Interdisciplinary Journal of Acute Care
Volume 2, January/June 2021, Issue 1, pp. 42-49
Journal homepage: www.ijac.lums.ac.ir

Predicting the Mortality of Patients with Leukemia Using Artificial Intelligence

Maryam Mahdavi[1], Shiva Ariainezhad [1], Parastou Kordestani-Moghadam[*2], Khadijeh Heidarizadeh [3], Ahmad Eskandarizadeh[2], Rasool Mohamadi[4]

[1] Student Research Committee, School of Nursing and Midwifery, Lorestan University of Medical Sciences, Khorramabad, Iran

[2] *Critical care and Emergency Nursing, Faculty of Nursing & Midwifery, Lorestan University of Medical Sciences, Khorramabad, Iran*

[3] Lorestan University of Medical Sciences, Khorramabad, Iran

[4] Nutritional Health Research Center, School of Health and Nutrition, Lorestan University of Medical Sciences, Khorramabad, Iran

ABSTRACT

Several factors must be considered when predicting mortality in patients with hematological malignancies. These factors and characteristics complicate the ability of doctors' and nurses to predict the prognosis of these diseases. This study aimed to develop a mortality prediction model for leukemia patients using artificial intelligence and a nearest-neighbor genetic algorithm. This retrospective study used the medical records of 235 patients with leukemia at the Ahvaz Oncology Center from 2016 to 2019. To provide a mortality prediction model, a genetic algorithm and nearest neighbor were used. A genetic approach was employed to identify the determinants of mortality, and the nearest-neighbor technique was utilized to enhance model accuracy. Ultimately, the diagnostic power of the mortality prediction model was assessed using accuracy, sensitivity, and specificity criteria. The laboratory values and variables incorporated into the genetic algorithm revealed that mechanical ventilation, hemodialysis, neutropenia, and bone marrow transplantation significantly influenced the mortality rate of patients with leukemia. The diagnostic accuracy of the genetic algorithm introduced in this study was 77.4%, with a sensitivity of 78.2% and specificity of 82%. The results showed the artificial intelligence algorithm in predicting mortality in leukemia patients.

Keywords: Artificial intelligence; Leukemia; Genetic Algorithm; K-nearest neighbor algorithm; Oncology; Intensive Care Unit

Introduction

The intensive care of patients represents a particularly complex and sensitive medical issue [1]. Despite the presence of sophisticated monitoring equipment and substantial medical resources within an intensive care unit (ICU), the mortality rate remains notably elevated. Moreover, decisions within these units are rendered amidst an environment characterized by instability and considerable complexity [3]. Indeed, even the most experienced personnel in ICUs may need to correct their assessments of patients' mortality risk, relying solely on their personal experience [4].

Given that patients with leukemia are usually hospitalized in ICU/department oncology during the acute phase of the disease [5], it is imperative to have tools that can assess prognosis at various stages of the disease [6]. Cancers are among the most critical issues in the healthcare system, necessitating a

* *Corresponding* author: Parastou Kordestani-Moghadam
Faculty of Nursing & Midwifery, Lorestan University of Medical Sciences, Khorramabad, Iran; Email: kparastou@yahoo.com
DOI: 10.22087/ijac.2021.207425

substantial number of resources and treatment facilities. The introduction of prediction models that are highly accurate in predicting the survival of these patients can assist in the identification of the factors that influence survival and, ultimately, the reduction of treatment costs [7].

Numerous efforts have been undertaken in this field, resulting in the development of acute and chronic health physiology models, including acute physiology and chronic health evaluation II (APACHE II), the simplified acute physiology score (SAPS II), and the mortality probabilities model (MPM) [8]. All of these models have undergone several updates, and prolonged and frequent utilization reveals that they continue to exhibit limitations [9]. The outcomes are varied if used in populations for whom the mortality prediction model was not intended [10]. Therefore, selecting variables and predicting factors in statistical modeling and artificial intelligence models are crucial [11].

In many instances, selecting the appropriate variable enhances model performance while simultaneously optimizing the data-gathering procedure and reducing costs [12]. Boxet et al. proposed that neural networks are appropriate for simulating complex departments, such as ICUs. Dybowski et al. discovered that their neural network approach had superior diagnostic accuracy in predicting mortality compared to existing technologies [13].

It is important to note that the outcomes of existing techniques vary according to the complexity and nature of the disease, as well as geographical location [14]. These models have great sensitivity; however, their positive predictive value is relatively insignificant. By including essential and pertinent factors using artificial intelligence (AI) algorithms, their ability to predict mortality and clinical deterioration levels can be further enhanced [15].

The genetic algorithm seeks the optimal solution by selecting potential responses to the relevant issues. Based on the value of each response, this subset of potential responses constitutes the initial population. The selection procedure from the initial population is conducted for the second generation. Selecting more prudent responses will yield a more precise estimate of the ultimate result [16].

Chen et al. utilized nine variables from the SAPS II tool, integrating them with a genetic algorithm, achieving an area under the curve of 94%, compared to the SAPS II tool's area under the curve of 72% [10]. Numerous studies indicate that customized models exhibit superior performance in terms of both discrimination and calibration, thereby necessitating the construction of mortality prediction models with optimal efficacy [17]. This study aimed to propose a strategy utilizing significant factors to predict mortality and death rates among leukemia patients through an AI algorithm.

Materials and Methods

This was a Part of a retrospective study in an Intensive care oncology center in Baqaee teaching hospital in Khuzestan province. This center comprises 238 beds designated for leukemia patients, while its adult oncology ICU contains 14 beds. The data from the patient's paper medical records was documented from 21 June 2016 to 20 April 2019. This study abides by the ethical code IRLUMS.REC.1399.130 from Lorestan University of Medical Sciences.

Data were gathered from the medical records of individuals over the age of 16, all of whom had a confirmed diagnosis of leukemia based on pathology tests and the assessment of an oncology specialist. The details of the medical record were documented during the initial hospitalization. Medical records lacking comprehensive information and containing diagnoses not pertinent to leukemia treatment, as well as instances of patient mortality within the initial 24 hours of admission, were not documented in the oncology ICU. The collected data encompassed 1- demographic variables such as age and gender. 2- Variables about disease status, including the type of cancer, blood type, stage of cancer, treatment stage/phase, reasons for admission, and the patient's condition upon admission. 3- Laboratory values encompassing concentrations of calcium, phosphorus, lactate, and albumin within the serum, the neutrophil to lymphocyte ratio, hemoglobin concentration, erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), red cell distribution width (RDW), red blood cell (RBC) count, and bilirubin levels.

Four variables associated with the patient's admission status in the oncology ICU, including the necessity for receiving mechanical ventilation, the administration of vasopressor medications, the requirement for hemodialysis, and the provision of cardiopulmonary resuscitation within the initial 24 hours of admission to the oncology ICU. The descriptive variables were explained through their mean values and frequency distributions. The K-nearest neighbor algorithm was employed for the statistical analysis of the AI algorithm, with the significance of the variables assigned weights within the intervals 0 and 1. Furthermore, the criteria of accuracy, sensitivity, and specificity were utilized to assess the validity and performance of the proposed AI model.

Results

Among 513 medical records of leukemia patients admitted to the oncology ICU between 1 December 2015, and 31 April 2018, 235 patients met the criteria for inclusion in the study. A total of 148 individuals, comprising 63% men, exhibited an average age of 55.25 years, with a standard deviation of 18.78 years. A total of 112 individuals were discharged, representing 48% of the cases, while the mortality rate within the ICU remained at 51.9%.

Initially, 30 distinct variables or attributes were organized. The attributes encompass age, gender, occupation, duration of hospitalization, family history, type of leukemia, stages of the disease upon admission to the oncology ICU, reason for admission to the oncology ICU, and the duration of prior hospitalization, Stem cell transplantation, duration of hospitalization prior to transfer to the oncology ICU, kidney replacement therapy, blood type, treatment stage/phase, mechanical ventilation status, age at the time of cancer diagnosis, presence of concurrent cancers/malignancies, cardiopulmonary resuscitation prior to transfer to the oncology ICU, presence of neutropenia, requirement for vasopressor administration, lactate, phosphorus and calcium concentrations within the serum, hemoglobin level as well as the values of ESR, CRP, RDW, RBC.

In the next step, the genetic algorithm determines the optimal classification value for each variable by the evaluation function. Initially, the population was established, with the values of the chromosomes representing the weights of the features being randomly initialized. We would then conceptualize the structure of the chromosome, represented as binary strings. As illustrated in Figure 1, the length of each chromosome is expressed as 14*k bits, with k denoting the number of variables. The rationale for employing 14 bits is to encompass the range of values between one and zero.

Each variable was assigned 14 bits, which were then indexed to a numerical range spanning from 0 to 1. The conversion process started by transforming the 14-bit string into a base-ten numeral, which was subsequently divided by (1-2^14) to yield a value from 0 to 1. Upon the generation of the initial population, the system executed the KNN algorithm, utilizing the weight assigned to each chromosome. Subsequently, the performance of each chromosome was assessed through the evaluation function inherent to the genetic algorithm. The number of classification errors when applying the KNN algorithm was derived from the evaluation dataset. In the next step, based on the assessment of the evaluation function, the value of each feature or variable was calculated and assigned a weight. The figure below delineates the weight attributed to each variable (Figure 2).

1	0	1	…	1	1	0	0	…	1
0	0	1	…	0	1	0	0	…	1
1	0	0	…	1	0	0	1	…	1
F1	F2	F3	…	Fi	Fi+1	Fi+2	Fi+3	…	Fk*14

Figure 1 - Gene structure

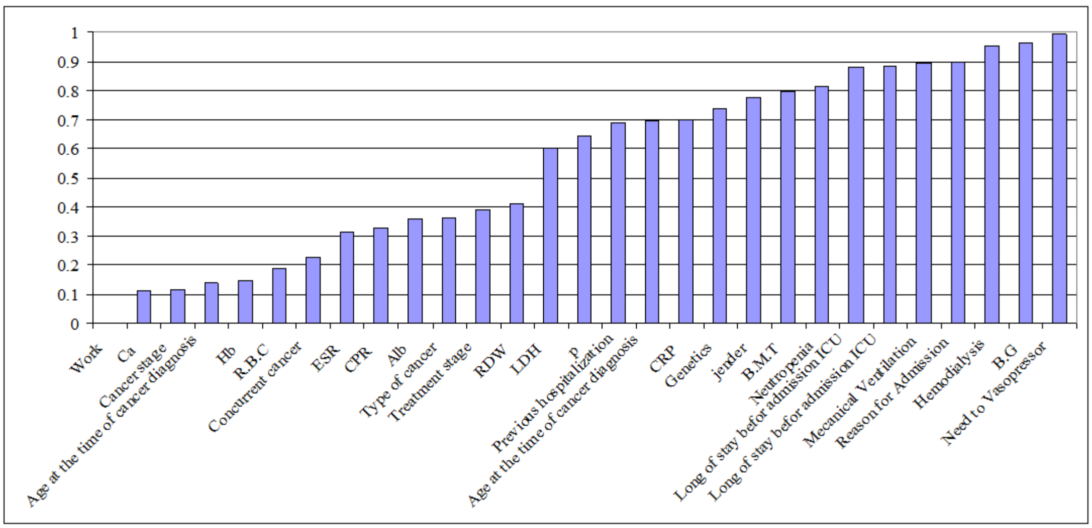

Figure 2- Examining the most effective weights in predicting mortality using a genetic algorithm

The figure clearly illustrates that within the variables incorporated in the genetic algorithm, variables such as the period of hospitalization, the reason for admission to the oncology ICU, kidney replacement therapy, blood type, requirement for vasopressor administration, mechanical ventilation, and presence of neutropenia exhibited greater weight and were correlated with increased mortality rates.

Upon acquiring the weight of each variable, the k-nearest neighbor algorithm was employed for recovery, dividing the data into three components. One component was designated for error measurement, while the remaining two were utilized for model fitting. In order to execute the algorithm, the dataset was partitioned into two training groups, comprising 165 patients, representing 70% of the data, and a test group of 70 patients, accounting for the remaining 30% of the data. This process was iteratively conducted until all three components were utilized for error assessment (Figure 3). The resultant errors were those associated with the estimated model. Figure 4 presents an illustration of the algorithm.

Upon determining the error derived from the model's test data, evaluated through accuracy, specificity, and sensitivity indicators, it was observed that in the absence of a genetic algorithm for variable weighting and under the assumption that all variables hold equal significance in predicting the mortality risk, the number of errors stood at 37. The diagnostic accuracy of the algorithm was recorded at 47.88, with specificity at 67.56 and sensitivity at 74.44. In the case involving weighting through the genetic algorithm, the number of errors of the variables amounted to 16, the diagnostic accuracy of the model reached 77.46%, the specificity was recorded at 81.25%, and the sensitivity stood at 72.22%.

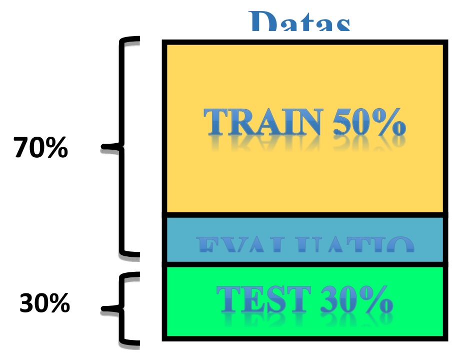

Figure 3- Dividing the dataset

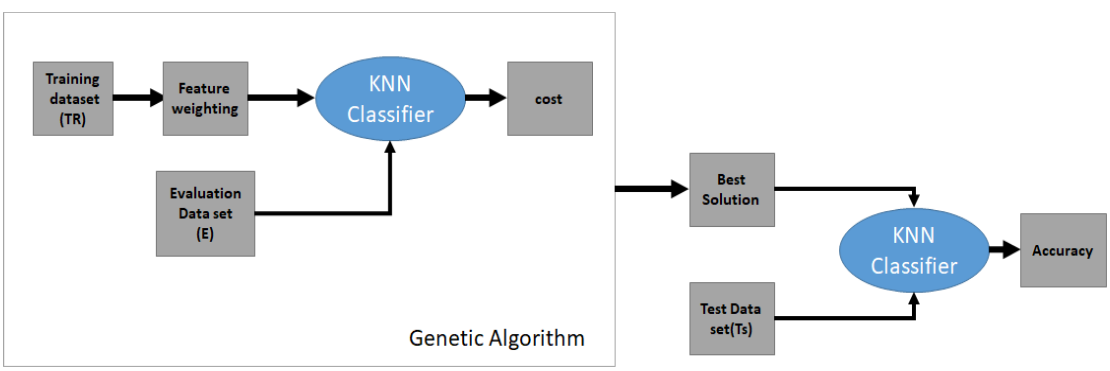

Figure 4- Combination of Genetic Algorithm and K-nearest Neighbor

Discussion

This study presents an algorithm for predicting disease severity and mortality among patients diagnosed with leukemia. The algorithm demonstrated a diagnostic accuracy of 77.4%, suggesting that enhancing the volume of data over time could further elevate its predictive precision. The variables requiring calculation are easily accessible. In the field of mortality prediction models within Intensive care departments, a persistent inquiry emerges regarding the identification of mortality and the factors associated with an elevated risk of mortality, as well as strategies to mitigate these risks.

On the other hand, the advancement and execution of death prediction models are not easy, and the pursuit of precise and scientific measurement presents considerable challenges, as certain variables might escape quantification, and some data necessitate continual updates within specific time intervals [18].

The results of Setareh et al.'s study on cancer patients showed that alongside AI methodologies, the selection of variables and their relevance to the study's purpose significantly influence the algorithm's accuracy [19]. Consequently, it is imperative to utilize pertinent characteristics and variables to ensure the model's precision.

The present investigation identified 31 variables previously linked to elevated mortality rates in other research projects.

The algorithm designed for predicting mortality incorporated variables such as mechanical ventilation, neutropenia, the necessity for dialysis, and the execution of bone marrow transplantation, all of which are assigned greater weight. Button's review study corroborated these findings, identifying the same variables as contributors to elevated mortality rates [20].

Furthermore, factors including the occurrence of leukaemia, the utilization of mechanical ventilation, and the administration of haemodialysis were documented as correlating with the heightened mortality rates observed in Vijenthira's study [21].

Therefore, it is prudent to consider these variables, as cancer patients admitted to intensive care face an elevated risk of mortality. Given that the risk of hospital mortality in patients with leukemia is two to three times greater than that of patients with other diagnoses, it is essential to refine and enhance mortality assessment systems for this specific patient population within intensive care units.

The proposed algorithm entails selecting features or variables, weighing these features, and developing a model to predict mortality guided by accuracy, sensitivity, and specificity indicators. In the suggested literature, it is advised that the dataset should be maximized for the application of genetic algorithms. In that regard, a notable study demonstrated that by augmenting the data from two thousand to 4804 entries, a prediction accuracy of 97% was achieved [22].

The implementation of artificial intelligence networks is significantly influenced by the specific algorithms and design choices employed, which affect the accuracy of predictions. Rahmani et al. determined the neural network weights by applying a genetic algorithm, successfully predicting the incidence of diabetes. The analysis encompassed a comprehensive dataset comprising 768 individuals, utilizing 8 distinct variables as inputs for the network, ultimately achieving a predictive accuracy of 92% regarding the incidence of diabetes [23].

In our study, the fold crosses method was employed to mitigate the bias of the variables, thereby minimizing errors and enhancing the model's fit. Conversely, minimizing variables is a crucial prerequisite in optimizing inputs within the genetic algorithm, significantly enhancing both the evaluation of the algorithm and the precision of its predictions. In Hosseini Teshnizi's study, introducing a significant variable, such as platelets, in predicting leukemia patients resulted in a neural network prediction accuracy of 78.9%, in contrast to the 50% accuracy achieved by the Cox regression method [24]. Shanbehzadeh's study aimed to enhance the precision of mortality predictions for COVID-19 patients by reducing 54 influential variables to 17. The nearest neighbor algorithm achieved a prediction accuracy of 93.7%, surpassing other machine learning algorithms [25]. A notable strength of their research lies in the comparative analysis of prediction accuracy against other artificial intelligence algorithms, including random forest and support vector machine, which represents a limitation in the current study.

The current study is constrained by a limited dataset and the fact that it is based on a single-center database. It may prove advantageous for the broader applicability of data and findings. Meanwhile, the presence of specific vacant and incomplete information fields, which resulted in data deletion, contributed to the diminished volume of data. To enhance the precision of mortality predictions for leukemia patients, it is recommended that studies that incorporate a larger dataset be undertaken and alternative artificial intelligence algorithms explored.

Conclusion

This paper introduced an algorithm designed to predict the mortality rates of patients diagnosed with leukemia. Our proposed methodology employed the genetic algorithm with the nearest-neighbor approach. The significance of laboratory values and various factors, including disease status upon admission to the oncology ICU, utilization of mechanical ventilation, administration of vasoconstrictor medications, the reason for admission to the oncology ICU, duration of hospitalization oncology ICU, and the application of hemodialysis, underscores the relevance of these variables and the algorithm in question. The genetic factors were identified, followed by the assessment of various variables based on their significance and influence on mortality. They ultimately employ the k-nearest neighbor algorithm, which allows for identifying the model that provides the most precise and reliable predictions regarding mortality. Accurate prediction of mortality presents an opportunity for improved patient outcomes and facilitates more effective management of intensive care unit resources. The results obtained can be utilized to discern the most significant variables influencing the survival of patients with leukemia, and the application software can be employed. It will also contribute to a streamlined system for healthcare professionals to efficiently assess the results of a patient's hospitalization in the specialized department by inputting essential variables, facilitating a more reasoned clinical evaluation, and timely decision-making concerning treatment protocols.

Acknowledgments

This paper is the result of a research project with the research code 1484 approved by the vice-chancellor for the research and technology department of Lorestan University of Medical Sciences and a part of the master's thesis of the critical care nursing field of this university in collaboration with vice-chancellor for research and technology department of Jundishapur university of Ahvaz.

Conflict of interests

The authors have no financial interest related to this article

References

1. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. The Lancet Respiratory Medicine. 2015 Jan 1;3(1):42-52.

2. Soleimani M, li Imani A, Bagheri S, Jafarpour H, Tangestani F. Evaluation of patient mortality in intensive care units using the APACHE IIscoring system. Koomesh. 2014 Jan 1;15(3):289-94.

3. Bruyneel A, Tack J, Droguet M, Maes J, Wittebole X, Miranda DR, Di Pierdomenico L. Measuring the nursing workload in intensive care with the Nursing Activities Score (NAS): A prospective study in 16 hospitals in Belgium. Journal of critical care. 2019 Dec 1;54:205-11.

4. Godinjak A, Iglica A, Rama A, Tančica I, Jusufović S, Ajanović A, Kukuljac A. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit. Mortality. 2016;29:10.

5. Maghsoudi B, Tabei SH, Zand F, Tabatabaee H, Akbarzadeh A. A model for decision making for intensive care unit admission in source limited hospitals. Iranian Red Crescent Medical Journal. 2014 Oct 5;16(10):e15497.

6. Liu J, Cheng Q, Yang Q, Li X, Shen X, Zhang L, Liu Z, Khoshnood K. Prognosis-related factors in intensive care unit (ICU) patients with hematological malignancies: a retrospective cohort analysis in a Chinese population. Hematology. 2015 Oct 1;20(9):494-503.

7. Fooladi S, Farsi H, Mohamadzadeh S. Detection of Malignancy Degree in Prostate and Breast Cancers by Using Deep Neural Network. Medical journal of Mashhad University of medical sciences. 2018 Nov 22;61(5):1178-87.

8. Hosseini M, Ramazani J. Evaluation of Acute Physiology and Chronic Health Evaluation II and sequential organ failure assessment scoring systems for prognostication of outcomes among Intensive Care Unit's patients. Saudi journal of anaesthesia. 2016 Apr 1;10(2):168-73.

9. Ayoubian A, Navid M, Moazam E, Hoseinpourfard M, Izadi M. Evaluation of standards in intensive care units in Isfahan hospitals, Iran. Iran J Milit Med. 2013;2013:14.

10. Chan CL, Ting HW. Constructing a novel mortality prediction model with Bayes theorem and genetic algorithm. Expert Systems with Applications. 2011 Jul 1;38(7):7924-8.

11. Tahamtan RA, Esmaeili MH, Ghaemian A, Esmaeili J. Application of artificial neural network for assessing coronary artery disease. J Maz Univ Med Sci. 2012;22(86):8-17.

12. Nielsen AB, Thorsen-Meyer HC, Belling K, Nielsen AP, Thomas CE, Chmura PJ, Lademann M, Moseley PL, Heimann M, Dybdahl L, Spangsege L. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. The Lancet Digital Health. 2019 Jun 1;1(2):e78-89.

13. Xie J, Su B, Li C, Lin K, Li H, Hu Y, Kong G. A review of modeling methods for predicting in-hospital mortality of patients in intensive care unit. Journal of Emergency and Critical Care Medicine. 2017 Aug 21;1(8).

14. Maqsood S, Badar F, Hameed A. Characteristics and outcomes of patients with hematological malignancies admitted for intensive care-a single centre experience. Asian Pacific journal of cancer prevention: APJCP. 2017;18(7):1833.

15. Chen CL, Wang ST, Cheng WC, Chen CY, Chen WC, Lin YC, Wu BR, Liao WC. Outcomes and Prognostic Factors in Patients with Hematologic Malignancies in the Intensive Care unit: A Single-Center Cohort Study of 233 Cases in Taiwan.

16. Momeny M, Latif AM, Agha Sarram M, Hajmirzazade K, Gharravi S, Seyed Mahammad N. Diabetes Prediction by Optimizing the Nearest Neighbor Algorithm Using Genetic Algorithm. Journal of Health and Biomedical Informatics. 2019 Jun 10;6(1):12-23.

17. Verplancke T, Van Looy S, Benoit D, Vansteelandt S, Depuydt P, De Turck F, Decruyenaere J. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. BMC medical informatics and decision making. 2008 Dec;8:1-8.

18. Asgari P, Atashi A, Meraji M, Miri M. The comparison of selected data-mining techniques in ICU mortality risk prediction in imam hossein hospital. Journal of Modern Medical Information Sciences. 2019 Dec 10;5(2):59-67.

19. Setareh S, Zahiri Esfahani M, Zare Bandamiri M, Raeesi A, Abbasi R. Using data mining for survival prediction in patients with colon cancer. Iranian Journal of Epidemiology. 2018;14(1):19-29.

20. Button E, Chan RJ, Chambers S, Butler J, Yates P. A systematic review of prognostic factors at the end of life for people with a hematological malignancy. BMC cancer. 2017 Dec;17:1-5.

21. Vijenthira A, Chiu N, Jacobson D, Freedman Z, Cheung MC, Goddard S, Fowler R, Buckstein R. Predictors of intensive care unit admission in patients with hematologic malignancy. Scientific reports. 2020 Dec 3;10(1):21145.

22. Nabovati E, Azizi AA, Abbasi E, Vakili-Arki H, Zarei J, Razavi AR. Using data mining to predict outcome in burn patients: a comparison between several algorithms. Health information management. 2014 Jan 21;10(6):789-99.

23. Mousavi S, Atabakhsh A, Moghadamniya M, Kazemnezhad Leili E, Rouhi Balasi L. Performance obstacle related to technology and equipment from employed nurses' viewpoint working in intensive care units. Journal of Holistic Nursing And Midwifery. 2015 Jun 10;25(2):65-72.

24. Hosseini Teshnizi S, Tazhibi M, Tavasoli Farahi M. Comparison of Cox regression and Artificial Neural Network models in prediction of survival in acute leukemia patients. Scientific Journal of Iranian Blood Transfusion Organization. 2013 Jun 1;10(2).

25. Shanbehzadeh M, Valinejadi A, Afrah R, KazemiArpanahi H, Orooji A, Kaffashian M. Comparison of machine-learning algorithms efficiency to build a predictive model for mortality risk in COVID-19 hospitalized patients. Koomesh. 2022 Feb 15;24(1).