <https://doi.org/10.5491/SHAW.2010.1.2.140>.
 15. Sun H, Guo J, Liu Y, Wang Z. Classification and regression tree analysis of patients with non-small-cell lung cancer treated with gefitinib after chemotherapy. *Thorac Cancer*. 2013;**4**(3):280-6. [PubMed ID: 28920234]. <https://doi.org/10.1111/1759-7714.12014>.
 16. Komura N, Mabuchi S, Isohashi F, Yokoi E, Shimura K, Matsumoto Y, et al. Radiotherapy for isolated recurrent epithelial ovarian cancer: A single institutional experience. *J Obstet Gynaecol Res*. 2019;**45**(6):1173-82. [PubMed ID: 30843318]. <https://doi.org/10.1111/jog.13947>.
 17. Clarke CL, Kushi LH, Chubak J, Pawloski PA, Bulkley JE, Epstein MM, et al. Predictors of Long-Term Survival among High-Grade Serous Ovarian Cancer Patients. *Cancer Epidemiol Biomarkers Prev*. 2019;**28**(5):996-9. [PubMed ID: 30967418]. [PubMed Central ID: PMC6500478]. <https://doi.org/10.1158/1055-9965.EPI-18-1324>.
 18. Shinagare AB, Balthazar P, Ip IK, Lacson R, Liu J, Ramaiya N, et al. High-Grade Serous Ovarian Cancer: Use of Machine Learning to Predict Abdominopelvic Recurrence on CT on the Basis of Serial Cancer Antigen 125 Levels. *J Am Coll Radiol*. 2018;**15**(8):1133-8. [PubMed ID: 29789232]. <https://doi.org/10.1016/j.jacr.2018.04.008>.
 19. Saki Malehi A, Hajizadeh E, Fatemi R. [Evaluation of prognostic variables for classifying the survival in colorectal patients using the decision tree]. *Iran J Epidemiology*. 2012;**8**(2):13-9. Persian.
 20. Navarro Silvera SA, Mayne ST, Gammon MD, Vaughan TL, Chow WH, Dubin JA, et al. Diet and lifestyle factors and risk of subtypes of esophageal and gastric cancers: classification tree analysis. *Ann Epidemiol*. 2014;**24**(1):50-7. [PubMed ID: 24239095]. [PubMed Central ID: PMC4006990]. <https://doi.org/10.1016/j.annepidem.2013.10.009>.
 21. Greene MZ, Hughes TL, Hanlon A, Huang L, Sommers MS, Meghani SH. Predicting cervical cancer screening among sexual minority women using Classification and Regression Tree analysis. *Prev Med Rep*. 2019;**13**:153-9. [PubMed ID: 30591857]. [PubMed Central ID: PMC6305684]. <https://doi.org/10.1016/j.pmedr.2018.11.007>.