

A DECISION MODEL BASED ON ARTIFICIAL INTELLIGENCE FOR DISEASE PREDICTION AND PATIENT TREATMENT PROCESS PLANNING

Serkan Dericici*¹, Nuri Özgür Doğan²

1. Faculty of Economics and Administrative Sciences, Department of Business Administration, Department of Quantitative Methods, Nevşehir Hacı Bektaş Veli University, Turkey
2. Faculty of Health Sciences, Department of Health Management, Istanbul Medeniyet University, Turkey

Correspondence: serkanderici@nevsehir.edu.tr

ABSTRACT

PURPOSE:

The aim of this study is to determine which patients are at risk of lung cancer with a sample of 1,000 people living in China and a data set consisting of 20 variables. In this direction, it is aimed to take preventive measures for individuals at risk of developing lung cancer, to protect human health, and to contribute to the effective and efficient use of resources. In this way, the model developed in the planning of preventive and therapeutic activities of professionals working in the health sector will be exemplary as a decision model.

METHODOLOGY:

A linear regression model was developed and applied to a public domain data set to predict individuals at high risk of developing lung cancer using machine learning on the Microsoft Azure Machine Learning Studio platform.

FINDINGS:

Lung cancer risk of the patients in the sample was predicted with the machine learning application. It was determined that these predictions made with machine learning gave an effective result and provided efficient policies. The developed model and application example is proposed as a modern, valid and artificial intelligence-based decision-making model for professionals.

ORIGINALITY:

There are few statistical analysis and machine learning applications for lung cancer prediction in the literature. In this paper, in addition to the prediction model, statistical analyses will be performed, preventive measures will be proposed, and the scope of the analyses will be made holistic by linking it to efficiency in resource utilization. As a result, this will enable professionals to make optimal decisions.

KEYWORDS

lung cancer, prediction, big data, artificial intelligence, machine learning.

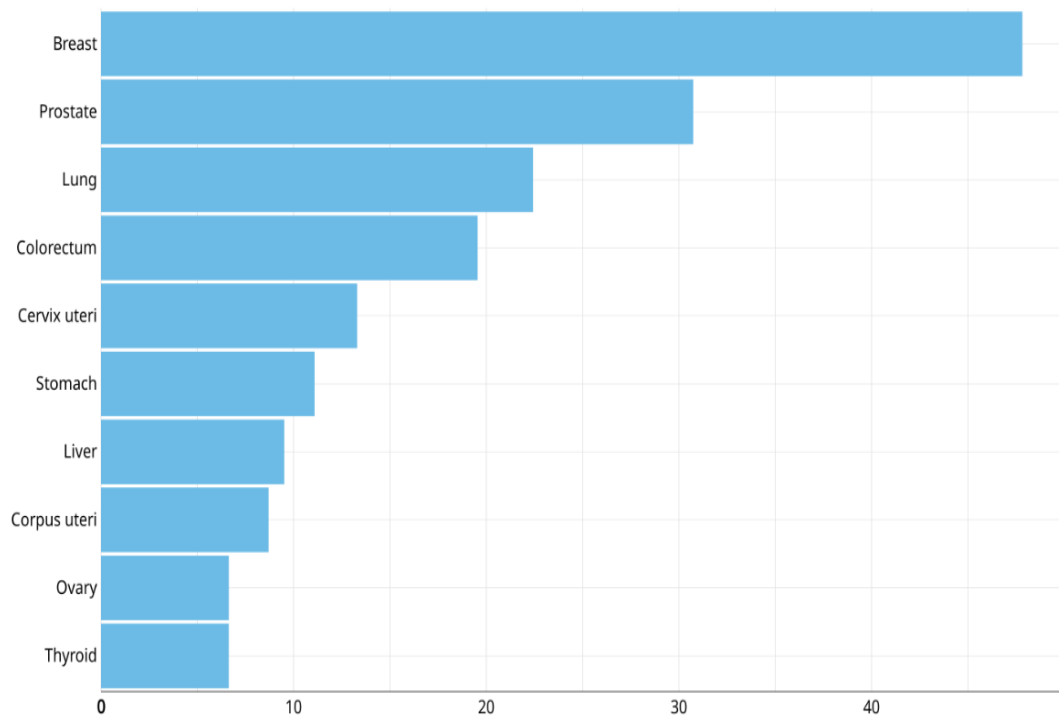
INTRODUCTION

The technological innovations that entered our lives with the last industrial revolution have led to the acquisition of big data in every sector. The large amount of data obtained in processes has created the problem of drawing useful conclusions from this data. The problem of analyzing big data has become one of the departments to which companies allocate the most resources. Transforming the raw data obtained in big data analysis into useful information and determining policies in line with this information directly affects the success and efficiency of the sectors. However, classical methods are insufficient in big data analysis [31]. At this point, machine learning emerges as an innovative approach based on artificial intelligence and is increasingly used in big data analysis. The use of machine learning algorithms in big data analysis is important for the efficient use of resources and the development of effective policies [32].

Machine learning algorithms are used for classification, clustering and prediction problems [33]. When the literature is examined, it is seen that machine learning has started to be widely used not only in marketing or production but also in all sectors. Artificial intelligence-based applications such as machine learning have started to be used in very important studies in the health sector [34]. Prediction models, especially robotic applications and visual processing, give successful results in the health sector. For this reason, in this study, a prediction model has been developed for patients to get lung cancer under certain variables. As shown in Figure 1, lung cancer is the third most common cancer in the world with 22.4% [Cancer Today (iarc.fr)]. Due to the prevalence of lung cancer, a machine learning application on lung cancer was realized in this study.

FIGURE 1: TYPES OF CANCER SEEN IN THE WORLD (2020) [CANCER TODAY (IARC.FR)]

Estimated age-standardized incidence rates (World) in 2020, World, both sexes, all ages (excl. NMSC)



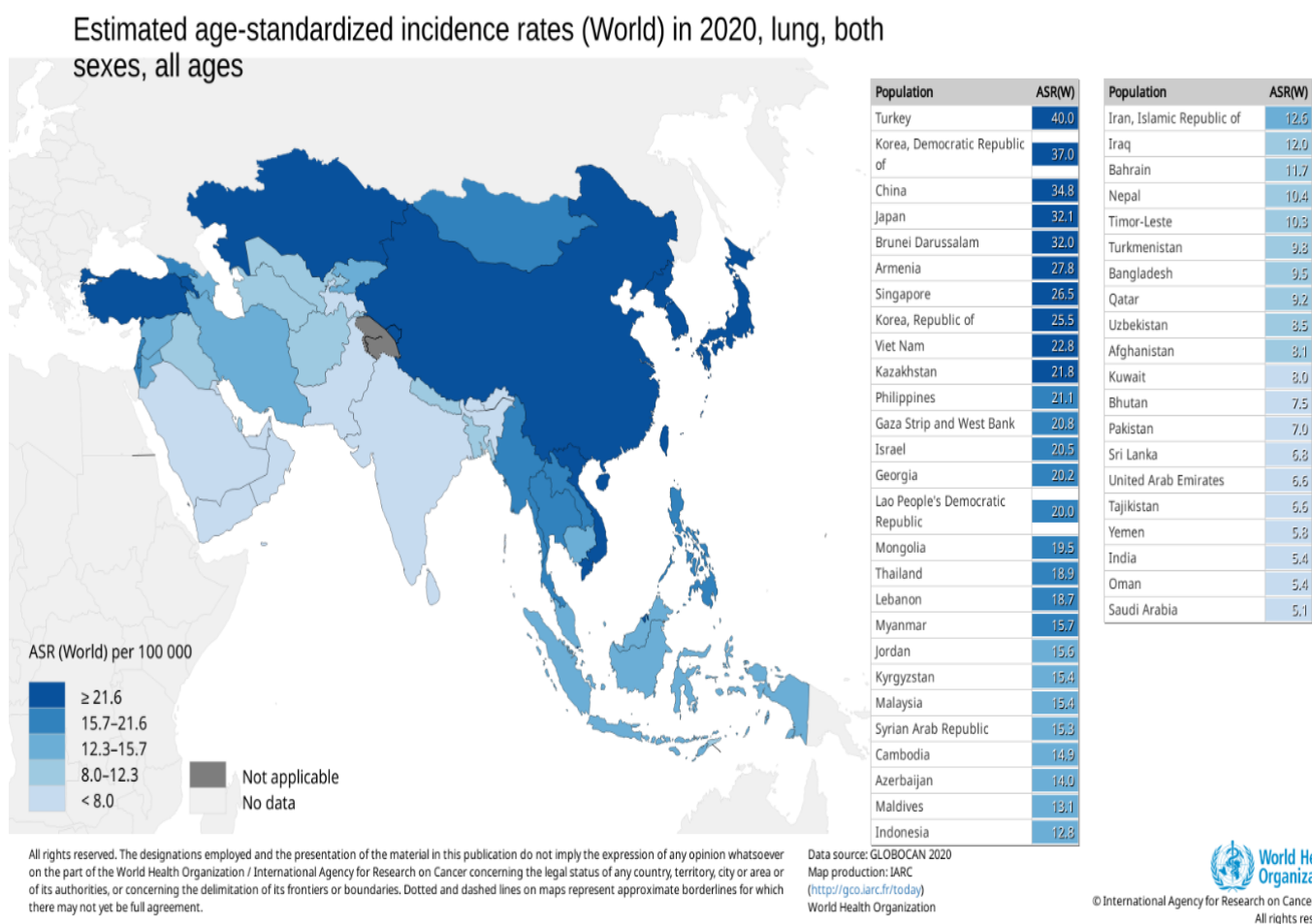
Data source: Globocan 2020
Graph production: Global Cancer
Observatory (<http://gco.iarc.fr>)

International Agency for Research on Cancer
World Health
Organization

Prediction models for lung cancer have been developed and applied with classical methods in the literature; however, few studies have been conducted with machine learning. Information on these studies is presented in the literature review section.

Early detection of lung cancer is very important for the success of treatment. AI-based prediction models are very successful in predicting the condition of a new patient by learning from previous observations. Therefore, the problem of predicting lung cancer in patients in a practical and fast way instead of lengthy examinations is very important. In this study, we investigate whether an AI-based model can actually provide effective results. According to World Health Organization data, 34.8% of lung cancer cases in Asia are in China, which is the reason for the analysis on China in this study. This information is shown in Figure 2 [Cancer Today (iarc.fr)].

FIGURE 2: DISTRIBUTION OF LUNG CASES IN ASIA BY COUNTRY [CANCER TODAY (IARC.FR)]



This paper is designed in four sections: introduction, literature review, application methodology and dataset, application and conclusion. These sections will be explained in the following sections, respectively.

This study consists of an introductory section followed by a literature review where the research model is presented, a methodology section where information about machine learning is presented, an application section where the results of the analysis are presented, and a conclusion and discussion section where the findings are presented.

LITERATURE REVIEW

Smart Objects, which entered our lives in the 21st century [5], began to develop at a very high speed with the last industrial revolution. In all sectors that affect our lives, large amounts of data are instantly obtained, stored, analyzed with artificial

intelligence components and transformed into useful information. Machine learning algorithms have been developed to analyze this data. Machine learning is a field of science developed in close relationship with data mining, which includes methods for computers to learn by analogy with the thoughts of individuals [13].

Machine learning minimizes statistical errors in big data analysis [35]. Since the aim of this study was to predict individuals who are likely to develop lung cancer, the literature was examined in this direction.

TABLE 1: LITERATURE REVIEW ON MACHINE LEARNING APPLICATIONS ON LUNG CANCER

Author(s)	Year	Method(s)	Application
Cai et al.	2015	Receiving Operating Curve Random Forests Maximum Relevancy	Classification of lung cancer types.
Danjuma	2015	Multilayer Perceptron Decision Tree Naive Bayes	Machine learning was applied to determine the probability of survival after surgery.
Burki	2016	Image evaluation	Machine learning application was made on the possibility of surviving lung cancer.
Podolsky et al.	2016	Support Vector Machine Naive Bayes Decision Tree	Determining the type of lung cancer
Kadir and Gleeson	2018	Prediction	Lung cancer prediction approaches evaluated.
Radhika, Nair & Veena	2019	Naive Bayes Support Vector Machine Decision tree Logistic Regression	Classification algorithms, early diagnosis of lung cancer.
Banerjee and Das	2020	Support Vector Machine Artificial Neural Network Random Forests	The issue of detecting early lung cancer has been studied.
Pradhan and Chawla	2020	Literature review	Lung cancer detection with IoT-based machine learning is discussed.
Raoof, Jabbar & Fathima	2020	Naive Bayes Support Vector Machine Logistic regression Artificial Neural Network	The factors causing lung cancer.
Abdullah and Ahmed	2021	Literature review	The detection of lung cancer.
Xie et al.	2021	Naive Bayes	Early lung cancer.
Dritsas and Trigka	2022	Rotation Forest	Lung cancer risk prediction
Sharma and Mittal	2024	Convolutional Neural Network	Prediction of lung cancer disease

It is clear from the literature review that machine learning algorithms have different applications in the health sector by providing many advantages. Machine learning algorithms; It has been the subject of studies such as diagnosis of lung cancer, prediction of deaths and recovery from diseases. The data set was determined in line with the literature review. Utilizing this literature review, the application model of the study is built with twenty different variables as shown in Table 3 in the following sections. Thus, a unique lung cancer prediction model was developed.

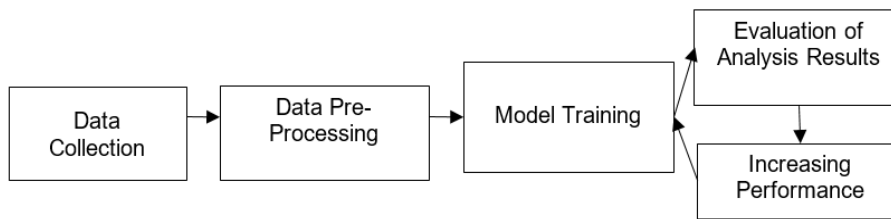
METHODOLOGY

The dataset is obtained from Kaggle¹ and consists of 1,000 observations and 20 variables from approximately 450,000 individuals with lung disease in China. The dataset was introduced to the Microsoft Azure Machine Learning Studio (AMLS²) platform and an attempt was made to predict patients at risk of developing lung cancer. The regression analysis and artificial neural networks used in the prediction are described in this section.

MACHINE LEARNING

The analysis of big data obtained through artificial intelligence components can only be done with artificial intelligence-based systems. Machine learning is a sub-branch of artificial intelligence, first defined by Arthur Lee Samuel in 1959 [24]. Theoretically, machine learning is a method that enables a system developed for a process to make predictions about future periods using information from previous processes [30]. In other words, machine learning is a method that enables computers to learn in a way similar to human thinking [13]. Machine learning applications generally follow the five steps shown in Figure 3:

FIGURE 3: STAGES OF MACHINE LEARNING [29]



Machine learning methods have been developed in the literature for machine learning applications depending on whether the data is labeled or not. The classification of these methods is shown in Table 2:

TABLE 2: CLASSIFICATION OF MACHINE LEARNING ALGORITHMS [15]

Machine Learning Type	Problems Type	Methods
Supervised Learning	Regression Classification	Decision Tree Random Forest Naive Bayes Artificial Neural Network Deep Learning Support Vector Machine K-Nearest Neighbor
Unsupervised Learning	Clustering	K-Means Clustering Hierarchical Clustering Soft Clustering
Semi Supervised Learning	Dimensional Reduction	Single Value Parsing Policy Component Analysis Independent Component Analysis
Reinforced Learning	Q-Learning	Q-Learning

An artificial intelligence-supported linear regression analysis was applied on the Microsoft Azure Machine Learning Studio platform to predict the risk of lung cancer in individuals in the data set within the scope of this study. Therefore, linear regression analysis is explained in the next section.

¹ <https://www.kaggle.com/>

² <https://azure.microsoft.com/en-us/products/machine-learning>

REGRESSION ANALYSIS

Regression analysis, first articulated by Francis Galton, is a supervised machine learning method. In regression analysis, a dependent variable (Y) is estimated by one or more independent variables (Xi) associated with it [3]. The mathematical notation of the regression analysis to express the i index independent variables and ϵ regale error is simply expressed as:
 $y = f(x) + \epsilon \dots\dots\dots(l)$

Regression analysis as a machine learning algorithm; It starts with the division of the data set into training and test data, and has a working principle of feature selection, determination of the regression model and obtaining the results by running the model [8- 15].

The literature review indicates that regression models are widely used in machine learning applications on lung cancer. Lung cancer, which is considered as the dependent variable in the context of regression analysis, is associated with different independent variables. Therefore, the data set used in the study has multiple regression analysis assumptions.

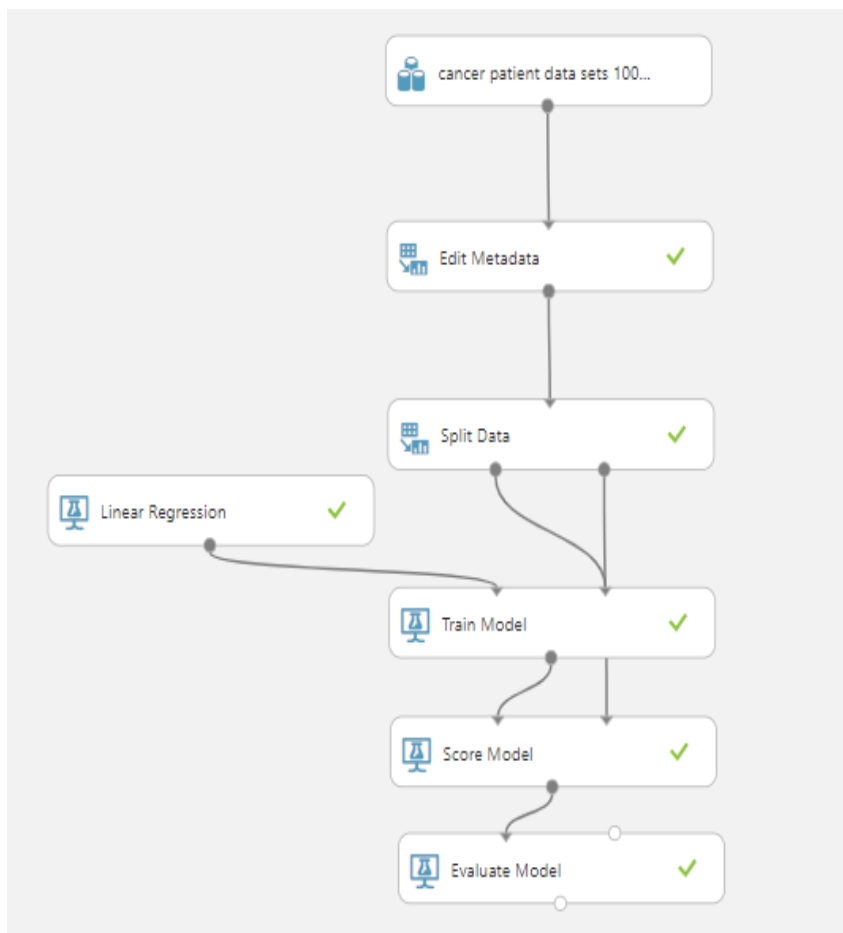
APPLICATION

The dataset used in this study is obtained from Kaggle. This dataset contains 23 factors that cause lung cancer for 1000 individuals living in China. A unique lung cancer prediction model was designed using 20 of these variables. The data was uploaded to the AMLS System and firstly, the data was preprocessed and missing data was completed by averaging on the same platform. In the analyses, the risk level variable was introduced to the system as a categorical dependent variable. The data belonging to 1000 individuals were divided into two parts as 70% training data and 30% test data to enable the system to learn the problem. In this step, it was ensured that the machines learned past disease experiences.

It is important for the protection of human health to identify individuals who pose a risk with this algorithm and to carry out the necessary medical tests and procedures. On the other hand, by following this path, an important planning will be made in the effective and efficient use of personnel, material expenses and medicines in health institutions. Misdiagnoses will be minimized and unnecessary procedures will be prevented in this way and time efficiency will be ensured. This will contribute to the efficient use of staff working hours and resources. On the other hand, following an effective diagnostic pathway and preventing misuse of drugs will increase efficiency in drug use.

The developed model was analyzed on the AMLS platform. The flow of the analysis on the AMLS platform is shown in Figure 4.

FIGURE 4: AMLS APPLICATION STEPS



The descriptive statistics of the variables used in the analyses were determined by the steps shown in Figure 4 and presented in Table 3 together with the results of the regression analysis.

TABLE 3: DESCRIPTIVE STATISTICS AND REGRESSION ANALYSIS RESULTS RELATED TO VARIABLES

Variables No	Independent Variables	Number	Mean	Standard Deviation	Min-Max
V1	Age	1000	37.17	12.01	14-73
V2	Gender	1000	1.40	0.49	1-2
V3	Air Pollution	1000	3.84	2.03	1-8
V4	Alcohol Use	1000	4.56	2.62	1-8
V5	Dust Allergy	1000	5.17	1.98	1-8
V6	OccuPational Hazards	1000	4.84	2.11	1-8
V7	Genetic Risk	1000	4.58	2.13	1-7
V8	Chronic Lung Disease	1000	4.38	1.85	1-7
V9	Balanced Diet	1000	4.49	2.14	1-7
V10	Obesity	1000	4.47	2.13	1-7
V11	Smoking	1000	3.95	2.50	1-8
V12	Passive Smoker	1000	4.20	2.31	1-8
V13	Chest Pain	1000	4.44	2.28	1-9
V14	Coughing of Blood	1000	4.86	2.43	1-9
V15	Fatigue	1000	3.86	2.25	1-9
V16	Weight Loss	1000	3.86	2.21	1-8
V17	Shortness of Breath	1000	4.24	2.29	1-9

V18	Wheezing	1000	3.78	2.04	1-8
V19	Frequent Cold	1000	3.54	1.83	1-7
V20	Dry Cough	1000	3.85	2.03	1-7
Dependet Variable: Risk Level		300	0.96	0.75	0-2
Mean Absolute Error: 0.30 Root Mean Squared Error: 0.41 Relative Absolute Error: 0.46 Relative Squared Error: 0.25 Coefficient of Determination (R ²): 0.79					

The data set consisting of 20 independent variables and one dependent variable used for lung cancer risk of individuals was analyzed. Descriptive statistics of the variables used are presented in Table 3. While creating the data set, the relevant risk factors were graded from 1 to increasing severity and accordingly, exposure levels and averages were determined. In the model developed for the lung cancer risk of individuals aged 14-73 years according to 20 risk factors, it is seen that the independent variables explain 79% of the dependent variable. This is an acceptable rate. Subsequently, the risk of developing lung cancer was estimated. After the risk level variable was expressed as a categorical variable (0: Low, 1: Medium, 2: High), predictions were made by applying machine learning on the AMLS platform. A cross-section of 20 individuals for these predictions is presented in Table 4.

The prediction results of these individuals showed that 10 patients were high risk, 5 patients were medium risk and 5 patients were low risk. Using the system developed as a result of this application, the risk level of patients can be easily found by entering values in this way. Because according to the data of 1000 patients entered into the system with the machine learning application, the computer has now recognized the symptoms of lung cancer. In this way, making a preliminary assessment and applying appropriate treatments plays an important role in terms of treatment and resource utilization efficiency.

TABLE 4: PREDICTING THE RISK LEVELS OF DEVELOPING LUNG CANCER

No	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	Risk Level
0	33	1	2	4	5	4	3	2	2	4	3	2	2	4	4	2	3	1	2	3	0.21*
1	17	1	3	1	5	3	4	2	2	2	2	4	2	3	3	7	6	2	1	7	1.13**
2	35	1	4	5	6	5	5	4	6	7	2	3	4	8	7	9	1	4	6	7	1.90***
3	37	1	7	7	7	7	6	7	7	7	7	7	7	8	2	3	4	5	6	7	1.86***
4	46	1	6	8	7	7	7	6	7	7	8	7	7	9	2	4	4	2	4	2	1.92***
5	24	1	6	8	7	7	6	7	7	3	8	7	9	6	2	5	3	2	1	7	1.98***
6	39	2	4	5	6	6	5	4	6	6	6	6	6	6	3	2	3	1	7	5	1.05**
7	27	2	3	1	4	2	3	2	3	3	2	2	4	2	2	3	1	5	2	6	0.17*
8	52	2	2	4	5	4	3	2	2	4	3	2	2	4	4	2	3	1	2	3	0.24*
9	14	1	2	4	5	6	5	5	4	6	5	4	6	5	3	2	4	7	2	1	0.99**
10	62	1	6	8	7	7	7	6	7	7	8	7	7	9	3	4	4	2	4	2	1.78***
11	61	2	2	3	4	2	1	1	2	4	3	2	1	5	1	3	1	3	2	1	0.09*
12	44	1	2	1	5	3	2	3	2	4	1	4	4	6	7	5	8	1	3	2	1.26**
13	32	2	1	6	7	5	3	2	6	2	3	3	2	3	3	7	4	8	7	7	1.99***
14	29	2	6	7	7	7	7	6	7	7	7	7	7	2	7	6	7	2	3	1	1.95***
15	53	1	6	8	7	7	6	7	3	8	7	9	5	6	2	3	2	1	7	6	1.96***
16	26	1	6	7	7	6	7	7	7	7	7	7	2	7	6	5	7	3	2	3	1.81***
17	47	1	6	7	7	7	7	5	6	7	5	8	2	5	5	4	6	1	2	4	1.91***
18	24	1	3	1	3	2	2	4	3	1	3	2	3	2	4	2	2	3	1	2	0.20 *
19	36	1	3	4	6	5	4	6	5	4	2	5	2	3	6	1	2	4	2	1	1.03**
20	18	1	3	1	4	2	3	1	2	2	3	1	2	1	3	8	6	2	1	2	1.14**

It was determined as "0<Low Risk<0.75<Medium Risk<1.50<High Risk".

*Patients at low risk

**Patients with intermediate risk

***Patients at high risk

RESULT AND CONCLUSION

Machine learning includes artificial intelligence-based algorithms that save time and automatically minimize errors by providing direct targeted results. Worldwide, lung cancer is the third most common type of cancer. Many preventive projects and campaigns such as reducing air pollution, quitting smoking and creating a clean air space are organized worldwide. An examination of the data on China used in this context reveals that lung cancer cases occur at a very high rate. In this respect, the Chinese sample is a very effective study area.

Predicting big data with machine learning algorithms and developing and implementing policies based on these predictions play an important role in achieving accurate and effective results. This is because errors are minimized and holistic analyses are performed using data from all variables of the process.

We used a sample of 1,000 people with lung disease in China to estimate the risk level of patients with lung cancer based on 20 different variables. Based on the analysis using the AMLS Platform, patients with high probability of lung cancer were identified by the artificial intelligence algorithm.

Thanks to such analyses to be made with machine learning applications, early detection of diseases and efficient treatment planning will be possible. Using the analyzes made in this direction, similar risk levels can be determined. This is an efficient way to prevent misdiagnosis and treatment and to protect human health. The use of machine learning algorithms in diagnosis provides speed, time and cost advantages in health institutions. On the other hand, it is an important guide for optimum use of resources and correct planning.

The decision-making processes with the last industrial revolution are succeeding as a result of the correct use of big data. Big data on patients in healthcare organizations is obtained in many ways, from recording a patient's patient cards on computers to the data of sensors. However, when the amount of data in question increases and reaches gigantic dimensions, analyzing it with classical methods and taking decision-making actions as a result of these analyzes are not the right practices. Because it has been clearly demonstrated in the literature that when big data is analyzed with artificial intelligence-based algorithms, more accurate results are obtained and especially the success of prediction models. Analysis of patient information with machine learning and making predictions will increase the success in the decision-making processes of professionals in the health sector. In addition to the machine learning algorithms used in this study, other purpose-built models can be used. Similar studies can be done for different disease groups as well as for different countries and sectors.

References

1. Abdullah DM & Ahmed NS. A review of most recent lung cancer detection techniques using machine learning. *International Journal of Science and Business*. (2021); 5(3), 159-173.
2. Agatonovic-Kustrin S & Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*. (2000); 22(5), 717-727.
3. Akin F., *Sosyal bilimlerde istatistik*. Ekin Kitabevi. (2013)
4. Banerjee N & Das S., Prediction lung cancer–in machine learning perspective. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). (2020, March); (pp. 1-5). IEEE.
5. Bozkurt E, Topçuoğlu Ö & Altiner A. Relationship between productivity and digitalization with Tobit model based on Malmquist index. *Verimlilik Dergisi*. 2022; 67-78.
6. Burki TK. Predicting lung cancer prognosis using machine learning. *The Lancet Oncology*. 2016; 17(10), e421.
7. Cai Z, Xu D, Zhang Q, Zhang J, Ngai SM & Shao J. Classification of lung cancer using ensemble-based feature selection and machine learning methods. *Molecular BioSystems*. 2015; 11(3), 791-800.
8. Danjuma KJ. Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients. 2015; arXiv preprint arXiv:1504.04646.
9. Fox C. Bayesian inference. *Data Science for Transport: A Self-Study Guide with Computer Exercises*. 2018; 75-92.
10. Hinton GE. How neural networks learn from experience. *Scientific American*. 1992; 267(3), 144-151.

11. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I & Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. Technical Report. 2012; arXiv:1207.0580.
12. Jordan M. Attractor dynamics and parallelism in a connectionist sequential machine. In Proceedings of the 1986 Cognitive Science Conference. 1986; Lawrence Erlbaum.
13. Kadir T & Gleeson F. Lung cancer prediction using machine learning and advanced imaging techniques. *Translational lung cancer research*. 2018; 7(3), 304.
14. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I & Chouvarda I. Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*. 2017; 15, 104-116.
15. Koç M, Balas C & Arslan A. Preliminary design of rubble mound breakwaters by using artificial neural networks. *Technical Journal of Turkish Chamber of Civil Engineers*. 2004; 15(4).
16. Kumar DP, Amgoth T & Annavarapu CSR. Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*. 2019; 49, 1-25.
17. McCulloch WS & Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943; 5, 115-133.
18. Mehrotra K, Mohan CK & Ranka S. *Elements of artificial neural networks*. MIT press. 1997.
19. Öztürk K & Şahin ME. A general view of artificial neural networks and artificial intelligence. *Takvim-i Vekayi*. 2018; 6(2), 25-36.
20. Podolsky MD, Barchuk AA, Kuznetsov VI, Gusarova NF, Gaidukov VS & Tarakanov SA. Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels. *Asian Pacific journal of cancer prevention*. 2016; 17(2), 835-838.
21. Pradhan K & Chawla P. Medical Internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*. 2020; 7(4), 591-623.
22. Radhika PR, Nair RA & Veena G. A comparative study of lung cancer detection using machine learning algorithms. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). 2019, February; (pp. 1-4). IEEE.
23. Raof SS, Jabbar MA & Fathima SA. Lung Cancer prediction using machine learning: A comprehensive approach. In 2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA). 2020, March; pp. 108-115. IEEE.
24. Rumelhart DE, Hinton GE & Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986; 323(6088), 533-536.
25. Samuel AL. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*. 1959; 3(3), 210-229.
26. Singh VK, Singh D & Singh TN. Prediction of strength properties of some schistose rocks from petrographic properties using artificial neural networks. *International Journal of Rock Mechanics and Mining Sciences*. 2001; 38(2), 269-284.
27. Stern HS. Neural networks in applied statistics. *Technometrics*. 1996; 38(3), 205-214.
28. Xie Y, Meng WY, Li RZ, Wang YW, Qian X, Chan C, ... & Leung ELH. Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational oncology*. 2021; 14(1), 100907.
29. Yegnanarayana B. *Artificial neural networks*. PHI Learning Pvt. Ltd. 2009.
30. Bilgin B. *Machine Learning Algorithms*. Daisy Publishing. 2018.
31. Mitchell TM. *Machine*. New York: McGraw-hill. 1997; Vol. 1, No. 9.
32. Tsai CW, Lai CF, Chao HC & Vasilakos AV. *Big Data Analytics*. 2016.
33. Al-Jarrah OY, Yoo PD, Muhaidat S, Karagiannidis GK & Taha K. Efficient machine learning for big data: A review. *Big Data Research*. 2015; 2(3), 87-93.
34. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O & Akinjobi J. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*. 2017; 48(3), 128-138.
35. Razzak MI, Naz S & Zaib A. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of decision making*. 2017; 323-350.
36. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R & Muharemagic E. Deep learning applications and challenges in big data analytics. *Journal of big data*. 2015; 2, 1-21.

37. Dritsas E & Trigka M. Lung cancer risk prediction with machine learning models. *Big Data and Cognitive Computing*. 2022; 6(4), 139.
38. Sharma A & Mittal S. Comparative Analysis of Various CNN Models for Lung Cancer Prediction. In 2024 3rd International Conference for Innovation in Technology (INOCON). 2024, March; pp. 1-6. IEEE. rce: Add source