Prediction of Lung Cancer Using Supervised Machine Learning

Moshera ElGohary[1], Hatem Abdelkader[2], Asma El said [3]

Department of information system, Faculty of computers and information
Menofia university
Menofia, Egypt
mosheraelgohary2020@gmail.com[1], hatem.abdelkader@ci.menofia.edu.eg [2], asmaa.elsayed@ci.menofia.edu.eg [3]

***Abstract*— Cancer of the lungs is a silent monster. It is discovered when it is far advanced, such as liver or pancreatic cancer. It can be difficult for doctors to recognize the disease in the beginning stages. For this reason, we focus on this topic, to help doctors and people to determine their cancer risk at a lower cost through an effective cancer prediction system and make appropriate decisions according to their cancer risk status. This paper's goal is to make a practical method for determining whether a patient has lung cancer or not. The proposal was tested with the Kaggle standardized data set Survey Lung Cancer. we used real data collected from real hospitals in Egypt. In our proposal, the two main processes are data pre-processing and prediction. Data preparation for the prediction process is known as data preprocessing. In the prediction process, we used techniques for machine learning to compare classifications between all these algorithms, which included Decision tree, Logistic regression, KNN, Support vector machine, and Naïve Bayes. The four criteria utilized for evaluating the techniques were accuracy, recall, precision, and F1-score. They were used to categorize the dataset, and the results were compared. The support vector machine achieved a maximum prediction accuracy of 98%.**

Keywords*— *Lung cancer, supervised Machine learning, ML algorithms, ML preprocessing and classifications.

I. Introduction

Lung cancer is a silent malignancy, and most cases are discovered at a late stage. It is the second most common cancer in both men and women, but it is also the leading cause of cancer death. It is usually detected in its final stage. According to the International Cancer Society, there will be around 238,340 new cases of lung cancer (117,550 men and 120,790 women) and approximately 127,070 fatal cases of lung cancer (67,160 men and 59,910 women) in the United States in 2023. Although there are Predicting the likelihood of the disease can be done by studying the disease's signs, it's hard to say if one has lung cancer unless one gets tested, it's very costly. but certainly, Predicting the possibility of the disease can be done by observing the symptoms of the disease. Early prediction of lung cancer can help reduce the effects of lung cancer, and possibly even mortality.

Machine learning techniques and their algorithms were used in this proposal, which is a field that relies on learning the machine to do tasks by itself. The machine learning techniques are classified into two types: supervised and unsupervised. In this study, supervised machine learning was applied. Our research aims to improve lung cancer prediction through supervised machine learning and the expansion of previous studies in this field to identify some relevant risk factors for lung cancer, to develop the previous research related to this topic, and also to know the main symptoms that will affect lung cancer patients also to compare supervised machine learning algorithms to conclude the best conclusion to predict the best results to enhance lung cancer prediction, and also to study some real cases in Egypt. Most of the previous work results were bad in the results of accuracy due to inaccuracy in the steps of the pre-processing dataset and selecting the important attributes. In this research, we get the best accuracy because of selecting the important attributes and the steps of pre-processing the dataset. supervised machine learning algorithms were used, and a comparison between classifications of these algorithms like Logistic regression, KNN, Decision tree, Naïve Bayes, and SVM was compared by splitting data to training and testing and making some preprocessing. The best accuracy of prediction was 98% was achieved by the support vector machine.

The classification of supervised machine learning algorithms helps in the prediction of lung cancer. Better results search happened after understanding our data, encoding some columns, visualizing and normalizing data, data pre-processing, determining the important features, using oversampling techniques to solve imbalanced data problems, splitting data, and then building a model. Anaconda Notebook was used for predicting lung cancer patients' results. In addition, preparing data is also important to get the best results, such as importing data, making data balanced by using oversampling techniques, cleaning data by removing null and duplicated data, Feature selection, feature scaling, splitting data into training and testing, and finally studying the models then gets a confusion matrix to get the accuracy, Recall, precision, and f1 score. finally, after preprocessing data and the comparison between supervised machine learning results, we found the best prediction accuracy of 98% was achieved by the Support vector machine to predict the disease.

II. Related Work And Limitations

This section presents the previous works on this topic and its limitations. In a study [1]. They use statistical techniques to analyze the dataset using pivot tables in Excel [1], Apriori Analysis, and Weka Data Mining Tool for subset evaluation and classification revealing a wealth of facts. A prediction accuracy of 92.6% was attained using the training set as the basis for assessment. Data mining techniques were successfully applied as suggested [1] to accomplish the goals. An accuracy of 92.6% was achieved in prediction. the limitation is they used 6 attributes from 15 like Allergy, swallowing difficulty, Alcohol consumption, Wheezing, Coughing, and Peer Pressure [1]. they removed some features that affect the disease like smoking which is the main reason for lung cancer, shortness of breath, yellow fingers, Anxiety, Chest pain, Fatigue, and Chronic disease.

A study by [2] discovered that the artificial Neural Network model can check for lung cancer with 96.6% accuracy. They used all attributes as input. They normalized only lung cancer, age, and gender. This research showed that the neural network can detect lung cancer. The limitation is they normalized only three attributes only but they should normalize all attributes to be fitter. Normalization of data is very important to reduce unnecessary data, minimize data update errors, and simplify the query procedure. Finally, normalization can improve productivity, increase security, and save expenses beyond just standardizing data. Also, they used the EasyNN function which was only used with neural networks. A study by [3], this project aims to develop a data mining classification algorithm that can determine a patient's propensity for lung cancer based on the data set. Utilizing various scenarios, algorithms, sampling techniques, and data approaches, various models were developed using the CRISP-DM methodology and the Rapid Miner program [3]. They used three scenarios in this study, First, they used all attributes with 30% of the data used for testing and 70% used for training and the result was 92% by Neural network algorithm with the 'SMOTE' Synthetic Minority Oversampling Technique which one of the most commonly used oversampling methods to solve the imbalance problem. In the second scenario, they applied all attributes except smoking, shortness of breath, and gender, and the neural network technique achieved a result of 93%. In the third scenario, they used all the attributes except smoking,yellow fingers, fatigue, shortness of breath, and gender, and the result was 91.4 %
By the neural network algorithm [3]. Regarding the different scenarios, the s2 scenario had slightly better results. The limitations we found in these searches in a study [2] that used the EasyNN function which was only used with neural networks, but we want the results of the other algorithms too. also found that the disadvantage was they normalized the values of attributes: lung cancer, gender, and age, but they must normalize all attributes to be fitter. In a study [1] they used 6 attributes from 15 and removed important features that affected the results. Also, in a study by [3], they made 3 scenarios, and the results were better than the others but they removed important attributes that affect cancer like smoking, shortness of breath, age, and gender. So, in the proposed work, we will solve this limitation and enhance the results.

III. Proposed Work

This section presents the steps of our work for Enhancing Lung Cancer Prediction Using supervised machine learning, and dataset description, The dataset was from the Kaggle site here https://shorturl.at/vwzM4.This section includes an overview of supervised learning, performance measure indices, some details about the algorithms they used, and the details of our work.

A. *Data description and preprocessing data*

Table 1 describes the dataset we used from the Kaggle site, Survey Lung Cancer. The dataset characteristics used in the study: gender, age, smoking, swallowing difficulty, peer pressure, yellow fingers, anxiety, chronic disease, allergy, wheezing, alcohol, coughing, Fatigue, shortness of breath, chest pain, and lung cancer. A total of 309 patients were evaluated, of which 270 had lung cancer and 39 did not.

TABLE 1 Dataset Attributes Description of Kaggel Site

#Attributes Name	scope
1. Gender	M(male), F (female)
2. Age	Age of patient
3. Smoking	Yes=2, No=1
4. Yellow fingers	Yes=2, No=1
5. Anxiety	Yes=2, No=1
6. Peer pressure	Yes=2, No=1
7. Chronic disease	Yes=2, No=1
8. Fatigue	Yes=2, No=1
9. Allergy	Yes=2, No=1
10. Wheezing	Yes=2, No=1
11. Alcohol	Yes=2, No=1
12. Coughing	Yes=2, No=1
13. Shortness of Breath	Yes=2, No=1
14. Swallowing Difficulty	Yes=2, No=1
15. Chest pain	Yes=2, No=1
16. Lung cancer	Yes, No

After importing libraries and the dataset, we found 16 columns and 309 rows. Then removed duplicated data which was 33 rows and there was no null data. so, the data became 276 rows and 16 columns. Of which 238 had lung cancer and 38 did not, so the data was imbalanced, we will handle this imbalance before applying the algorithms. Then we will encode lung cancer and gender. Then we will determine the important features from the visualizations in Fig 1, It is obvious that A relation between lung cancer and some features. So, let's eliminate all of these characteristics. and clean up that dataset. Fig. 2 explains the importance of the attributes we used in the dataset which affected lung cancer.

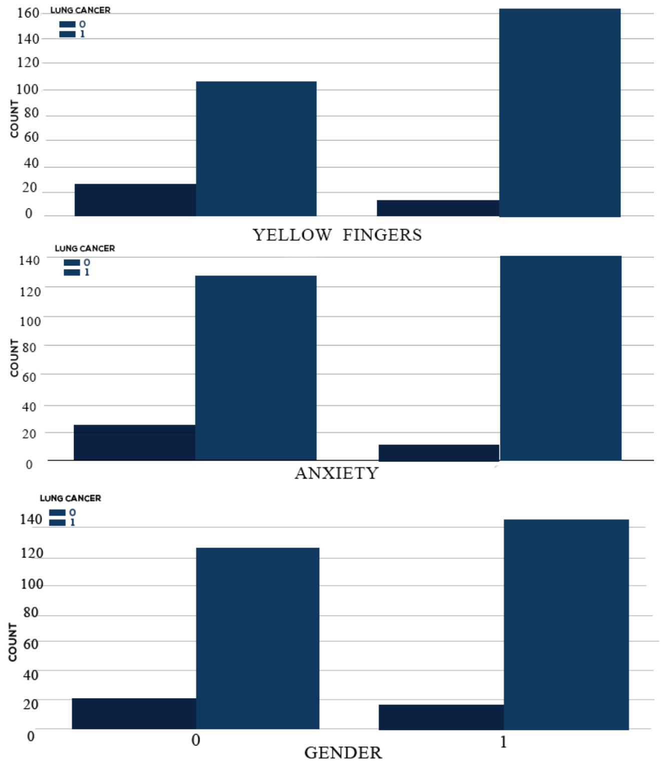

Fig. 1. *The relationship between lung cancer and some features*

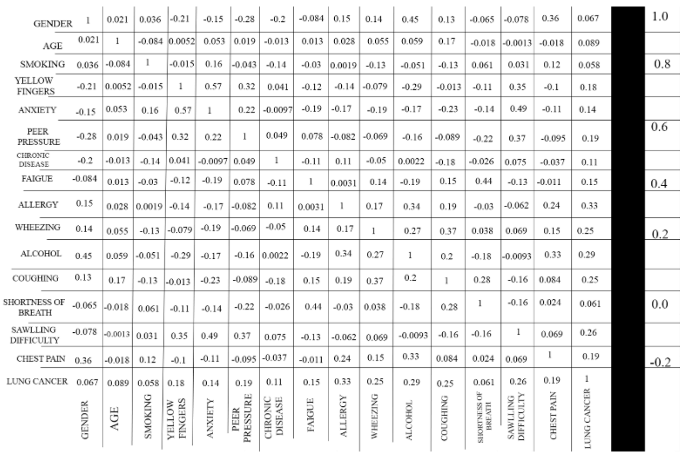

Fig. 2. *Important attributes that affected lung cancer*

After analyzing the data, we found the important attributes in Fig. 2, All attributes were important, especially smoking which is the main reason for lung cancer. Then splitting data into training and testing. Then solve the problem of imbalanced data by using oversampling techniques for imbalanced data. Then apply models. And Fig.3 shows the methodology of our work.

B. Performance Measure Indices

When we get the data, after the data clean-up and pre-processing, the first step is to pass it on to an exceptional model and, of course, to get results in probabilities. Our objective is to achieve better effectiveness and performance. The Confusion matrix is the focal point here. The Confusion Matrix is a tool for measuring performance in machine learning classification. Accuracy, precision, recall, and F1 score are the metrics used to evaluate the utilized classifiers. A confusion matrix is a table that determines the performance

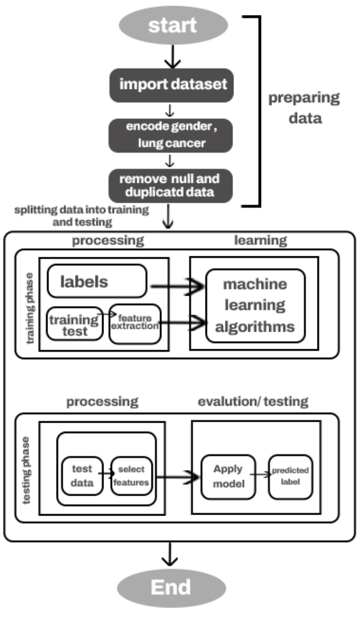

Fig. 3. *Methodology of preprocessing data of our data*

of a classification algorithm First, some terminologies are discussed:

1) True Positive (TP): The present value is positive, and the machine learning model predicts that it will remain positive.

2) True Negative (TN): When the expected and actual values have the same negative.

3) False Positive (FP): A test result that suggests an individual has a particular medical issue while the person does not.

4) False Negative (FN): a result in which the model predicts the negative class wrongly.

5) Accuracy is the level at which a measurement is close to the true or accepted value. The higher the accuracy value, the better the model's performance as defined in Eq. (1).

$$\text{Accuracy} = (TP + TN)/(TP + TN + FN + FP) \quad (1)$$

6) Precision, as defined in Eq. (2), quantifies the number of positive class predictions that are positive class predictions. It refers to how close two measurements of the same item are. Precision is distinct from accuracy.

$$precision = TP/(TP + FP) \quad (2)$$

7) Recall, also known as Sensitivity, is the proportion of correctly predicted positive observations to total positive observations. In mathematical expression; (3).

$$\text{“}sensitivity\text{” Recall} = TP\ /\ (TP+FN) \quad (3)$$

8) In Eq. (4), The F-score combines the precision and recall of the approach, which can be described as the balanced average of the model's precision and recall.

$$F1-Score = 2*(\text{Recall}*precision)\ /\ (\text{Recall}+ precision) \quad (4)$$

9) Confusion matrix as defined in Fig (4)

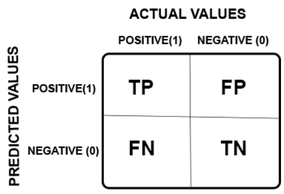

Fig. 4. *Confusion matrix*

We used supervised machine learning and its algorithms to compare classifications between all these algorithms, which include Logistic regression, KNN, Decision tree, Naïve Bayes, and SVM. Finally, supervised machine learning algorithms were used to classify the dataset, and the results were compared. The maximum accuracy of prediction of 98% was achieved by a support vector machine. with splitting data into training data and testing.

C. The proposed Supervised machine learning algorithms

Machine learning is a subfield of AI and computer science that uses algorithms and data to teach how to learn from experiences to do some tasks to improve performance measures without the need for a human. To correctly classify data or predict outcomes. Gradually improving its accuracy.

supervised learning is divided into two types of problems: classification and regression. This section briefly overviews the supervised machine learning algorithms classification used in our research.

1. *KNN algorithm*

is a non-parametric technique that categorizes data points based on their proximity to other relevant data. It attempts to compute the distance between data points, usually with Euclidean distance and then assigns a classification based on the most popular or mean group. The KNN algorithmcan rival the more accurate models because it makes very precise predictions. For predictions that don't require a human-readable model, the KNN algorithm can be used. The quality of the forecast is dependent on the distance measure.

2. *Decision tree algorithm*

It is an algorithm that can work with both discrete and continuous variables. It makes use of a tree-like graph or model of decisions and their probable outcomes. It is aconditional control statement that demonstrates an algorithm.Its nodes each contain a test on a particular feature, and the branches represent the test's results. [11].

3. *Support vector machine algorithm*

The technique employs supervised machine learning with binary sets that are linearly separable. The overarching goal is to create a hyperplane that can classify all training vectors. This produces the widest potential disparity between the two classes. The support vector machine is a linear model that can be used to solve regression and classification problems. It is useful for a wide variety of applications since it can solve both linear and nonlinear problems. SVM is based on a fundamental concept: The method produces a row or hyperlink that categorizes the data. If the dimensions surpass the number of samples, they are best used in higher-dimensional areas.

4. *Logistic regression algorithm*

It is the classification problem is just like a regression problem except that values y we now want to predict take on only a smaller number of discrete values.

5. *Naïve Bayes algorithm*

The proposition of Naïve Bayes and a supposition of predictor independence serve as the foundation of this bracket strategy. To put it simply, a Naive Bayes classifier believes that the actuality of a single trait in a class doesn't affect the actuality of any other features. Naive Bayes is designed to break multi-class problems. However, it can perform better than other models and bear significantly lower training data, If the supposition that features are independent is true. Naïve Bayes is more applicable for categorical input variables than numerical variables.

IV. Results And Discussion

We compared with five supervised ML algorithms like logistic regression, KNN, Naïve Bayes, SVM, and Decision tree. A confusion matrix includes the actual and predicted classification methods performed. After calculating the confusion matrix of logistic regression is: [[٦٤ ٤] [٨ ٥٩]] and the accuracy was 9١%, the confusion matrix of the decision tree is: [[48 0] [3 44]] and the accuracy is 96%, the confusion matrix of KNN is: [[68 0] [7 60]] with 95% accuracy, the confusion matrix of naïve Bayes is: [[48 0] [3 44]] with 94% accuracy finally the confusion matrix of the support vector machine is: [[68 0] [3 64]] with 98% accuracy. We noticed that the best accuracy was achieved by the support vector machine by splitting data into training and testing. Preprocessing data is the main reason to make the results better. Table 2 shows the comparison prediction between supervised machine algorithm classifications.

TABLE 2 COMPARISON OF PREDICTION ACCURACY,AND PRECISION.RECALL.

Algorithm name	**Performance measure indices**			
	precision	***recall***	***Fi score***	***accuracy***
Logistic regression	% 91	% 91	% 91	% 91
Decision tree	% 96	% 96	% 96	% 96
KNN	%95	% 95	% 95	% 95
naïve Bayes	% 94	% 94	% 94	% 94
SVM	% 98	% 98	% 98	**% 98**

Micro-averaging gives each instance equal weight and displays the average performance across all predictions. Micro-averaged precision, recall, and accuracy are all the same in multi-class classification.

A. *Case study: Real data of Egyptian people suffering from lung cancer.*

1. *Data Characteristics and Descriptions of dataset*

According to the most recent WHO data published in 2020, the number of fatal cases of lung cancer in Egypt reached 5,677, which represents 1.06% of all fatalities. Egypt ranks 115th in the world, with an age-adjusted mortality rate of 8.02 per 100,000 people. Lung cancer ranks first among cancers in men and second in women.So, in this research study, we collected some data from real people who suffered lung cancer in Egypt in the period from [2015_2022]. We will predict the grade of lung cancer. Because of the silence of diseases, the grade was primary and moderate only. The attribute names were 'gender', 'location', 'type _of_ tumor', 'specimen','primary or secondary, and 'grade'. As shown in Table 3. We preprocessed this sample data and showed the relationship between the number of people and the age, gender, and grade of the tumor as shown in Fig .5 We found the number of males than females in Egypt according to these few people. The grade primary is more than moderate. And people who suffer from lung cancerare between [63 and 70]. That is the concentration that the lung cancer is the silent lung.

TABLE 3 SAMPLE OF DATA FOR EGYPTIAN PATIENTS

# Attributes	scope
1. gender	Male or female
2. age	Age of patient
3. location	Location of disease
4. type_of_tumor	Tumor type
5. specimen	core biopsy
6. primary or secondary	Primary or secondary
7. grade	Primary or moderate

1. *Results of the case study*

We compared a few algorithms and found a decision tree and support vector machine with 75% accuracy, KNN with 50% finally, Naïve Bayes and logistic regression with 75 % accuracy. The results were better when we used supervised machine learning

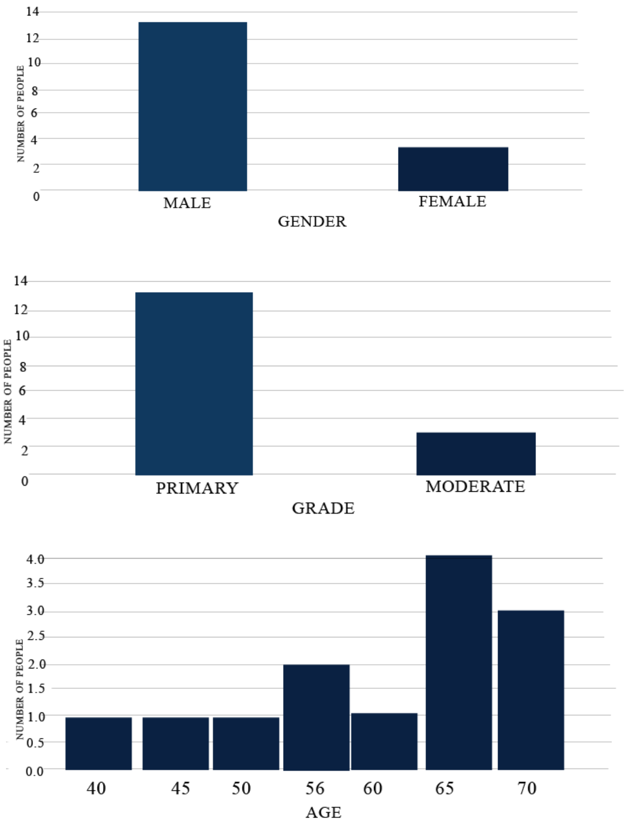

Fig. 5. *The relationship between the number of people and some attributes*

V. Conclusion And FutureWork

In this study, we enhance lung cancer prediction by using supervised machine learning classifiers. The dataset has 309 records of patients who reported having lung cancer obtained from a database. A total of 309 patients were evaluated, of which 270 had lung cancer and 39 did not. so, the data were imbalanced, so we used oversampling techniques to get the results fitter. to solve the problem of imbalanced data. Using Supervised machine learning classification by the programmed Anaconda notebook to analyze the dataset. A wealth of information was uncovered by analysis andclassification. It was learned about important variables that serve as biomarkers for the development of lung cancer. These were then incorporated into a prediction model to calculate thelikelihood that patients will get lung cancer. The training set was used to evaluate the prediction model, The support vector machine method gave 98% accurate predictions. As a result, machine learning techniques were more effective than data mining in achieving the goals. Preprocessing data and fitting itwas the primary factor in improving results.

Future work: We intend to construct a system using theEgyptian lung cancer illness dataset to integrate text mining and data mining to read any report for patients in Egyptian hospitals to determine whether or not the patients have lung cancer.

References

[1] N.Vijayalakshmi and J.PolleyAmilya, "Application of Knowledge Engineering for Prediction of Lung Cancer,"International Journal of Computer Sciences and Engineering,Vols. -6, no. -7, pp. no. 957-960, Jul-2018.

[2] S. S. A.-N. Ibrahim M. Nasser, "Lung Cancer Detection Using Artificial Neural Network," International Journal of Engineering and Information Systems (IJEAIS), no. 3, pp. 17-23, 2019.

[3] d. f. C. n. a. a. j. M. Eduarda Vieira, "Data mining Approach to classify cases of lung cancer," in WorldCIST, 2021.

[4] I. S. Z. Samy S. Abu Naser, "Knowledge-based systems that determine the appropriate of students major," World Wide Journal of Multidisciplinary Research andDevelopment, pp. 26-34, 2016.

[5] I. A. B. A. M. H. R. A. Saadaldeen Rashid Ahmed Ahmed, "Lung cancer classification using data mining and supervised learning algorithms on the multi-dimensional data set," Periodicals of Engineering and Natural Sciences, vol. 7,pp. 438-447, 2019.

[6] A. J. S. S. A. N. Massoud El Agha, "Polymyalgia Rheumatic Expert System," International Journal of Engineering and Information Systems (IJEAIS), vol. 1, no. 4,pp. 125-137, 2017.

[7] S. S. A. N. &. M. M. Al-Hanjori, "An expert system for men genital problems diagnosis and treatment," InternationalJournal of Medicine Research, pp. 83-86, 2016.

[8] S. Jadhav, "Lung Cancer Detection Using ClassificationAlgorithms," National College of Ireland,p. 1_21, 2019.

[9] E. S. N. Joshua1*, M. Chakravarthy and D. Bhattacharyya2, "An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques," International Information and engineering technology association, p. 351_359, 2020.

[10] M. B. S. Ismail, "Lung Cancer Detection and Classification using Machine Learning," Turkish Journal of Computer and Mathematics Education, vol. 12, pp. 7048- 7054, 2021.

[11] Matthew Brendel, Chang Su, Zilong Bai, Hao Zhang, Olivier Elemento, Fei Wang, “Application of Deep Learningon Single-cell RNA,” *Genomics Proteomics Bioinformatics,*vol. 20, no. 5, pp. 814-835, 2022.

[12] Fayez Eid Alazemi, Babar Jehangir, Muhammad Imran, Oh-Young Song, Tehmina Karamat, “An Efficient Model for Lungs Nodule Classification Using,” *Journal of Healthcare Engineering*, pp. 11, 2023.

[13] Yongbing Zhao, Jinfeng Shao, Yan W. Asmann, "Assessment and Optimization of Explainable Machine Learning Models Applied to Transcriptomic Data,"*Genomics Proteomics Bioinformatics,* vol. 20, no. 5, pp. 899-911, 2022.

[14] Stefan Stanojevic, Yijun Li,Aleksandar Ristivojevic,Lana X. Garmire, "Computational Methods for Single-cell Multi-omics Integration and Alignment," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, pp. 836-849, 2022.

[15] Sen Yang, Tao Shen, Yuqi Fang, Xiyue Wang, Jun Zhang, Wei Yang, Junzhou Huang, Xiao Han, "DeepNoise: Signal and Noise Disentanglement Based on ClassifyingFluorescent Microscopy Images via Deep Learning," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, pp. 989-1001, 2022.

[16] Yawei Li, Xin Wu, Ping Yang, Guoqian Jiang, Yuan Luo, “Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis,” *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, p. 850–866, 2022.

[17] N.Vijayalakshmi and J.PolleyAmilya, "Application of Knowledge Engineering for Prediction of Lung Cancer,"International Journal of Computer Sciences and Engineering,Vols. -6, no. -7, pp. no. 957-960, Jul-2018.

[18] S. S. A.-N. Ibrahim M. Nasser, "Lung Cancer Detection Using Artificial Neural Network," International Journal of Engineering and Information Systems (IJEAIS), no. 3, pp. 17-23, 2019.

[19] Yuguo Zha, Hui Chong, Pengshuo Yang, Kang Ning, "Microbial Dark Matter: From Discovery to Applications," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, p. 867–881, 2022.

[20] Haodong Xu, Zhongming Zhao, "NetBCE: An Interpretable Deep Neural Network for Accurate Prediction of Linear B-cell Epitopes," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, p. 1002–1012, 2022.

[21] Hui Wan, Liang Chen, Minghua Deng,"scEMAIL: Universal and Source-free Annotation Method for scRNA- seq Data with Novel Cell-type Perception," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, p. 939–958, 2022.

[22] Alexandra J. Lee, Dallas L. Mould, Jake Crawford, Dongbo Hu, Rani K. Powers, Georgia Doing, James C. Costello, Deborah A. Hogan, Casey S. Greene, "SOPHIE: Generative Neural Networks Separate Common and Specific Transcriptional Responses," *Genomics Proteomics Bioinformatics,* vol. 20, no. 5, p. 912–927, 2022.

[23] Yiran Shan, Qian Zhang, Wenbo Guo, Yanhong Wu, Yuxin Miao, Hongyi Xin, Qiuyu Lian, Jin Gu, "TIST: Transcriptome and Histopathological Image Integrative Analysis for Spatial Transcriptomics," *Genomics ProteomicsBioinformatics,* vol. 20, no. 5,p. 974–988, 2022.

[24] Raja E. Altarazi, Malak S. Hamad, Rawan Elbanna, DinaElborno, and Samy S. Abu-Naser, "A CLIPS-Based Expert System for Brain Tumor Diagnosis," *International Journal ofAcademic Engineering Research (IJAER),* vol. 7, no. 6, pp. 9-15, 2023.

[25] Fadi N. Qanoo, Raja E. N. Altarazi, Samy S. Abu-Naser,"A CLIPS-Based Expert System for Heart Palpitations Diagnosis," *International Journal of Academic Information Systems Research (IJAISR),* vol. 7, no. 6, pp. 10-15, 2023.

[26] Husam Abd Rahim Eleyan,Mohammed Almzainy,Shahd Albadrsawai,Samy S. Abu-Naser4, "An Expert System for DiagnosingWest Nile virus Problem Using CLIPS," *International Journal ofAcademic Information Systems Research (IJAISR),* vol. 7, no. 6, pp. 27- 37, 2023.

[27] Dina F. Al-Borno and Samy S. Abu-Naser, "A ProposedExpert System for Vertigo Diseases Diagnosis,"*International Journal of Academic Information Systems Research (IJAISR),* vol. 7, no. 6, pp. 1-9, 2023.

[28] Mohammed M. Almzainy, Shahd J. Albadrasawi, Jehad M. Altayeb, Hassam Eleyan, Samy S. Abu-Naser, "Development and Evaluation ofan Expert System for Diagnosing Tinnitus Disease," *International Journal of Academic Information Systems Research (IJAISR),* vol. 7, no. 6, pp. 46-52, 2023

[29] Walid F. Murad, Samy S. Abu-Naser, "An Expert System for Diagnosing Mouth Ulcer Disease Using CLIPS,"*International Journalof Academic Engineering Research(IJAER),* vol. 7, no. 6, pp. 30-37, 2023.

[30] Abedeleilah S. Mahmum, Nidaa Wishah, Waleed Murad, Dina F. Al-Borno, Samy S. Abu-Naser, "An Expert System for Diagnosing Whooping Cough Using CLIPS," *International Journal of Engineering and InformationSystems (IJEAIS),* vol. 7, no. 6, pp. 1-8, 2023.