# The Gulf Journal of Oncology

Indexed By PubMed and Medline Database

Issue 40, September 2022
ISSN No. 2078-2101


The Official Journal of the Gulf Federation For Cancer Control

# Table of Contents

## Original Articles



## Review Article



## Case Reports



## Conference Highlights/Scientific Contributions

Original Article

# Using Data Mining and Association Rules for Early Diagnosis of Esophageal Cancer

Seyed Mohammad Saleh Hadavi[1,3], Shahram Oliaei[1], Sandra Saidi[1], Elham Nadimi[2], Mohammad Hassan Kazemi–Galougahi[1]

[1]Aja University of Medical Sciences, Tehran, Iran
[2]Shiraz University of Medical Sciences, Shiraz, Iran
[3]Shiraz University of Technology, Shiraz, Iran

## Abstract:

From 17,000 new cases of esophageal cancer worldwide during last year, 16,000 proved to be fatal. Late or incorrect diagnosis of esophageal cancer cases increases its fatality rate. Today, a data–mining technique can predict the course of the disease with the help of an up–to–date technology. With this knowledge, we can reduce esophageal cancer mortality. This study aims to find an association between general characteristics, screening tests, and esophageal cancer based on raw data from the Cancer Research Center within–person interviews, using data mining and classification techniques on mortality.

The 5–year medical records of 512 esophageal cancer patients and those with problems related to this cancer, with 50 functional characteristics, were included in this model. In order to provide a prognostic and rule discovery model for esophageal cancer suffering, we used preprocessing EM Algorithm. After accurate identification of the data, WEKA Software tools and Java programming language was used to create Association Rule Classifier and Apriori algorithm for the associated rule discovery.

We created 6 significant rules of the association for classification generated by rule miner with 95% and 91% confidence based on screening tests and general attributes, respectively. These substantial rules showed significant association between age, history of medication, smoking, gender, carcinoembryonic antigen (CEA), creatinine, WBCs, and Platelets.

The findings of this study can be used as a clue for physicians to consider patients with these characteristics as people who are more likely to develop esophageal cancer and help them for early diagnosis of patients.

**Keywords:**Data mining, esophageal cancer, association rule, healthcare

## Introduction:

The capability to collect and store data has grown at a dramatic rate in all disciplines over the past decade and with the new techniques being developed for effective storage and analysis.

Also, there is no exception in healthcare. The shift toward evidence–based research presents important opportunities to extract significant information and transform it into knowledge from clinical data[(1)].

Esophageal cancer begins in the inner layer of the esophageal wall and grows outward. If spread through the esophageal wall, it can help the lymph nodes, as well as blood vessels in the chest and other adjacent organs. Esophageal cancer can also spread to the lungs, liver, stomach and other parts of the body[(2)].

It is estimated that 16,080 deaths (3,060 women and 13,020 men) from this disease happen in a year. The overall relative 5–year survival rates over time increase gradually in women and men; for example, the rate was under 2% in 1995 to more than 10% in 2018 (SEER). Esophageal cancer is the sixth leading cause of death among all cancers[(3)].

Understanding data across multiple systems is challenging and various integration techniques with varying levels of difficulty have been proposed to solve the problem of data integration and storing. However, research reveals that the current designs are not efficient for data sets with large numbers of attributes that vary over time[(4)].

**Corresponding Author:** Mohammad Hassan Kazemi–Galougahi (PhD), Department of Social Medicine, Faculty of Medicine, Aja University of Medical Sciences, Tehran, Iran,
Email Address: farshad.kazemi@yahoo.com,
Phone number: +989122405076

The characteristics of clinical data as it originates during the process of clinical documentation include issues of data availability and complex representation models that can make data mining process stimulating. Therefore, data preprocessing and transformation are required before one can apply data mining algorithms on clinical data[5].

The stored data in the database would provide a basis for the analysis of disease risk factors.

There is an interdisciplinary area which focusing on methodologies for extracting useful knowledge from data named KDD (Knowledge Discovery and Data Mining). The examining knowledge extraction from data revision after researching statistics, databases, pattern recognition, machine learning, data visualization, optimization, and high–performance computing, to provide advanced business intelligence and web discovery solutions.[6]

CRISP–DM is an abbreviation forCross–Industry Standard Process for Data Mining. The CRISP–DM methodology makes available a structured approach to planning a data mining project. It is a robust and well–proven methodology; however, we do not claim any ownership over it. As shown below the CRISP–DM model (Figure 1)[7].

Data mining method has been a common practice in evidence–based medicine, which is an approach where a clinician is aware of the evidence in support of the clinical practice, and its associated strength. Generally, clinicians and researchers do not care how this method works, but the results are important to them[8].

Association rule mining is being applied to search for hidden relationships between prevention factors by finding the factors frequently appearing together in the cancer database[9]. An association rule has two parts: a precursor (if) and a consequent (then). A precursor is an item found within the data by the machine. A consequent is an item found in combination with the antecedent. Rules are type of the most human–understandable knowledge; therefore, it is most suitable for deciphering new rules conforming to the data associated with medical applications. Association rule mining is valuable for analyzing and predicting future events among large setoff data items. It also is a general–purpose rule observation scheme that has been widely used for observing the rules in medical applications[10].

Since the Apriori association is one of the most important algorithm, which states"that if" an item–set is infrequent, all of its supersets must be infrequent[11].

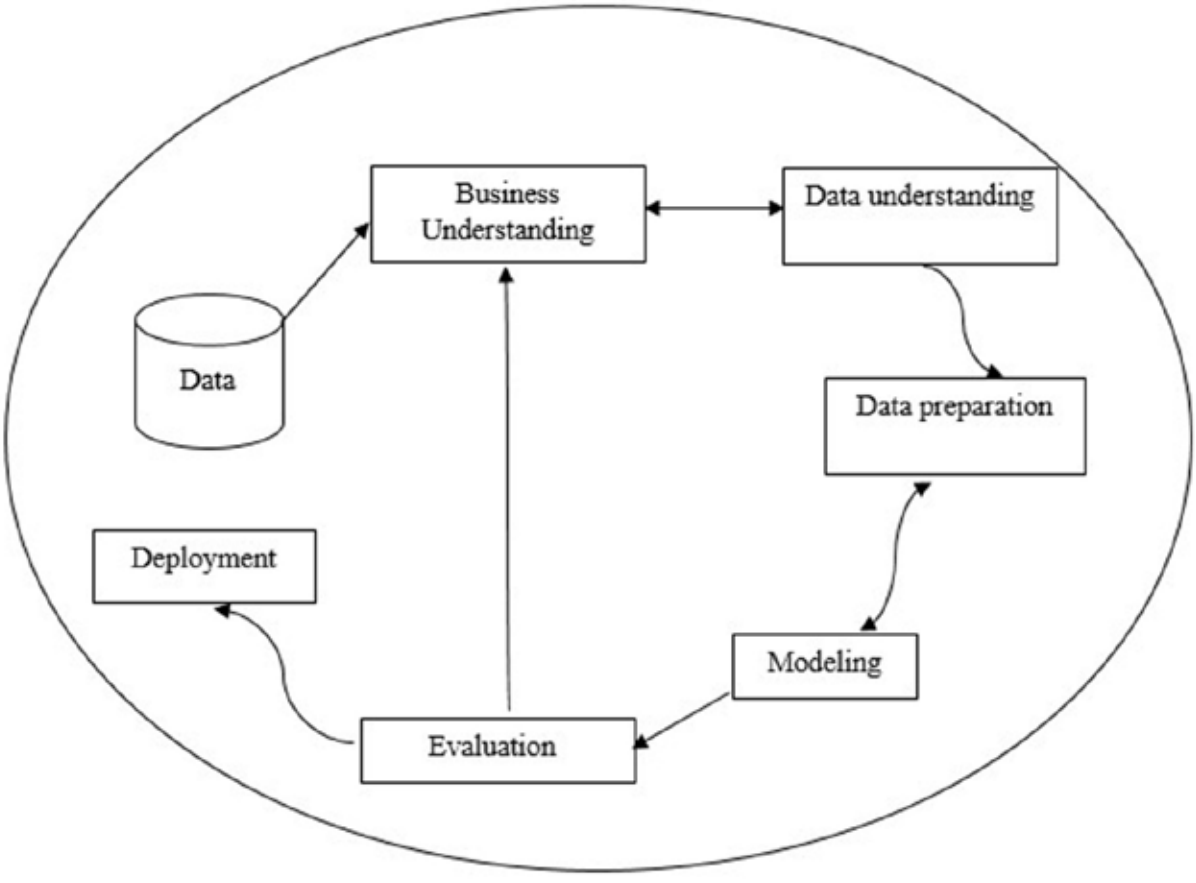


**Figure 1:** The diagram of CRISP methodology

The objective of this study is to present a data mining process that can be used to predict esophageal cancer from raw esophageal cancer patient's datain the primary stage and observe association rules based on clinical diagnostic parameters.

Therefore, in this paper, we researched the adequacy of the classifier for clinical specialists to foresee regardless of whether there are esophageal cancer and the existing knowledge regarding disease co–occurrences and also to discover new disease relationships that could potentially lead to improved clinical health care.

## Methods:

The required data were gathered from clinic of cancer records in KhorasanRazavi, Iran, from June 2018 to September 2019, which were organized into a relational data base. The data of 512 patients with probableesophageal cancer who referred to this center were examined by an oncologist, of them 352 cases were confirmed.

In this study, a method similar to the CRIPS–DM methodology was used, which is an ideal sequence of events (Figure 1)[12]. Indeed, in learning the algorithm for data, many tasks were executed sequentially, while it was necessary to revert to previous tasks and repeat actions were repeated. This model tries to perform a method through the data mining process for high accuracy and fast.

The process of data mining consists of three steps including:

**1. Data Preparation Preprocessing:**It is one of the most important stages in the discovery process because it depends on the quality of decision making and the methodology of the algorithms like expectation–maximization (EM). This includes data collection, data cleaning, and data transformation. Data cleaning includes filling the missing values, smoothing random data and detecting or deleting data. The transformation for data is selection of incorporated forms of data and identification of incorrect values based on recommended scale. The independent variables that we utilized are demonstrated as a part of Tables 1 and 2. The dataset was cleaned by handling the missing values, noise, identifying and correcting inconsistencies by using Expectation Maximization (EM) method.

| Variable | Describe |
|---|---|
| Gender | Male, Female |
| Age | Between 13 And 92 |
| Marital status | Single, Married |
| Birthplace | Some Important City In Iran |
| Ethnicity | Fars, Kord, Tork, Lor |
| Education level | Illiterate, Under diploma, Diploma, BSc, MSc,Ph.D. |
| Job | Housewife, Farmer, Worker, Employed, Retired, Others |
| Place of living | Urban, Rural |
| Body Mass Index (BMI) | ($Kg/M^2$) |
| Family history of cardiovascular diseases, hypertension or respiratory diseases | Yes, No |
| Cancer in first–degree relatives | Yes, No |
| Number of first–degree relatives with cancer | Between 1 and 3 |
| History of cardiovascular diseases, hypertension or respiratory diseases in patient | Yes, No |
| History of medication(medicines related to cardiovascular diseases, hypertension or respiratory diseases) | Yes, No |
| History of drinking alcohol | Yes, No |
| History of narcoticsconsumption | Smoking, Opium, Hookah, Syrup, Non |
| Final diagnosis (Esophageal cancer) | Yes, No |

**Table 1:** Variables used for preprocessing by general attribute in esophageal cancer

| Variable | Describe | Normal Range |
|---|---|---|
| Gender | Male, Female | – |
| Age | Between 13 and 92 | – |
| Marital Status | Single, Married | – |
| Body Mass Index (BMI) | Between 11.7 and 46.4 | 18.5–25 |
| Organ that has been biopsied | Esophagus, Lung, Stomach, Cardia | – |
| Treatment Area | Thorax, Esophagus, Chest, Liver, Breast, Stomach | – |
| Family history of cardiovascular diseases, hypertension or respiratory diseases | Yes, No | – |
| Cancer in first–degree relatives | Yes, No | – |
| Number of first–degree relatives with cancer | Between 1 and 3 | – |
| History of cardiovascular diseases, hypertension or respiratory diseases in patient | Yes, No | – |
| White Blood Cell (WBC) | 1400–3800/µl | 4500–11000/µl |
| Hemoglobin (HB) | 4.74–16 g/dl | Male: 14–17.5 g/dl<br>Female 12–15.5 g/dl |
| Platelets (PLT) | 200–55900/µl | 150000–450000/µl |
| Blood Urea Nitrogen (BUN) | 9–91 mg/dl | 8–25 mg/dl |
| Creatinine | 0.5–1.8 mg/dl | Male: 0.7–1.4 mg/dl<br>Female: 0.6–1.3 mg/dl |
| Carcinoembryonic antigen (CEA) | 0.1–18 | <3 |
| Aminotransferase (AST) | 9–56.45 | Male:<37<br>Female:<31 |
| Alanine Aminotransferase (ALT) | 5–107 | Male:<41<br>Female:<35 |
| Alkaline Phosphatase (ALP) | 102–850 | Male:80–306<br>Female:64–306 |
| C–Reactive Protein (CRP) | Negative/Positive | |
| phosphate (P) | Negative/Positive | |
| Creatine Kinase (CK) | Negative/Positive | |
| History Of Drug | Yes, No | |
| History of Alcoholic | Yes, No | |
| Final diagnosis (Esophageal cancer) | Yes, No | |

**Table 2:** Variables used for preprocessing by screening tests attribute in esophageal cancer

To fill the missing data, first the relation of each of the properties to each other was found and then missing data were completed with the relations, after that each attribute which was include 25% missing data, were obtained by EM algorithm separately and finally were completed by meaningful and real data.

In statistics, an expectation–maximization (EM) algorithm is an iterative method to find the maximum likelihood or maximum posteriori evaluations of the parameters in statistical models, where the model depends on undetected covert variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log possibility evaluated using the current estimate for the parameters, and a maximization (M) step, which computes the parameters maximizing the expected log–likelihood found on the E step. These parameter–estimates are then used to determine the distribution of the latent variables in the next E step[13].

Preprocessing in this study depends on two tables separately, including general attribute like lifestyle, medical history information, and screening tests such as common clinical tests (Table 1 and 2).

**2. Data Modeling:**After determining the percentage of training test, validation test and test sets,data analyzed for frequent if–then patterns to create association rules and using the measures support and confidence to identify the most important relationships. Confidence in this algorithm shows the number of times the if–then statements are found true; also support is an indication of how frequently the items appear in the data. A third metric, called Lift, can be used to compare confidence with expected confidence[14].

The law is scrutinized with two scales, one of which is the amount of "Confidence" and the second one is "Support" which are calculated in the formula below.

Confidence $(X \rightarrow Y) = P(Y|X) = P(X \cap Y)/P(X)$(15)

Support $(X \rightarrow Y) = P(X \cap Y)$(15)

Lift $(X \rightarrow Y) = P(X \cap Y)/P(X)*P(Y)$[15]

Association rules are calculated from item–sets, which are made up of two or more items. "If" rules are made from analyzing all the possible item–sets, there could be so many rules that the rules hold slight meaning. With that, association rules are usually created from rules well–signified in data[16].

In this study, after using *Apriroir algorithm* and *Association Rule Classifier* from this dataset we created six significant rules and all the frequent item set gained with at least 60% support standards were exposed for the observation of association rules. The confidence to presume rule was set to at least 85% and the process was performed with precedent iteration rate of value 10.

Apriori algorithm find the frequent item–sets that have minimum support– A subset of a frequent item–set must also be frequent item–set, i.e. if {AB} is a frequent item–set, both {A} and {B}should be a frequent item–set– Iteratively find frequent item–sets with cardinality from 1 to k (k–tem–set). Use the frequent item–sets to generate the association rules[17].

*Association Rule Classifier* is the model generated by an AC and used to label new records consists of association rules, where the consequence corresponds to the class label. As such, they can also appear as a list of "if–then" clauses: if the record matches some conditions (expressed in the left side of the rule, also called antecedent), it is then labeled accordingly to the class on the right side of the rule (or consequent)[18].

We have applied Weka software tool to experiment with these algorithms. Weka is a gathering of machine learning algorithms for data mining tasks. These algorithms can either be connected directly to a dataset or called from

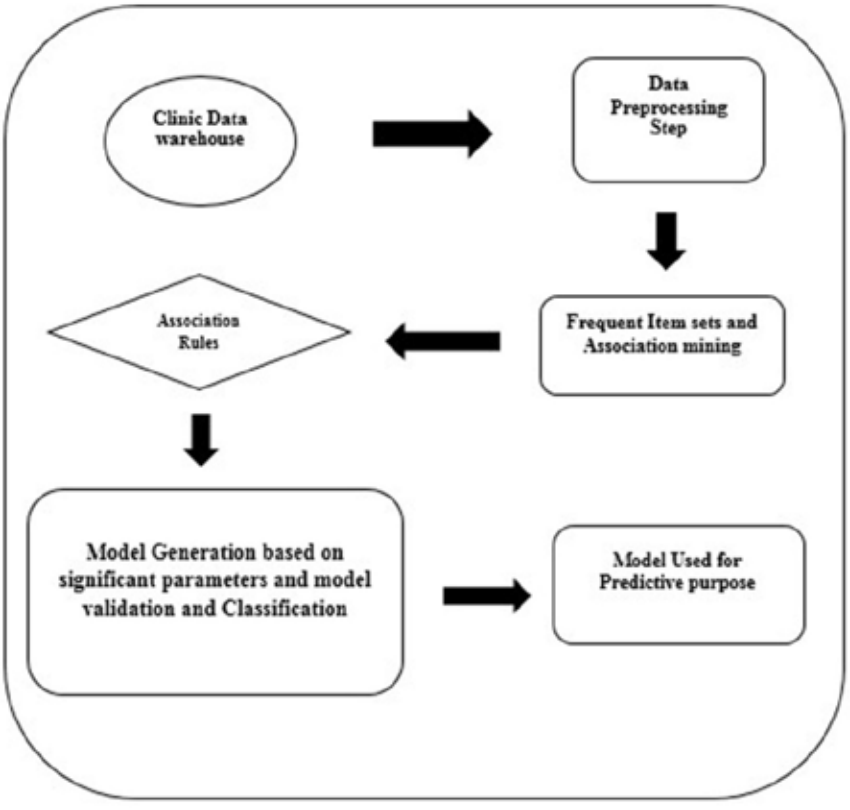


**Figure 2.** The process of data analyzing and data mining method

a Java code. Weka contains apparatuses for data pre–processing, classification, regression, association rules, and visualization are contained by Weka. it is open source software supplied under the GNU General Public License[19].

**3. Feature selection:** Finally, in the feature selection step, the parameters were found to be associated with the suffering of cancer, selected to build a predictive model using a normalized regression approach and the best rules were recognized by the specialist.

## Results:

After initialdata analysisfor preparation and presentation of mean, standard division and missing value, datawere filled the missing values, smoothed random data and detected or deleteddata. Marital status, cancer in first–degree relatives, opium consumption, smoking, CEA, CR, and PLT are the parameters that showed support of at least70% (Table 3).

After the final analysis, the results are shown in two sections. The first was the general factors attribute, including birthplace, gender, age, marital status, body mass index, drug use history, smoking and opium consumption which are associated with confidence leading to esophageal cancer (table 4). Furthermore, the important model achieved is shown in Table 4.

The second one was screening tests attribute containing gender, treatment area, creatinine, marital status, history of cardiovascular diseases, hypertension or respiratory diseases, WBC, CEA, PLT, BMI and CRand significant rules are shown in Table5.

The various association rules within the specified mining criteria causes esophageal cancerwere shown in table 4, such as gender(male), age (more than 35), history of medication (yes), history of esophageal cancer family (yes), marital status (married), history of the disease (yes), history of drinking alcohol(yes), smoking (yes) and opiumconsumption (yes).

On the other hand, some important relations with gender (male), history of esophageal cancer family (yes), creatinine (between 0.89 and 1.02), WBC (less than 4500), HB (12.622 to13.748), PLT (223694.8 to 279579) and CEA (more than 5.4) were discovered that were associated with esophageal cancer.

Another rule was discovered by the algorithm is gender, marital status (married), age (more than 68), ALT (less than 20.6) and ALP (more than 401.2).

| Variable | Mean | SD | Missing Value (%) |
|---|---|---|---|
| Age | 65.1 | 13.1 | 0 |
| BMI | 22.2 | 5.1 | 7 |
| Number of relatives with cancer | 1.1 | 0.6 | 1 |
| WBC | 7164 | 2491 | 30 |
| HB | 12.2 | 2.0 | 28 |
| PLT | 222558.5 | 95334.1 | 19 |
| BUN | 33.1 | 17.6 | 29 |
| Creatinine | 1.0 | 0.2 | 25 |
| CEA | 6.7 | 5.0 | 25 |
| AST | 24.3 | 14.1 | 42 |
| ALT | 20 | 18 | 25 |
| ALP | 247.5 | 100 | 38 |

**Table 3:** Mean, standard deviation and missing value of quantitative variables in participants

| Rules | Result | Confidence | Support |
|---|---|---|---|
| Birthplace=rural, age>35, History of medication=Yes, Marital status= married, opiumconsumption=Yes Smoking=Yes | Final diagnosis = Esophagus cancer | 91% | 75% |
| Family history of diseases=Yes, History of medication=Yes, Marital status= married, Age>35, BMI≥26.5 | Final diagnosis = Esophagus cancer | 90% | 70% |
| Gender = male, Cancer family= Esophagus, Marital status= married, History of the disease= Yes, History ofdrinking alcohol=Yes, Smoking=Yes, Opiumconsumption= Yes | Final diagnosis = Esophagus cancer | 89% | 70% |

**Table 4:** Association rules based on general attribute

| Rules | Result | Confidence | Support |
|---|---|---|---|
| Gender = Male, Treatment area= Esophagus, Creatinine=0.89–1.02, Age>68, History of the disease= Yes, BMI□26.5 | Final diagnosis = Esophageal cancer | 95% | 75% |
| Gender= Male, Cancer family= Esophagus, Creatinine=0.89–1.02, WBC□4500, HB=12.622–13.748, CEA□5.4 | Final diagnosis = Esophageal cancer | 91% | 70% |
| Gender= Male, Marital status= married, CEA□5.4, ALT□20.6, BUN=25.4–33.6 | Final diagnosis = Esophageal cancer | 90% | 60% |

**Table 5:** Association rulesbased on screening tests attribute

## Discussion:

The findings of this study can be used as a clue for physicians to consider patients with these characteristics as people who are more likely to develop esophageal cancer, but it should be noted thatthis is not a confirmed diagnosis.

According to these factors, identifying a prognostic model is proposed to forecast the possible situation of an individual whether suffering from tumor among the associative rules in table4 with 91% confidence (birthplace, age, history of medication, marital status, smoking and opiumconsumption) and associative rules in table 5 with 95% confidence (gender, creatinine, age, history of the disease and BMI).

Nasrollahzadeh et al. used conditional logistic regression models foropium, smoking and tobaccoconsumption to compute odds ratios and 95% confidence intervalsadjusted for potential confounders. The odds of esophageal squamous cell carcinoma (ESCC)in those who used both tobacco and opium were increased compared with those who used neither tobacco nor opium. They were revealed all forms of tobacco use (cigarettes, hookah, and nass) were associated with higher risk of ESCC [20].

The association between opium consumption and drinking alcohol and an increased risk of esophagus, stomach, larynx, lung, bladder, and pancreas cancer have been determined[21].In this study the general factors attribute such as smoking and opium consumption are also associated with confidence leading to esophageal cancer.Furthermore, tobacco smoking may influence DNA methylation in the esophagus and raise the opportunity that these risk factors affect Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC) that have a biologically reasonable role in cancer formation is related with specific demographic and behavioral factors, including gender, fatness/higher body mass index (BMI).[22]

Grabowski JP et al. were founded that median age at diagnosis was 59 years and the effect ofBMI<20 and BMI>30 on cancers. they also were showed impact of BMI on chemotherapy–associated toxicity in ovarian cancer patients[23], which our results were showed 90% confidence and 70% support about generalfeaturesas well.

In this assay, we have three roles according to screen attribute with over 90% confidence and 60% support. Effective factors in this part were elements similar to creatinine which is an analysis product of creatine phosphate in the muscle. Indeed, some risk factors such as postoperative acute kidney injury, serum creatinine level, smoking history, and hypertension were established after esophageal cancer surgery[24].

Plus, it was shown that high hemoglobin levelsare related to raised numbers or sizes of WBCs, another feature in our screen attribute,can also be caused by exposure to high altitudes, dehydration, smoking, advanced lung disease, and assured tumors[25]. LA Tjon–Kon–Fat et al were revealed platelets have pivotal roles in several phases of cancer metastasis[26].

The level of Carcinoembryonic Antigen(CEA) usually is very low (about 9.9 ng/ml) in blood of healthy adults and the age (more than 57) and CEA greater than 5.57 ng/ml is significantly related to shortened survival in patients with esophageal carcinoma[27, 28].

Alanine transaminase (ALT) is a transaminase enzyme which is also called alanine aminotransferase (ALT). MirceaChirica et al. were shown that severe oropharyngeal injuries, increased levels of blood lactate, increased levels of bilirubin, low platelet counts increased, and also levels of alanine aminotransferase (ALT) were prognostic of digestive necrosis[29],which was confirmed in our results.

Methodology of our study, in addition to the method of association relations on this data was conducted for

the first time in Iran. However, the some incomplete data,relatively small sample size and lack of information about the stage of esophagealcancers were our study limitations that are suggested to be considered in future studies.

## Conclusion:

This study mostly focused on observation of clinical factors that can be data mining methods in large and raw datasets with various patients' attributes include general and screening tests, so that a number of significant rules can be exported for predicting recurrence in esophageal cancer more accurately. In our analysis, we built up a prediction model for clinical oncology specialiststouses as a clue for diagnosis ofesophageal cancer. Based on the discoveries made in this study, a prognostic model is proposed for its early identification. Our future work goals are to find some more relevant attributes and manage the missing values that help to forecast esophageal cancer.

## Funding:

This study was supported by research deputy of Aja University of Medical Sciences.

## Conflict of Interest:

The Authors declare that they have no conflict of interests.

## References

1. Chen CP, Zhang C–Y. Data–intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences. 2014;275:314–47.
2. Pawar S. Esophageal cancer.
3. Zhang Y. Epidemiology of esophageal cancer. World journal of gastroenterology: WJG. 2013;19(34):5598.
4. March ST, Hevner AR. Integrated decision support systems: A data warehousing perspective. Decision Support Systems. 2007;43(3):1031–43.
5. Yoo I, Alafaireet P, Marinov M, Pena–Hernandez K, Gopidi R, Chang J–F, et al. Data mining in healthcare and biomedicine: a survey of the literature. Journal of medical systems. 2012;36(4):2431–48.
6. Fayyad UM, Piatetsky–Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining: AAAI press Menlo Park; 1996.
7. Wirth R, Hipp J, editors. CRISP–DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining; 2000: Citeseer.
8. Li J, Fu AW–c, He H, Chen J, Jin H, McAullay D, et al., editors. Mining risk patterns in medical data. Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining; 2005: ACM.
9. Nahar J, Tickle KS, Ali AS, Chen Y–PP. Significant cancer prevention factor extraction: an association rule discovery approach. Journal of medical systems. 2011;35(3):353–67.
10. Verykios VS, Elmagarmid AK, Bertino E, Saygin Y, Dasseni E. Association rule hiding. IEEE Transactions on Knowledge & Data Engineering. 2004(4):434–47.
11. Inokuchi A, Washio T, Motoda H, editors. An apriori–based algorithm for mining frequent substructures from graph data. European conference on principles of data mining and knowledge discovery; 2000: Springer.
12. Rennolls K, Al–Shawabkeh A. Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. Intelligent Data Analysis. 2008;12(2):147–63.
13. Lawrence CE, Reilly AA. An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences. Proteins: Structure, Function, and Bioinformatics. 1990;7(1):41–51.
14. Muyeba M, Khan MS, Coenen F, editors. Fuzzy weighted association rule mining with weighted support and confidence framework. Pacific–Asia Conference on Knowledge Discovery and Data Mining; 2008: Springer.
15. Wang H, Liu X, editors. The research of improved association rules mining apriori algorithm. 2011 eighth international conference on fuzzy systems and knowledge discovery (FSKD); 2011: IEEE.
16. Hornik K, Grün B, Hahsler M. arules–A computational environment for mining association rules and frequent item sets. Journal of Statistical Software. 2005;14(15):1–25.
17. Singh S, Yadav M, Gupta H. Finding the chances and prediction of cancer through Apriori algorithm with transaction reduction. Int J Adv Comput Res. 2012;2(2):23–8.
18. Tajbakhsh A, Rahmati M, Mirzaei A. Intrusion detection using fuzzy association rules. Applied Soft Computing. 2009;9(2):462–9.
19. Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, et al. Weka–a machine learning workbench for data mining. Data mining and knowledge discovery handbook: Springer; 2009. p. 1269–77.
20. Nasrollahzadeh D, Kamangar F, Aghcheli K, Sotoudeh M, Islami F, Abnet C, et al. Opium, tobacco, and alcohol use in relation to oesophageal squamous cell carcinoma in a high–risk area of Iran. British journal of cancer. 2008;98(11):1857.

21. Shakeri R, Kamangar F, Mohamadnejad M, Tabrizi R, Zamani F, Mohamadkhani A, et al. Opium use, cigarette smoking, and alcohol consumption in relation to pancreatic cancer. Medicine. 2016;95(28).

22. Kaz AM, Wong C–J, Varadan V, Willis JE, Chak A, Grady WM. Global DNA methylation patterns in Barrett's esophagus, dysplastic Barrett's, and esophageal adenocarcinoma are associated with BMI, gender, and tobacco use. Clinical epigenetics. 2016;8(1):111.

23. Grabowski JP, Richter R, Rittmeister H, Chekerov R, Woopen H, Sehouli J. Impact of Body Mass Index (BMI) on Chemotherapy–associated Toxicity in Ovarian Cancer Patients. A Pooled Analysis of the North–Eastern German Society of Gynecological Oncology (NOGGO) Databank on 1,213 Patients. Anticancer research. 2018;38(10):5853–8.

24. Wang W, Wang T, Feng X, Sun L. Incidence and risk factors of acute kidney injury after esophageal cancer surgery: A nested case–control study. International Journal of Surgery. 2017;39:11–5.

25. Padmanaban G, Venkateswar V, Rangarajan P. Haem as a multifunctional regulator. Trends in biochemical sciences. 1989;14(12):492–6.

26. Tjon–Kon–Fat L–A, Sol N, Wurdinger T, Nilsson RJA, editors. Platelet RNA in cancer diagnostics. Seminars in thrombosis and hemostasis; 2018: Thieme Medical Publishers.

27. Tuncer İ, Dülger H, Uygan İ, Öztürk M, Kotan Ç, Sekeroglu R. Comparison of serum cytokeratin–18, CEA and CA 19–9 levels in esophageal and gastric cancers. Eastern Journal of Medicine. 2004;9(2):72.

28. Bagaria B, Sood S, Sharma R, Lalwani S. Comparative study of CEA and CA19–9 in esophageal, gastric and colon cancers individually and in combination (ROC curve analysis). Cancer biology & medicine. 2013;10(3):148.

29. Chirica M, Resche–Rigon M, Zagdanski AM, Bruzzi M, Bouda D, Roland E, et al. Computed tomography evaluation of esophagogastric necrosis after caustic ingestion. Annals of surgery. 2016;264(1):107–13.