Journal homepage: https://medicinskiglasnik.ba

Med Glas (Zenica) 2026;23(1)
https://doi.org/10.17392/1970-23-01

Med (Zenica)
Glas

ORIGINAL ARTICLE

Bayesian model prediction for lung cancer survival based on demographic and laboratory results: a retrospective analysis

Islam Bani Mohammad[1], Muayyad M. Ahmad[2]

[1]Department of Nursing, Faculty of Nursing, Al-Balqa Applied University, [2] Clinical Nursing Department, School of Nursing, University of Jordan, Amman; Jordan

ABSTRACT

Aim To examine the likelihood of predicting lung cancer survival versus death using Bayesian model based on demographic and laboratory data.

Methods A predictive design using electronic health records from 2012 to 2023 was implemented. IBM SPSS Statistics version 29.0 was used for data descriptive analysis and prediction models were built using SPSS Modeler version 18.0. Among the eight generated models, the Bayesian model demonstrated the highest accuracy (71.9%) and the best area under the curve (AUC) at 80.304, showcasing its superior predictive performance for lung cancer outcome.

Results A total of 1,843 patients without missing values were used. Males constituted 64.2 % of total sample. About 70 % of the patients were aged between 46 and 99 years. The Bayesian network identified seven key predictors for determining patient outcome (survival versus death). Among these, age was found as the most significant predictor of survival outcome.

Conclusion The Bayesian network outperformed other models in predicting lung cancer survival versus death probability. The integration of routine laboratory testing and demographic data in the machine learning model can help in the prediction of lung cancer survival versus death.

Keywords: death, machine learning, prediction, survival

INTRODUCTION

Lung cancer has become the most frequent disease globally and continues to be a major public health concern (1). In 2022, lung cancer has accounted for about 2.5 million new cases globally (2). Furthermore, lung cancer is a leading cause of illness worldwide (1). In the United States of America (USA), there is an increase in the incidence of lung cancer, which is accounted for 33 cases per 100,000 persons in 2022 (2). Lung cancer is considered to be a leading cause of death in the US with a mortality rate of 18.9 cases per 100,000 persons in 2022 (2). In Jordan, lung cancer is the fourth most frequent type of cancer and the average age at lung cancer diagnosis is 63.8 years (3).

Lung cancer development is significantly affected by demographic factors such as age and gender (4). Age is considered an important factor that affects the risk of developing lung cancer because this cancer becomes more common in older patients (5). Although lung cancer is uncommon in patients under 40 years of age, health care providers should not rule it out in younger patients (4). Lung cancer ranks first in men and second in women for both incidence and mortality (2). Regardless of the age at which it is recognized, men are more likely to have this malignancy than women (6). For both genders, lung cancer is a common cause of cancer-related death (7). In 2022, mortality rate among men was 18% higher than in women with lung cancer (2).

Laboratory tests are considered to be an important predictor of lung cancer such as white blood count (WBC), as higher WBC values were associated with lung cancer (8,9). Furthermore, low red blood cells (RBC) and haemoglobin (Hb) values were associated with lung cancer (10). Higher serum creatinine variation during hospitalization was associated with an increased risk of one-year mortality among patients with lung cancer (11). Manual diagnosis techniques for lung cancer are imaging tests, needle biopsy, sputum cytology, bronchoscopy, and lab tests (1).

In a recent study conducted in 2024, researchers used regression approaches, including the Cox proportional hazards (CPH) model to predict lung cancer survival (12). However, missing covariate data in clinical datasets is a significant barrier for

---

***Corresponding author:** Islam Bani Mohammad
Department of Nursing, Faculty of Nursing, Al-Balqa Applied University
PO Box 19117 Al-Salt, Jordan
E-mail: eslam.banimohammad@bau.edu.jo
ORCID ID: https://orcid.org/0000-0003-3569-7875

| Submitted: 04. Feb 2025. Revised: 18. Jul 2025, Accepted: 04. Aug 2025.

regression-based models (12). Incomplete predictors cannot be included in regression-based predictive models. Machine learning (ML) algorithms imply that missing values occur randomly. However, if there is a specific missing pattern, such as a non-random process, ML algorithms may result in incorrect conclusions. Due to restrictions, commonly used imputation methods may not be accurate enough for datasets with missing values. To address the challenge of unavailable data in clinical risk assessment, we applied Bayesian network, which successfully manages data by creating a complex network structure with multiple components (12). The Bayesian network uses evidence to determine the posterior probability distribution of query variables (13). Furthermore, Bayesian Network is commonly utilized for predicting illness onset and progression, as well as evaluates treatment effectiveness. In cancer treatment, Bayesian network was used to estimate patient survival and evaluate treatment options, guiding the development of optimal strategies (12).

The aim of this study was to examine the likelihood of predicting lung cancer survival versus death using Bayesian network based on demographic and laboratory data.

PATIENTS AND METHODS

Patients and study design

Predictive and retrospective designs were used in this study. This study was conducted at Jordan University Hospital over the course of six months, from January 2024 to July 2024. Parients' information was managed with confidence. Each record was assigned an ID, allowing for anonymous processing of patient records. The extracted data were kept in a password-protected file at a secured computer in the researcher's office. The study data were only accessible to the researchers.

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki, Good Clinical Practices, and the Indian Council of Medical Research (ICMR). We included the electronic health records of patients who were diagnosed with lung cancer. The inclusion criteria specified patients with lung cancer regardless of the age. We excluded patients with any cancer other than lung cancer. The Jordan University Hospital Ethics Committee approved the study protocol prior to its initiation.

Methods

During the study period from January 2024 to July 2024, data on patients with lung cancer between 2012 and 2023 were obtained from Jordan University Hospital using a part of their electronic health records system (EHS). Extracting the data took four months due to the complexity of the organizational processes and the electronic health system. This dataset contains 1,843 records and eight variables. The selection of these variables was based on previous literature indicating that the choice of variables included in a study or model was guided by findings from previous publications. This approach is common in research to ensure that relevant and important factors are considered for predicting the outcome of lung cancer. Each record represents a particular patient within the database. These variables provide socio-demographic and lung cancer specific information. The variables that were requested included the stage of lung cancer, age, gender, RBC, Hb, WBC, and creatinine values, governorate, smoking, family history of cancer and patient outcome (survival versus death). The values such as the stage of lung cancer, smoking, and family history of cancer were excluded because they were not available in the electronic health records.

Statistical analysis

To build the Bayesian network, the following steps were performed as data preparation: checking for missing data cleaning by eliminating noisy data from the dataset and removing the duplicate and inconsistent data. All laboratory data were standardized using one international unit for analysis. All data were originally on five Excel sheets exported to SPSS files. Then, the files were merged into one data sheet by matching the cases with the ID numbers. Descriptive statistics were conducted using Statistical Package for Social Sciences, version 29.1 (14). IBM SPSS Statistics (Version 29.0) and IBM SPSS Modeler (Version 18.0) were employed for data manipulation, statistical analysis, and visualization in this study. These software applications enable effective data presentation for statistical and predictive analysis, as well as data management for descriptive and predictive modelling (15). Descriptive modelling found the major predictive variables for predicting lung cancer survival versus death. Predictive modelling was chosen based on accuracy and Area Under Curve (AUC) factors to create an appropriate model.

For applying data mining step, SPSS Modeler version 18.0 was used (15). Seventy and 30 percent of database records were selected as training and testing data, respectively. Training data were utilized to develop a predictive model, whereas testing data were used to evaluate the model's performance (16). The primary criteria for selecting the most effective AI model were the overall accuracy and AUC (17). The accuracy is the percentage of all the used datasets that are properly predicted out of all the instances (18). The AUC represents the performance metrics that determine the predictive ability of machine learning (ML) models and it measures the overall performance of the model (19).

The Bayesian network is a probabilistic graphical model that describes variables and their relationships using an acyclic graph with a directed structure (13). They are particularly useful in medical applications such as predicting lung cancer because they can model complex relationships between risk factors and symptoms, incorporating uncertainty and prior knowledge effectively. A probabilistic graphical model for information prediction about an uncertain area is called the Bayesian networks. According to Kim and Lee (13) this model has nodes that represent random variables and edges that reflect the conditional probability for the related random variables. Furthermore, the Bayesian network was selected due to several reasons other than achieving the highest performance, such as the likelihood of survival versus death calculated using the existing variables and their conditional likelihood correlations, effectively addressing the difficulty of risk assessment with partial knowledge (13). The Bayesian network supports expert knowledge to determine the conditional independence of risk variables. Additionally, they provide an intuitive visual representation of the relationships between survival and death parameters (20). Because the model is based on easily accessible covariates from daily clinical practice, it can be utilized as a predictive tool for particular lung cancer patients, assisting doctors in decision-making process (13). Bayesian networks use a probabilistic framework to generate predictions, enabling outcomes to be interpreted in terms of probabilities.

RESULTS

The study included 1,843 patients between 2012 and 2023 of which 1,8184 (64%) were males. Age groups were categorized as children (2-18 years, 3.4%), young adults (19-45 years, 26.2%), and older adults (46-99 years, 70.4%). Laboratory assessments included haemoglobin (Hb), red blood cell (RBC) count, white blood cell (WBC) count, and creatinine serum levels. Below-normal Hb and WBC levels were observed in 628 (34.0%) and 569 (31.0%) of patients, respectively. Geographically, 1,084 (59%) were from Middle Governorates, followed by Northeast 423 (23.0%), South 175 (9.5%), and North Governorates 161 (8.7%). Notably, 1,479 (80%) of the patients belonged to the survived class (Table 1).

Table 1. Characteristics of the study sample (N=1,843).

Variable	No of patients (%)
Patient outcome	
Dead	364 (19.8)
Survived	1,479 (80.2)
Gender	
Male	1,184 (64.2)
Female	659 (35.8)
Age group	
Children (2-18)	63 (3.4)
Young adult (19-45)	482 (26.2)
Old adult (46-99)	1,298 (70.4)
WBC	
Normal	1,274 (69.1)
Above normal	569 (30.9)
Hb	
Normal	1,215 (65.9)
Below normal	628 (34.1)
Creatinine	
Normal	1,093 (59.3)
Abnormal	750 (40.7)
RBC	
Normal	1,133 (61.5)
Below normal	710 (38.5)
Governorates	
Middle	1,084 (58.8)
Northeast	423 (23.0)
South	175 (9.5)
North	161 (8.7)

Hb, haemoglobin; RBC, red blood cell; WBC, white blood cell

Among the eight generated models, the Bayesian network was the most effective one, achieving the highest overall accuracy score (80.304%) and the highest AUC score (0.719) (Table 2) followed by logistic regression (80.304%, AUC = 0.704), CHAID model (80.250, AUC = 0.692), neural network model (80.250, AUC = 0.685). Then, C5 model, Quest model, and C&R Tree model had the same performance (80.250, AUC = 0.5). Decision list model (70.912, AUC= 0.584) had the lowest performance.

Table 2. Characteristics of the study's models

Model	Overall accuracy (%)	Area under the curve
Bayesian network 1	80.304	0.719
Logistic regression 1	80.304	0.704
CHAID 1	80.250	0.692
Neural net 1	80.250	0.685
C5 1	80.250	0.5
Quest 1	80.250	0.5
C&R tree 1	80.250	0.5
Decision list 1	40.912	0.584

This study used the Bayesian network for binary classification to distinguish between deceased and surviving patients. The patient outcome node (survival versus death) in the graph represented a random variable, with death encoded as 0 and survival as 1. This encoding enabled the identification of probabilistic dependencies among discrete variables, facilitating the analysis of relationships within the dataset. The Bayesian network model contains eight nodes, including the parent node. It has 13 edges, representing the parameters that define the relationships between these nodes (Figure 1). Each node in the network represents a random variable of interest. For outcome (survival versus death) prediction of lung cancer, the predictor variables were age, gender, governorate, RBC, Hb, WBC, and creatinine values. The parent node is a node that has direct edges that lead to one or more child nodes (13). Directed edges between the nodes reflect the probability link among the variables in the network (20).

The Bayesian network found seven important predictors to predict the outcome of lung cancer (survival versus death). The age predictor's importance was 0.26, which is the most important predictor followed by RBC predictor's importance of 0.19, Hb predictor's importance was 0.16, gender predictor's importance 0.13, governorate predictor's importance 0.11, WBC predictor's importance 0.10, and creatinine predictor's importance was 0.06, which is the least important predictor (Figure 2).

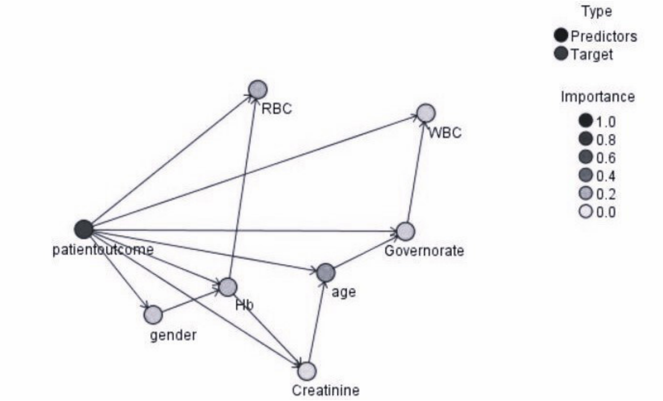

Figure 1. The structure of Bayesian network model
Hb, haemoglobin; RBC, red blood cell; WBC, white blood cell

For patients who had below normal RBC and Hb values, the conditional probability of death was 85%, while for patients who had normal RBC and Hb values, it was 80%. Survival probabilities were greatest among individuals with normal RBC and Hb values (0.81%). Survival probabilities among patients with normal creatinine value (0.69%) were higher

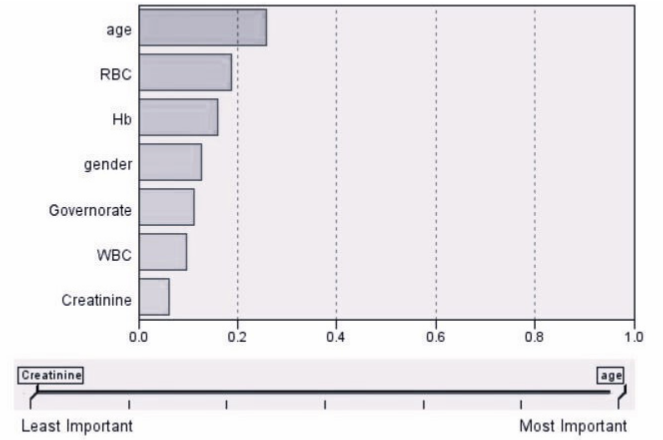

Figure 2. Predictors' importance for the outcome of lung cancer
Hb, haemoglobin; RBC, red blood cell; WBC, white blood cell

than in patients with abnormal creatinine value (0.31%). Furthermore, survival probabilities among patients who live in the North Governorates and had above normal WBC (0.14%) were lower than that in patients with normal WBC (0.86%). The older adults who live in the Middle Governorates had the highest probability of death (0.75%), while older adults residing in the North Governorates demonstrated the lowest recorded probability of death at just 0.04%. However, when WBC values were normal, the highest death probability was in the North Governorates (0.83%) and the Northeast Governorates (0.70%). Males with normal Hb levels had a higher survival probability (0.78%) compared to those with below-normal Hb levels (0.22%). Additionally, the probability of death among males was 0.42% higher than that of females (Table 3).

DISCUSSION

Machine learning is currently one of the most popular approaches to create a prediction model. It has been widely employed in medical science to assist healthcare professionals with prognosis, diagnosis, and other factor analysis. Machine learning algorithms are capable of managing massive amount of data and create prediction models for illness development, risk factors, and management, making them very useful. Using ML algorithms to analyse data can improve patient outcomes, identify needs, and enhance quality of life (21). In this paper, Bayesian network was used to predict survival versus death of lung cancer based on risk factors. Furthermore, Bayesian network is an extremely advanced modelling technique with numerous parameters (22). It is commonly used to forecast illness onset and progression, as well as evaluate treatment effectiveness, and for risk prediction of lung cancer (12).

Many researches have assessed the usefulness of machine learning algorithms in predicting the risk of cancer, but few of them used Bayesian network to predict the survival of lung cancer (12,20,23). However, the findings of this study contribute to clarifying the effectiveness of ML algorithms for the prediction of survival versus death likelihood among patients with lung cancer. The study model shows correlations between the outcome variable (whether the individual survived or died) and the seven predictors. This demonstrates how changes in the parent nodes' statuses might alter the probabilities of outcomes in the child nodes.

The Bayesian network provides a thorough description of the interactions and effects of a number of variables on patient outcomes (12). A recent study using a Bayesian network model to

Table 3. The Bayesian network model's determination of the probabilities of survival versus death and predictors

Parents Nodes	**Conditional probabilities of RBC**		**Conditional probabilities of creatinine**	
Dead or Survived/ Hb	**Normal**	**Below normal**	**Normal**	**Abnormal**
Survival/normal	0.81	0.19	0.69	0.31
Death/normal	0.80	0.20	0.57	0.43
Survived/below normal	0.28	0.72	0.43	0.57
Death/below normal	0.15	0.85	0.46	0.54

	Conditional probabilities of age (years)			**Conditional probabilities of WBC**	
Dead or Survive/ Governorate*	**Children** (2-18)	**Young adult** (19-45)	**Old adult** (46-99)	**Normal**	**Above normal**
Survived /1	0.46	0.44	0.61	0.68	0.32
Dead/1	0.57	0.57	0.75	0.50	0.50
Survived/2	0.14	0.30	0.23	0.67	0.24
Dead/2	0.14	0.20	0.15	0.70	0.30
Survived/3	0.16	0.15	0.07	0.86	0.14
Dead/3	0.14	0.09	0.04	0.83	0.17
Survived/4	0.23	0.12	0.09	0.77	0.32
Dead/4	0.14	0.15	0.06	0.56	0.44

	Conditional probabilities of Hb		**Conditional probabilities of gender**	
Dead or survived/ Gender	**Normal**	**Below normal**	**Male**	**Female**
Survived/male	0.78	0.22	0.62	
Death/male	0.60	0.40	0.71	
Survived/female	0.53	0.47		0.38
Death/female	0.40	0.60		0.29

*1, middle, 2, Northeast, 3, North, 4, South; Hb, haemoglobin; RBC, red blood cell; WBC, white blood cell evaluate lung cancer survival prognosis showed that internal validation demonstrated robust predictive performance, with the model achieving strong discrimination and calibration metrics, reflected by an AUC of 0.896 (12). The Bayesian network model had the highest overall accuracy score (80.304%) and the highest AUC score (0.719) in our study, whereas Bayesian networks had a low AUC score (0.614), which is utilized to predict survival among individuals diagnosed with lung cancer (23). The AUC decreases as the percentage of missing values in the validated dataset increases (12). This could mean that Bayesian network requires sufficient and accurate data for learning reliable conditional probabilities. Furthermore, the predictive study aimed to assess the potential value of blood inflammation in the diagnosis of lung cancer using a machine learning algorithm. It showed that the naive Bayes algorithm outperformed the other five ML systems in predicting adult lung cancer, with an AUC of 0.84 and an accuracy of 0.87 (10).

Age was the most important predictor for lung cancer, with incidence and mortality probability rising dramatically with age. Older patients (46-99 years) have the highest chance of mortality, which is consistent with evidence suggesting that increasing age is associated with significantly higher death outcomes in lung cancer (2). Researchers found that age was an important predictive factor for predicting lung cancer survival and prognosis using Bayesian network, as the median age of the patients with lung cancer was 63 years and it ranged from 22 to 92 years (12).

Laboratory tests can determine the survival and death probability of lung cancer. In our study, below normal red blood cell (RBC) and haemoglobin (Hb) were considered as important predictors for low survival probability among patients with lung cancer. Furthermore, a cross-sectional study found that Hb and RBC were significantly lower in lung cancer compared with healthy individuals (10).

Gender is the fourth important predictor. The study's findings are consistent with other trends found in the literature on lung cancer. According to global data, males are somewhat more likely to develop lung cancer than females, which is consistent with our study's higher male prevalence (2). The study's probability of survival and death was higher among males than females, which is consistent with global trends, particularly in high-income nations where early identification and advanced therapies are more widely available (2). A study of survival rates by gender found that treatment, healthcare systems, and comorbidities are likely to play significant effects (24).

In our study, the death probability was high in the Northeast Governorates (0.70), which is consistent with a Jordanian study revealing that Al- Mafraq city (northeast Jordan) has a high probability of lung cancer, which is possibly related to gaseous highly radioactive Radon element in Al- Mafraq city (25).

Other laboratory test that have been frequently obtained from patients was WBC, which is the sixth important predictor in this study. Elevated WBC is an alarming risk associated with lung cancer (8). In our study, above normal WBC was associated with poor survival rates and high death rate. Similarly, a predictive study aimed to create a prediction model that predicts future lung cancer diagnoses based on the available laboratory and clinical such as WBC, found that

WBC is an essential factor for predicting lung cancer death and has the ability to help prevent deaths from lung cancer through early identification utilizing a machine learning algorithm (9).

Creatinine serum value is considered as the least important predictor. Furthermore, abnormal serum creatinine is a predictor for low survival probability in our study. Creatinine could be a good biomarker for detecting early lung cancer (26). Researchers found that serum creatinine fluctuation increased the chance of death in patients with lung cancer (11).

The use of Bayesian network in lung cancer prediction incorporates previous knowledge and beliefs about causal links and conditional probabilities, as well as handling incomplete or uncertain data. The graphical structure of Bayesian network makes it easier to understand and convey variable interactions than more sophisticated ML models.

One of our study's strengths was that the Bayesian network efficiently overcomes the issues of missing data in forecasting patients' outcomes, while retaining high accuracy in prediction. Furthermore, the ML method was used in the study to analyse a large dataset. This study was especially unique in that it examined all patients with lung cancer, regardless of age, rather than merely the elderly.

Our research has a few limitations. While the aim was to include key variables relevant to predicting patient outcomes, such as lung cancer stage and medical imaging, these were excluded due to unavailability in the electronic health system. Additionally, the dataset's class imbalance, with most patients classified as survivors, may influence the results.

Prospective research is required to verify the use of the Bayesian network in future investigations. Further research is needed to compare the effectiveness of our ML model to the most used models for predicting lung cancer outcomes. Researchers often utilize Bayesian network in conjunction with medical data to investigate the links between risk variables and disease outcomes. Studies may focus on network structure refinement, inference algorithm improvement, and prediction validation against clinical data.

Healthcare providers can use the Bayesian network to evaluate survival versus death probabilities and guide hospital-based lung cancer treatment decisions, fostering personalized treatment approaches using routine demographic and laboratory data. In lung cancer survival prediction, the Bayesian network identifies the most influential predictors, such as age, gender, governorate, RBC, Hb, WBC, and creatinine values. This enhances model interpretability and aligns with clinician expertise, validating its practical utility.

Nursing practice implications involve utilizing predictive models to improve information systems, support clinical decisions, streamline documentation, and estimate death probabilities. By incorporating routinely available data into health records, the Bayesian network can deliver precise risk predictions. This AI approach elucidates complex relationships between lung cancer predictors, including demographics and laboratory results. Given its simplicity and interpretability compared to other ML methods, the Bayesian network is increasingly adopted in healthcare and can be seamlessly integrated into clinical workflows.

The Bayesian network is a robust tool for predicting lung cancer survival versus mortality. Its ability to manage uncertainty and integrate prior knowledge makes it highly beneficial for medical decision support systems. This study has examined the likelihood of predicting lung cancer survival versus death using Bayesian network based on demographic and laboratory data. Abnormal laboratory tests and older age were considered as important predictors for low survival probability. Incorporating demographic information and routine laboratory tests such as RBC, WBC counts, Hb, and creatinine levels enhance the model's capacity to predict survival outcomes, supporting improved clinical decision-making for lung cancer care.

AUTHOR CONTRIBUTIONS STATEMENT

Conceptualization, IM, MA; Methodology, IM, MA; data collection: IM; analysis and interpretation, MA; manuscript writing, IM, MA; approval of the final article, IM, MA; accountability for all aspects of the work, IM, MA.

FUNDING

No specific funding was received for this study.

TRANSPARENCY DECLARATION

Competing interests: None to declare.

REFERENCES

1. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: Current status and future trends. Nat Rev Clin Oncol 2023; 20 (9):624-39.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74 (3):229-63.
3. MOH. Cancer incidence in jordan – 2018. https://moh.gov.jo/ebv4.0/root_storage/ar/eb_list_page/%D8%A7%D9%84%D8%AA%D9%82%D8%B1%D9%8A%D8%B1_%D8%A7%D9%84%D8%B3%D9%86%D9%88%D9%8A_%D9%84%D8%A7%D8%B5%D8%A7%D8%A8%D8%A7%D8%AA_%D8%A7%D9%84%D8%B3%D8%B1%D8%B7%D8%A7%D9%86_%D8%A7%D9%84%D9%85%D8%B3%D8%AC%D9%84%D8%A9_%D9%81%D9%8A_%D8%A7%D9%84%D8%A7%D8%B1%D8%AF%D9%86_%D9%84%D8%B9%D8%A7%D9%85_2018.pdf (accssed: 1, Jun, 2025).
4. Bradley SH, Kennedy MP, Neal RD. Recognising lung cancer in primary care. Adv Ther 2019; 36 (1):19-30.
5. Candal-Pedreira C, Ruano-Ravina A, de Juan VC, Cobo M, Trigo JM, Rodríguez-Abreu D, et al. A Comparison of clinical and genetic characteristics between younger and older lung cancer patients. Arch Bronconeumol 2024; 60 (2):88-94.
6. Mazouzi C, Kerboua E, Benyahia R, Ferrah N, Aftis R. Lung cancer:"Epidemiology and risk factors". Int J Med Phar Drug Re 2024; 8:1.
7. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Ann Glob Health 2019; 85 (1).
8. Wong JY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, et al. White blood cell count and risk of incident lung cancer in the uk biobank. JNCI Cancer Spectr 2020; 4 (2):pkz102.
9. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine learning for early lung cancer identification using routine clinical and laboratory data. Am J Respir Crit Care Med 2021; 204 (4):445-53.
10. Wang Q, Liang T, Li Y, Liu X. Machine learning for prediction of non-small cell lung cancer based on inflammatory and nutritional indicators in adults: A cross-sectional study. Cancer Manag Res 2024: 527-35.
11. Li L, Yang H, Zhang Y, Liu J, Li S, Wang L, et al. Association of serum creatinine variability and risk of 1-year mortality among patients with cancer. Res Sq 2024; https://doi.org/10.21203/rs.3.rs-4639262/v1.
12. Zhong L, Yang F, Sun S, Wang L, Yu H, Nie X, et al. Predicting lung cancer survival prognosis based on the conditional survival bayesian network. BMC Med Res Methodol 2024; 24 (1):16.
13. Kim A, Lee D. Dynamic bayesian network-based situational awareness and course of action decision-making support model. Expert Syst Appl 2024; 252:124093.
14. IBM. IBM SPSS statistics for windows. Version 29.1. Armonk. NY: IBM Corp; 2023.
15. Giri A, Paul P. Applied marketing analytics using spss: Modeler, statistics and amos graphics. PHI Learning Pvt. Ltd.; 2020.
16. Verma K, Bhardwaj S, Arya R, Islam MS, Bhushan M, Kumar A, et al. Latest tools for data mining and machine learning. IJITEE 2019; 8 (9):18-23.
17. Lee KH, Choi GH, Yun J, Choi J, Goh MJ, Sinn DH, et al. Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: A multi-center study. NPJ Digit Med 2024; 7(1):2.
18. Joloudari JH, Saadatfar H, Dehzangi A, Shamshirband S. Computer-aided decision-making for predicting liver disease using pso-based optimized svm with feature selection. Inform Med Unlocked 2019; 17:100255.
19. Muschelli J. Roc and auc with a binary predictor: A potentially misleading metric. J Classif 2020; 37(3):696-708.
20. Ji J, Yang C, Liu J, Liu J, Yin B. A comparative study on swarm intelligence for structure learning of bayesian networks. Soft Comput 2017; 21:6713-38.
21. Mohammad EB, Ahmad M. A systematic evaluation of big data-driven colorectal cancer studies. Med Glas 2024; 21 (1).
22. Pathan RK, Shorna IJ, Hossain MS, Khandaker MU, Almohammed HI, Hamd ZY. The efficacy of machine learning models in lung cancer risk prediction with explainability. PLoS One 2024; 19(6):e0305035.
23. Osong B, Dekker A, Wee L, van Soest J, Bermejo I. Bayesian network structure for predicting two-year survival inpatients diagnosed with non-small cell lung cancer. Ann Oncol 2023; 33; S107-S107.
24. Araghi M, Fidler-Benaoudia M, Arnold M, Rutherford M, Bardot A, Ferlay J, et al. International differences in lung cancer survival by sex, histological type and stage at diagnosis: An icbp survmark-2 study. Thorax 2022; 77(4):378-90.
25. Al-Zubaidy NN. Increase in the probability of lung cancer disease in some homes of al-mafraq city, jordan caused by radon. Hikari Ltd Adv Theor Appl Mech 2012; 5(5):199-207.
26. Xu Y, Dong X, Qin C, Wang F, Cao W, Li J, et al. Metabolic biomarkers in lung cancer screening and early diagnosis. Oncol Lett 2023; 25(6):1-11.