

PAPER • OPEN ACCESS

## Multi-Criteria Decision Support System for Lung Cancer Prediction

To cite this article: Baidaa Al-Bander *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1076** 012036

View the [article online](#) for updates and enhancements.

You may also like

- [Multi-criteria decision approach with stakeholders for food waste management](#)  
R Abu, M A A Aziz, N Sapuan *et al.*
- [Multiple Criteria Decision Analysis under uncertainty in sustainable construction: a neutrosophic modified best-worst method](#)  
M Scherz and A Vafadarnikjoo
- [A comprehensive review of hybrid game theory techniques and multi-criteria decision-making methods](#)  
Muhammad Akram Ramadhan Ibrahim, Nor Izzati Jaini and Ku Muhammad Naim Ku Khalif



**ECS** The Electrochemical Society  
Advancing solid state & electrochemical science & technology

### 242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Early hotel & registration pricing ends September 12

Presenting more than 2,400 technical abstracts in 50 symposia

The meeting for industry & researchers in

**BATTERIES**  
**ENERGY TECHNOLOGY**  
**SENSORS AND MORE!**

 Register now!

 **ECS Plenary Lecture featuring M. Stanley Whittingham,**  
Binghamton University  
Nobel Laureate –  
2019 Nobel Prize in Chemistry



# Multi-Criteria Decision Support System for Lung Cancer Prediction

Baidaa Al-Bander<sup>1</sup>, Yousra Ahmed Fadil<sup>1</sup>, Hussain Mahdi<sup>1</sup>

<sup>1</sup> Department of Computer Engineering, College of Engineering, University of Diyala, Iraq

E-mail: baidaa.q@gmail.com

**Abstract.** Lung cancer is one of the most common deadly malignant tumours, with the most rapid morbidity and death worldwide. Cancer risk prediction is a challenging and complex task in the field of healthcare. Many studies have been carried out by researchers to analyse and establish lung cancer symptoms and factors. However, further improvements are vital and required to be conducted in order to overcome the persistent challenges. In this study, a multi-criteria decision support system for lung cancer risk prediction based on a web-based survey data has been presented and realised. The proposed framework aims to incorporate the powerful of analytical hierarchy process (AHP) with artificial neural network for constituting lung cancer prediction model. The multiple criteria decision-making strategy (AHP) assigns a weight to each individual cancer symptom feature from survey data. The weighted features are then used to train multi-layer perceptron artificial neural network (ANN) to build a disease prediction model. Experimental analysis and evaluation performed on 276 subjects revealed promising prediction performance of developed lung cancer prediction framework in terms of various classification metrics.

## 1. Introduction

Globally, lung cancer is the most common cause of cancer death accounting for 19% of all cancer deaths; and following the breast cancer, the most spread form of cancer. In 2012, GLOBOCAN estimated that 1, 242, 000 new diagnosed lung cancer cases among men, which is almost 17% of all cancer types (excluding skin cancer) and 583 000 (9%) of new cancer cases among women. Lung cancer also accounts for 19% of all cancer deaths [1, 2]. In people aged less than 40 year, the incidence rate of lung cancer is low among both genders, but it starts rising up in age between 75-80 years [3]. Several risk factors have been found to increase the chance of lung cancer incidence and make the subject more likely to developing the cancer. In several conducted studies, a high positive family history of lung cancer has been found to be a risk factor for developing lung cancer [4]. One another major risk factor is tobacco smoking which can cause all other major histological categories of lung cancer. The duration of smoking and second-hand exposure effects of tobacco have been demonstrated and recognised as the strongest determinant of cancer by public health authorities [5, 6]. Number of studies suggested many other risk factors such as dietary supplements [7], alcohol consumption [8], chronic obstructive pulmonary disease [9], occupational exposures (examples; asbestos, radon and silica) [10, 11], and indoor air pollution (example; coal burning) [12] and explained their key and significant roles in developing lung cancer.



High risk individuals are screened to avoid deaths from lung cancer, as early detection increases the survival rate of patients with lung cancer. The physicians utilise different tests, including Computed Tomography (CT) scans, blood tests, X-rays, Positron Emission Tomography (PET) scan, and bone scan for reliably diagnosing the severity of lung cancer [13]. Many lung cancer prediction systems using an electronic health record data including subjects' health conditions, medical history of patients, socio-demographic factors, genetical and biological criteria, and behavioural and lifestyle conditions have been presented and reported in the literature. Authors in [14, 15] proposed logistic regression-based methods for early lung cancer detection using set of risk factors. Moreover, Markaki et al. [16] presented multivariable analysis of 36 risk predictors, utilising feature selection with Cox regression for prediction lung cancer in smoker of all ages.

Machine and deep learning algorithms are widely used for automatically lung cancer detection, screening, and diagnosis aiming to save physicians' time and efforts. Luna et al. [17] used an effective machine learning tool, random forest to recognize known and new predictors of symptomatic radiation pneumonitis, which is introduced as a radiotherapy dose limiting toxicity for advanced NSCLC. The authors in [18] proposed Internet of Things (IoT) approach combined with a fuzzy clustering technique to predict lung cancer through monitoring by providing commands aiming towards improving healthcare. Using 400 samples from cancer and non-cancerous data, Ahmed et al. [19] developed a lung cancer prediction model using k-means clustering algorithm and decision tree. Lynch Chip and colleagues [20] used multiple machine learning techniques including decision tree, linear regression, SVM, gradient boosting machines to identify patient's survival. Cha et al. performed and evaluated a trained deep convolutional neural network model to identify operable cancer with chest radiographs (CXRs) [21]. In [22], the authors provided an evidence that deep convolutional neural networks can be used for death-rate risk prediction using computed tomography (CT) form Non-small-cell lung cancer (NSCLC) patients. Recently, Ahmad and Mayya [23] designed a lung cancer prediction tool based on risk factors using random forest classifier. In the most recent research paper [24], lung cancer was identified in the elderly people using risk factors and deep neural networks.

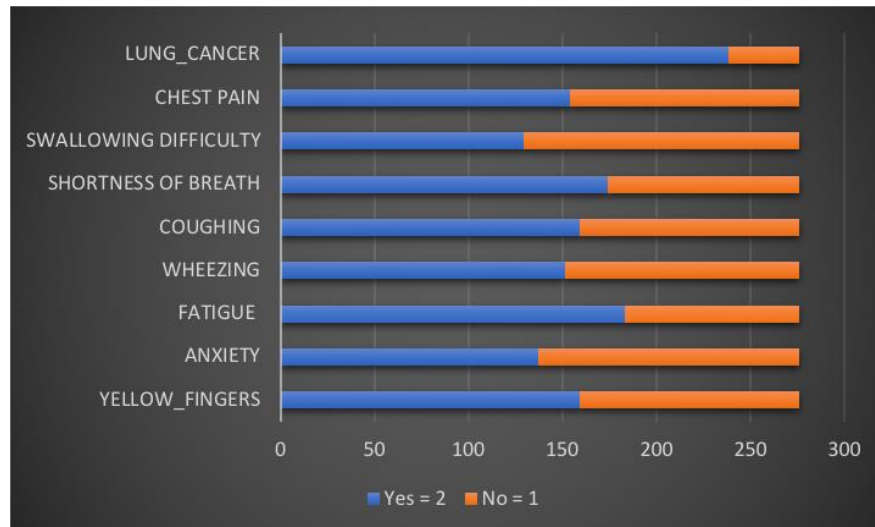
In addition to risk factors, symptoms of lung cancer have been considered for lung cancer prediction [25] including anxiety, yellow fingers, fatigue, wheezing, coughing, shortness of breath, swallowing difficulty, chest pain, loss of weight, and loss of appetite [26, 27]. The symptoms of lung cancer are not specific and difficult to be recognized and diagnosed. If the symptoms have not been early detected and recognized, this would lead to degradation in the patient's health and subsequently difficulty of treatment [28]. This concludes that very few studies were conducted to explore the symptoms for lung cancer prediction. In this paper, we propose a multi-criteria decision aid system to predict lung cancer from web survey-based data which includes description of lung cancer symptoms collected from 276 patient subjects and control. The presented framework proposes the analytical hierarchy process (AHP) methodology for lung cancer feature weighting and selection and multi-layer perceptron (MLP) for symptoms classification and disease prediction. This incorporation of purely statistical-based method with machine learning method allows to leverage disease feature selection to help with the learning of the neural network, alleviating the effect of low significant features in the disease prediction. The rest of this paper is presented and structured as follows. In section 2, the data and proposed methodology are presented and described. Results of the proposed framework are presented and discussed in Section 3. Finally, the research work is concluded in Section 4.

## 2. Materials and Method

### 2.1. Materials

The data was collected from a website survey using a feedback received from the participant [29]. The number of subjects participated in the survey was 276 with eight attributes representing the risk factors, two attributes representing socio-demographic (age and gender), and six attributes representing risk factors. In this study, we focus on symptoms of cancer for lung cancer prediction including anxiety, yellow fingers, fatigue, wheezing, coughing, shortness of breath, swallowing difficulty, and chest pain.

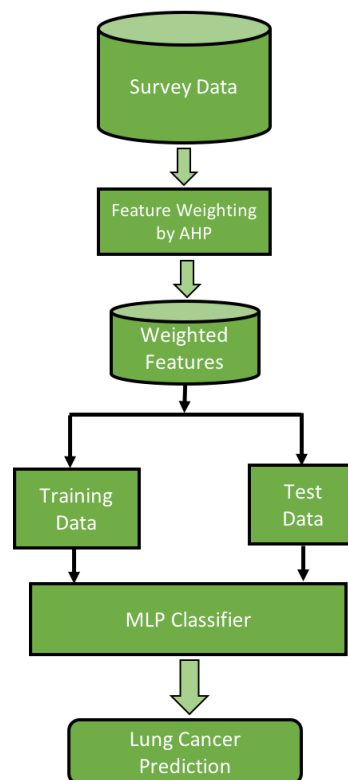
There are 134 female and 142 male subjects their age between (21-87), including 38 control cases and the rest with lung cancer. The distribution of symptoms among subjects is explained in Figure 1.



**Figure 1.** The distribution of symptoms among 276 subjects. Yes: refers to the presence of symptom, No: refers to the absence of symptom.

## 2.2. Method

The block diagram of our proposed framework is depicted in Figure 2. The components of proposed methodology are explained as follows:

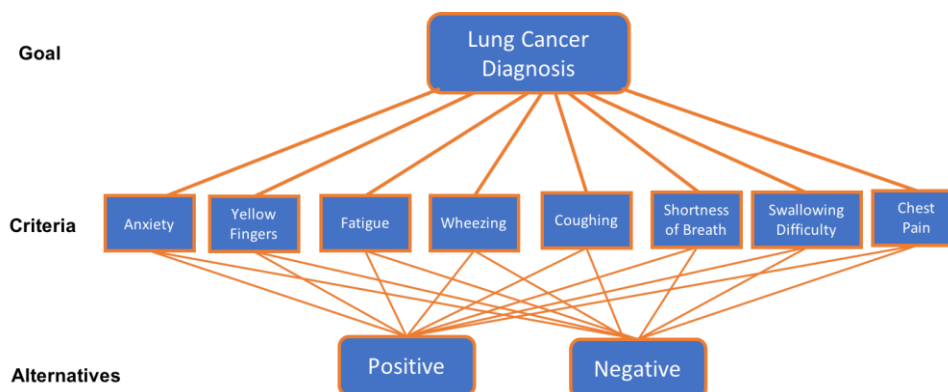


**Figure 2.** The block diagram of developed lung cancer prediction system.

### 2.2.1. Feature Weighting by AHP

The analytic hierarchy process (AHP) method is a multi-criteria decision making/aid (MCDA) technique developed by Saaty [30] attempting to determine the importance of each attribute/variable on a pair basis using matrix called decision matrix. The AHP method comprises three main stages which are: construction of hierarchy (decomposition), analysis of data by determining the priority/preference, and validation of consistency. Saaty [30] suggests defining and splitting up the problem and deciding the goal as the first step of the AHP analysis which is in our study lung cancer diagnosis/prediction. In the constructed hierarchical structure, where the goal of problem pinned at the top, criteria (symptoms in our study) are placed at the intermediate level and, finally alternatives at the base (presence or absence of lung cancer), as shown in Figure 3.

For preference/priority analysis, a pairwise comparison matrix (PWC), shown in Figure 4, is created using priority scale proposed by Saaty [30] as described in Table 1. The relative ratio scale of measurement derived from paired comparisons of the symptoms is used to help the expert/physician in establishing and judging preferences of criteria over others. The scale for pairwise comparisons ranges from 1 to 9. A verbal statement related to each scale is used to define the importance of individual criteria (symptom) over other. The symbolic expression of scales along with the numeric representation can be defined as equal (1), moderate (3), strong (5), very strong (7), and extreme (9). While 2, 4, 6, and 8 are utilised between the aforementioned symbolic expression judgment. To represent the inverse comparison, the reciprocals are used, for instance, if X is very strongly more important than Y, then X is 7 times as important as Y, while Y is 1/7 as important as X.



**Figure 3.** Hierarchy structure of decision support system based on disease's symptoms.

$$PWC = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}$$

**Figure 4.** Pairwise comparison matrix (PWC). Value of (a) refers to the scale value established by expert.

**Table 1.** Verbal description of scale values for AHP preferences

Scale	Degree of Preference
1	Equal preferred
3	Moderate preferred of one symptom over another
5	Strong preferred
7	Very strong preferred
9	Extreme preferred
2, 4, 6, 8	Intermediate values between two scales
1/3	Moderately less preferred
1/5	Strongly less preferred
1/7	Very strongly less preferred
1/9	Extremely less preferred
1/2, 1/4,	Intermediate values between two scales
1/6, 1/8	

To verify the consistency of the preference rating conducting by the expert in the constructed PWC, consistency ratio (CR) should be determined. Saaty [30] defines that the acceptable threshold value of CR is less than 0.1; otherwise, the rating conducted by expert should be considered inconsistent as follows:

$$CR = \begin{cases} < 0.1 & \text{Acceptable} \\ \geq 0.1 & \text{Unacceptable} \end{cases} \quad (1)$$

In case the comparison matrix is inconsistency, the judgments by expert should be reviewed and revised until obtaining a consistent matrix. The formula of consistency ratio can be defined as:

$$CR = \frac{CI}{RI} \quad (2)$$

Where: (RI) is the random consistency index suggested by [30] and can be obtained from table 2.

**Table 2.** Random consistency index (RI) associated with pairwise comparison matrix size (number of criteria n)

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46

The consistency index (CI) is determined using the maximum eigenvalue ( $\lambda_{max}$ ) computed from pairwise comparison matrix as follows:

$$CI = \frac{(\lambda_{max} - n)}{n - 1} \quad (3)$$

Where n represents the number of criteria (symptoms).

Eigenvector and  $\lambda_{max}$  are calculated using the following procedure:

- Step 1:** Determine the column sum of PWC matrix
- Step 2:** Divide the elements of each individual column of PWC matrix by its corresponding column sum to produce a normalized PWC matrix.
- Step 3:** To find the normalized eigenvector (which is also called weight vector or priority vector), the average of normalized row elements in PWC matrix is determined.
- Step 4:** Maximum eigenvalue ( $\lambda_{max}$ ) is obtained by summation of the products between the elements of the sum of columns (Step1) and each element of normalized eigenvector (Step3).

Finally, the obtained eigenvector is represented a weight vector and each individual value in eigenvector represent the weight of each corresponding criteria (symptom feature).

### 2.2.2. MLP for Lung Cancer Prediction

Multi-layer perceptron (MLP) [31] is a feedforward artificial neural network (ANN) that can be trained to classify the input patterns to desired output. For our ANN model, we designed three layers network with 8 neurons in the input layer, 10 neurons in the hidden layer and 2 neurons in the output layer. Our ANN has been trained on the lung cancer data using a backpropagation algorithm to adjust the weights of the network with gradient descent as an optimization method. The cross-entropy loss function is used to tune the weights by reducing the error in each epoch aiming to improve the performance of ANN model. The gradient value is computed and updated in an incremental model (updated after each input) and the soft-max activation function is used in the output layer. Input symptoms data to the network has been weighted using the weights generated in the eigenvectors of AHP and then passed into the network. The output layer represents the lung cancer presence (1) or absence (0).

## 3. Experimental Results and Discussion

To conduct the experiments, the pairwise comparison (PWC) matrix of the AHP, shown in Table 3, are obtained with the help of doctor judgment where symptoms are abbreviated as CP: chest pain, SD: swallowing difficulty, WZ: wheezing, YF: yellow fingers, AN: anxiety, FG: fatigue, CO: coughing, SOB: shortness of breath. This pairwise comparison matrix has been considered after many revisions to achieve the consistency condition. The achieved consistency ratio of comparison matrix is 0.09 (< 0.1, consistent) which is computed as follows:

**Table 3.** Pairwise comparison matrix (PWC).

	CP	SD	WZ	YF	AN	FG	CO	SOB
CP	1	3	1	5	4	5	1/3	1/4
SD	1/3	1	1/3	2	3	3	1/5	1/9
WZ	1	3	1	6	7	8	1	1
YF	1/5	1/2	1/6	1	1/3	1/4	1/7	1/8
AN	1/4	1/3	1/7	3	1	1/2	1/5	1/6
FG	1/5	1/3	1/8	4	2	1	1/8	1/6
CO	3	5	1	7	5	8	1	1/2
SOB	4	9	1	8	6	6	2	1
Sum	<b>09.98</b>	<b>22.17</b>	<b>4.77</b>	<b>36.00</b>	<b>28.33</b>	<b>31.75</b>	<b>05.00</b>	<b>03.32</b>

**Table 4.** Normalized pairwise comparison matrix (NPWC).

	CP	SD	WZ	YF	AN	FG	CO	SOB	Sum	Avg. (Eigenvector)
CP	0.10	0.14	0.21	0.14	0.14	0.16	0.07	0.08	<b>1.02</b>	<b>0.13</b>
SD	0.03	0.05	0.07	0.06	0.11	0.09	0.04	0.03	<b>0.48</b>	<b>0.06</b>
WZ	0.10	0.14	0.21	0.17	0.25	0.25	0.20	0.30	<b>1.61</b>	<b>0.20</b>
YF	0.02	0.02	0.03	0.03	0.01	0.01	0.03	0.04	<b>0.19</b>	<b>0.02</b>
AN	0.03	0.02	0.03	0.08	0.04	0.02	0.04	0.05	<b>0.29</b>	<b>0.04</b>
FG	0.02	0.02	0.03	0.11	0.07	0.03	0.02	0.05	<b>0.35</b>	<b>0.04</b>
CO	0.30	0.23	0.21	0.19	0.18	0.25	0.20	0.15	<b>1.71</b>	<b>0.21</b>
SOB	0.40	0.41	0.21	0.22	0.21	0.19	0.40	0.30	<b>2.34</b>	<b>0.29</b>

The value of  $\lambda_{max}$  is found by the sum of product between the eigenvector and sum of each individual column in comparison matrix giving value of 8.895. Given  $n = 8$  symptoms,  $CI = (8.895 - 8)/(8 - 1) = 0.1279$  and  $RI = 1.41$  (from Table 2), then  $CR = 0.1279/1.41 = 0.09$ . The values of eigenvector represent the weight of each lung cancer symptom as follows (symptom, weight): (CP, 0.13), (SD, 0.06), (WZ, 0.20), (YF, 0.02), (AN, 0.04), (FG, 0.04), (CO, 0.21), (SOB, 0.29). The highest weight is assigned to shortness of breath symptom, followed by coughing symptom. The lowest weight is allocated to yellow finger, followed by anxiety and fatigue. To determine the importance and contribution of each symptom over others in predicting the lung cancer, the provided lung cancer data is multiplied by its allocated weight value before feeding them into the neural network for training. The most important attribute (symptom criterion) has the highest weight value and vice versa. It is worth mentioning that the minority class in the dataset (negative classes here) has been oversampled during the training to handle the problem of label unbalancing. The data is divided into 60% training, 15% validation, and 25% testing. The performance of proposed framework for lung cancer prediction has been evaluated by comparing the predicted lung cancer label with the ground truth provided in the dataset using many evaluation metrics. The evaluation metrics include accuracy, sensitivity, specificity, and F1-score, which are defined as follows:

$$Accuracy (AC) = \frac{tp + tn}{tp + tn + fp + fn} \quad (4)$$

$$Sensitivity (Sen) = \frac{tp}{tp + fn} \quad (5)$$

$$Specificity (Sp) = \frac{tn}{tn + fp} \quad (6)$$

$$F1\ Score = \frac{2tp}{2tp + fp + fn} \quad (7)$$

Where  $tp$ ,  $fp$ ,  $tn$  and  $fn$  represent true positive, false positive, true negative, and false negative, respectively. The trained neural network has achieved 77.7%, 71.2%, 87.5%, 82.8% in terms of accuracy, specificity, sensitivity, and F1 score, respectively on the test data without weighting; whereas achieved 80.7%, 75.3%, 89.9%, 86.4% using the same metrics on the test data weighted by AHP weights. The obtained results indicate that the weighting method represented by AHP is an effective strategy in the medical cases which are controlled by various criteria but the priority (importance) of the individual criterion over others is not well defined. It can be noticed that the specificity in both models is slightly lower than other reported evaluation metrics which due to the limited number of negative examples (control subjects) in the dataset. However, this would not degrade the performance of the prediction model which focuses on improving the predicting of positive cases than the negative cases. The performance of lung cancer prediction system can be further improved by training the neural network on a larger dataset.

In terms of comparing our study performance with the state-of-the-art methods, our developed method reveals a competitive and promising performance. For example, the model proposed by Petousis et al. [32] achieved accuracy of 65% using dynamic Bayesian model on population of 25,486 subject. Likewise, Wang et al. [33] reported lung cancer prediction accuracy of 67% using conditional Gaussian Bayesian network on population of 961 subject. Recently, Kaviarasi [34] presented system based on Gaussian classifier for lung cancer prediction from risk factors but the author did not report the performance of the accuracy metric. They claim that their propose method gives area under the ROC curve of 88.1% on population of 321 subject.



#### 4. Conclusion

In this paper, an automated method to predict the lung cancer from web-based survey data has been presented. The benefits of weighting and selectin features from symptoms lung cancer have been designed, realized and evaluated. The developed system, which is merging AHP with MLP, has proved to be efficient in feature weighting, learning and identification of lung cancer in a group of subjects. Furthermore, the obtained results revealed that the proposed methodology is promising in predicting lung cancer efficiently and accurately in terms of accuracy, sensitivity, specificity, and F1-score. The proposed lung cancer prediction framework is generic and can be easily developed to other diseases' prediction models.

#### References

- [1] Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1), 5.
- [2] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*, 136(5), E359-E386.
- [3] Malvezzi, M., Bosetti, C., Rosso, T., Bertuccio, P., Chatenoud, L., Levi, F., & La Vecchia, C. (2013). Lung cancer mortality in European men: trends and predictions. *Lung Cancer*, 80(2), 138-145.
- [4] Bailey-Wilson, J. E., Sellers, T. A., Elston, R. C., Evens, C. C., & Rothschild, H. (1993). Evidence for a major gene effect in early-onset lung cancer. *The Journal of the Louisiana State Medical Society: official organ of the Louisiana State Medical Society*, 145(4), 157.
- [5] Smoke, T., & Smoking, I. (2004). IARC monographs on the evaluation of carcinogenic risks to humans. *IARC, Lyon*, 1-1452.
- [6] Doll, R., Peto, R., Boreham, J., & Sutherland, I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *Bmj*, 328(7455), 1519.
- [7] Lam, T. K., Gallicchio, L., Lindsley, K., Shiels, M., Hammond, E., Tao, X. G., & Guallar, E. (2009). Cruciferous vegetable consumption and lung cancer risk: a systematic review. *Cancer Epidemiology and Prevention Biomarkers*, 18(1), 184-195.
- [8] García-Lavandeira, J. A., Ruano-Ravina, A., & Barros-Dios, J. M. (2016). Alcohol consumption and lung cancer risk in never smokers. *Gaceta sanitaria*, 30, 311-317.
- [9] Mayne, S. T., Buenconsejo, J., & Janerich, D. T. (1999). Previous lung disease and risk of lung cancer among men and women nonsmokers. *American journal of epidemiology*, 149(1), 13-20.
- [10] Cogliano, V. J., Baan, R., Straif, K., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., & Galichet, L. (2011). Preventable exposures associated with human cancers. *Journal of the National Cancer Institute*, 103(24), 1827-1839.
- [11] International Agency for Research on Cancer. IARC Monographs, Chemical Agents and Related Occupations. Vol 100F (2012)[displayed 13 October 2017].
- [12] Afaj, A. (2011). IARC Monograph on the Evaluation of Carcinogenic Risk to Humans, Volume 95: Household Use of Solid Fuels and High-temperature Frying.
- [13] Ruddle, R. W. (2007). *Cancer biology*. Oxford University Press.
- [14] Iyen-Omofoman, B., Tata, L. J., Baldwin, D. R., Smith, C. J., & Hubbard, R. B. (2013). Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax*, 68(5), 451-459.
- [15] Bach, P. B., Kattan, M. W., Thornquist, M. D., Kris, M. G., Tate, R. C., Barnett, M. J., & Begg, C. B. (2003). Variations in lung cancer risk among smokers. *Journal of the National Cancer Institute*, 95(6), 470-478.
- [16] Markaki, M., Tsamardinos, I., Langhammer, A., Lagani, V., Hveem, K., & Røe, O. D. (2018). A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a HUNT study. *EBioMedicine*, 31, 36-46.
- [17] Luna, J. M., Chao, H. H., Diffenderfer, E. S., Valdes, G., Chinniah, C., Ma, G., & Simone II, C.

- B. (2019). Predicting radiation pneumonitis in locally advanced stage II–III non-small cell lung cancer using machine learning. *Radiotherapy and Oncology*, 133, 106-112.
- [18] Palani, D., & Venkatalakshmi, K. (2019). An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification. *Journal of medical systems*, 43(2), 21.
- [19] Ahmed, K., Kawsar, A. A., Kawsar, E., Emran, A. A., Jesmin, T., Mukti, R. F., & Ahmed, F. (2013). Early detection of lung cancer risk using data mining.
- [20] Lynch Chip, M., Behnaz, A., Fuqua Joshua, D., de Carlo Alexandra, R., Bartholomai James, A., Balgemann Rayeanne, N., & Frieboes Hermann, B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International journal of medical informatics*, 108, 1-8.
- [21] Cha, M. J., Chung, M. J., Lee, J. H., & Lee, K. S. (2019). Performance of deep learning model in detecting operable lung cancer with chest radiographs. *Journal of thoracic imaging*, 34(2), 86-91.
- [22] Hosny, A., Parmar, C., Coroller, T. P., Grossmann, P., Zeleznik, R., Kumar, A., & Aerts, H. J. (2018). Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS medicine*, 15(11), e1002711.
- [23] Ahmad, A. S., & Mayya, A. M. (2020). A new tool to predict lung cancer based on risk factors. *Heliyon*, 6(2), e03402.
- [24] Chen, S., & Wu, S. (2020). Identifying Lung Cancer Risk Factors in the Elderly Using Deep Neural Networks: Quantitative Analysis of Web-Based Survey Data. *Journal of Medical Internet Research*, 22(3), e17695.
- [25] Okoli, G. N., Kostopoulou, O., & Delaney, B. C. (2018). Is symptom-based diagnosis of lung cancer possible? A systematic review and meta-analysis of symptomatic lung cancer prior to diagnosis for comparison with real-time data from routine general practice. *PloS one*, 13(11), e0207686.
- [26] Corner, J., Hopkinson, J., Fitzsimmons, D., Barclay, S., & Muers, M. (2005). Is late diagnosis of lung cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis. *Thorax*, 60(4), 314-319.
- [27] Birring, S. S., & Peake, M. D. (2005). Symptoms and the early diagnosis of lung cancer.
- [28] Hamilton, W., Peters, T. J., Round, A., & Sharp, D. (2005). What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax*, 60(12), 1059-1065.
- [29] Staceyinrobert. Data World. Available Online: <https://data.world/sta427ceyin/survey-lung-cancer>. [Accessed: 15/ 07/ 2020].
- [30] Saaty, T. L. (1994). How to make a decision: the analytic hierarchy process. *Interfaces*, 24(6), 19-43.
- [31] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
- [32] Petousis, P., Han, S. X., Aberle, D., & Bui, A. A. (2016). Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network. *Artificial intelligence in medicine*, 72, 42-55.
- [33] Wang, K. J., Chen, J. L., & Wang, K. M. (2019). Medical expenditure estimation by Bayesian network for lung cancer patients at different severity stages. *Computers in biology and medicine*, 106, 97-105.
- [34] Kaviarasi, R. (2019). Accuracy enhanced lung cancer prognosis for improving patient survivability using proposed Gaussian classifier system. *Journal of medical systems*, 43(7), 201.