IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Lung Cancer Prediction from Text Datasets Using Machine Learning

SURAJ RASTOGI[1] ,SAURABH SINGH[2] ,SHIVAM SINGH[3] ,UJJWAL KUMAR SINGH[4], RAVIN KUMAR[5]

[1]B.Tech Scholar Department of CSE, GNIOT, Greater Noida, India
[2]B.Tech Scholar Department of CSE, GNIOT, Greater Noida, India ,
[3]B.Tech Scholar Department of CSE, GNIOT, Greater Noida, India
[4]B.Tech Scholar Department of CSE, GNIOT, Greater Noida, India ,
[5]Assistant Professor Department of CSE, GNIOT, Greater Noida, India ,

Abstract

Lung cancer is the leading cause of cancer.related death in this generation and is expected to remain so in the future. Lung cancer treatment is possible if the symptoms of the disease are caught early. The design model of cancer treatment can be created using current advances in intelligence in inclusion without affecting the environment. It saves time and money as it minimizes the waste of resources and the work required to complete the book's work.Use the machine learning.based support vector machine (SVM) to improve the detection process of pulmonary clinical data. Use the SVM classifier to classify cancer patients based on their symptoms while adding patterns using the Python programming language. The performance of our SVM model was evaluated according to various variables. Multiple cancer databases from the UC Irvine library were used to test the model. Thanks to the acceptance of this research, smart cities will be able to provide better health care to their citizens.Cancer patients can access real treatment at a good price, anywhere, anytime with minimal effort and slowness. The proposed model is compared with existing SVM and SMOTE methods. Compared to the current system, the plan achieves 98.8% accuracy

1. Introduction

In other words, lung cancer is the leading cause of death among men and women worldwide [1].
According to other research, lung cancer accounted for 13 percent of all cancer diagnoses in the United States in 2015. According to the American Cancer Society, lung cancer accounts for approximately 27% of all cancers [2]. Therefore, appropriate evaluation and monitoring of pulmonary nodules in early development is necessary. In this study, researchers studied cancer development and growth using ML and DL methods to predict cancer growth and development. The predictive models discussed here have been developed using various input and data models and various supervised machine learning algorithms.
Using the image operator LBP, an image can be converted to an integer array or image, called the local binary pattern. This text is used for further image analysis and is usually displayed as a histogram. Due to its unique capability and ease of use, LBP is widely used in user fabric [2]. Histogram then uses these symbols to perform a more detailed analysis of the image. Cancer deaths from lung disease are still higher in the past three years than deaths from prostate cancer or breast cancer in both men and women [3].The main reason for this is the complexity and process of prostate and breast cancer prognostic models developed in recent years. Therefore, it is necessary to make a reliable diagnosis of lung cancer as early as possible [4]. SVM is a good predictor in both linear and non.linear scenarios and is widely used in many industries, including medicine [8]. Although SVM is a good classifier, cancer diagnostic models are still being developed [9]. Choosing the right treatment for the patient depends on the outcome of the transplant [10], which has gained prominence in clinical trials.
In addition to scanning, a direct connection can be used to detect unexpected changes during scanning. Genetic mutations in the epidermal growth factor receptor (EGFR) have been identified and can be used to identify genetic changes in lung cancer [13]. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have been shown to outperform their counterparts [14]. Because most unfair decisions

carry more weight than small decisions, the majority will de cide more than the minority. Classification algorithms base d on traditional operating systems do not perform as well as they should [15].In this paper, we optimized the detection p rocess in cancer data using the SVMbased machine learning model. Use the SVM classifier to classify cancer patients b ased on their symptoms while adding patterns using the Pyt hon programming language. Different tumors have different diagnoses. But there are only a few ways to count the popu lation in it. In this article, not only the methods for the diag nosis of cancer, but also the studies that need to be done to calculate their size, shape and location will be discussed. Thus, not only can tumors be detected, their types can be ea sily identified by counting and retrieving, and the correct in structions for handling them can be calculated.

2. Related Works

The authors [3] conducted an experiment to evaluate the effect of referral in patients with suspected cancer and the adverse effects of delay in delivery of prompt treatment, and to determine whether the delay is disease-related. Stage and Tumor features, Formed organizations and many extensions are corrected again. A total of 565 patient revision charts were collected for this study. Altogether, 51 percent of the participants had lung enlargement, while the other half (Step 8).5% had multiple injuries and 111 (19.6%) of these had radiological abnormalities that were not life-threatening. In the case of hemoptysis, the first-line waiting time is shorter than in other conditions. During the policy making process, RODP was created to facilitate the review process. It is estimated that most patient delays are due to delays in primary and secondary care.[11] investigated several methods to measure lung growth. There are several of them, including the use of neural networks, image processing, linear correlation analysis (LDA), and self-maps (SOM). In conclusion, it is suggested to use SVM as a representation technique. When you use machine learning, you use vector machines to look at data and recognize patterns. At the beginning of his research [10], he developed a method for investigating lung enlargement.In this way, the data is preprocessed to start the image development process. When the data is ready to be tested in both data mining and nervous systems, which is important for distinguishing between different methods for rehabilitation, this is the main language they can try. Using a backpropagation neural network to classify images as malignant or benign, researchers have successfully developed a predictive network (BPNN). Those who practice examining the stage of malignancy help themselves during diagnosis. This study [20] uses a web-based biomarker analysis and is set up to improve quality to discover and suggest signatures associated with lung cancer and related pathways.They found that in addition to what previous discoveries in these areas had predicted, the data revealed many new and surprising features of the body's replication ability that previous findings in these areas had not predicted. [21] developed a web-based system for managing smoking recommendations and categorized the advantages associated with lung tumor survival and the advantages associated with the non-smoker group, and analyzed all lung tumor survival outcomes and lung tumor survival outcomes. group was well related to the non-smoker group. Six major markers associated with smoking have been found to predict lung cancer risk and survival. Smokers can see and identify lung enlargement if these positive signs are used. [22] used data mining and simplification methods to extract findings from big data to investigate lung enlargement.To improve integration, feed-forward and feedback neural networks are used to sort objects. It cannot be denied that the author [5] is working on the use of various artificial intelligences to diagnose diseases and provide medical treatment [5]. Artificial neural networks (ANNs) is used to detect breast cancer data. Data from microarrays and the UCI machine learning repository can be used to detect cancer using an ANN-like multilayer feedforward neural network. Back propagation rules are used to plan the system.Using cross validation, different files with different numbers of hidden layers and centers connected to the same data can be tested against each other. If an event (chest pain) occurs from the UCI dataset, the accuracy of the framework should be improved due to the various combinations of latent processes and affiliated centers. The number of hubs and layers in the NCBI dataset has increased, increasing the accuracy of the analysis. If you use a similar nervous system, you can predict how sick the patient will be. This can be done with the help of automated decision making.[24] used computer aided analysis, fuzzy weighted preprocessing and anti-fraud system in all studies. The framework is divided into three phases. Although the document contains 57 key points, only four of them can be checked by conventional analysis. A weighting method using a fuzzy weight-based prior was used as a preparatory step before using the previous distribution. Third, fake security certificates are used to identify counterfeit products.The lung dataset was used to test the programmed approach to tumor analysis developed by the researchers. Seeing the performance base at 100% is encouraging for future integration, which is great news. According to [25], an opinion-based recommendation based on the integration of users from social network information, explicit knowledge and information requirements from user reviews can be used to generate recommendations [26]. With this technology, the accuracy of forecasts and recommendations can be increased. The authors [26] developed a reliable estimate to reduce the probability.Although this method does not provide as much efficiency as more deterministic methods, it provides a more efficient design while maintaining learning speed. The pi-sigma organization thought that they would use a rare marginal polynomial in their system of marginal polynomial, which is a rare marginal polynomial. The possibility of a different name type is expressed by the RPN and considered as the RPN response, contained within this framework. By using RPN, additional development methods are provided for specific processes. Based on the evidence, the system appears to have a high computational efficiency allowing optimization and continuous learning.[27] have found a way to predict pneumonia, detect it early, and treat pneumonia at a young age. Some important parts of the picture have been omitted to better prepare for pneumonia. Efficiency-based modeling has been shown to have an impact on the ability to predict lung growth. Imaging techniques were used to evaluate previous experience with lung oncology needs. Computer-based imaging tools are available to predict and manage lung enlargement and can help you do this.The literature [29] proposed the evolutionary process of genetic and molecular clustering using the MLPNN algorithm to analyze the features of lung CT images. It has been shown to be a reliable source of information compared to a CT scan of the lungs. Using visual acuity for anxiety is associated with better results and using preliminary images to aid

subtraction. We can extract this brightness using the MAD method. Personal information is selected with the help of GAPSO.The final image is created using the GAPSO-MLPNN algorithm. As a starting point, many of the test materials have good geometrical accuracy, high Classification accuracy and low error rate that goes for everything. In addition to its excellent performance, the technology can also identify lung diseases. [30] developed a model to predict the value of the recovered material via neural networks. Next, the author introduces a self-transforming trading strategy that follows futures trading rules and test data. Finally, new ideas are compared to traditional methods to show how their ideas have evolved.Tests show that his answers outperformed those of other researchers on a standardized test. Its strategy outperformed its competitors in terms of rewards and risks. After [31], a hybrid feature extraction was proposed, thereby improving the security of ECG evidence. A further development is the creation of a common recognition standard for ECGs, which improves the recognition of performance at different ECG positions. After testing, it was found that the authentication process was successful.The authors of [32] proposed a hybrid system that they tested on the real Sina Weibo dataset. Those who have seen the results say that the new hybrid consensus algorithm outperforms its predecessors in terms of performance and outperforms other algorithms. An overview of the current machine learning model used in this study is shown in Figure 1.

3. Proposed Method

In this section, the SVM classifier helps classify cancer patients based on their symptoms. Preprocessing in the plan takes place before data is collected. Selected candidates were trained and evaluated a second time on the benchmark data using a 10-fold cross-validation procedure. Analyze and evaluate data to determine the best way to diagnose cancer. Figure 2 shows the high-level process structure.All previously provided information is provided as an entry to the order. The difference between the data on which the inputs are based also requires testing a large number of different volumes of data and extracting the data needed based on their results. This preprocessing transforms and processes the input data into desired groups. The information for this study is divided into several groups. Of these Division operations that calculate and divide the basic structure size and value operation of data. After finally splitting them, delete their children's apps.

3.1. Data Acquisition. In this research, we used data from the online database Lung Cancer at the University of California, Irvine. In total, the dataset contains 32 samples, 57 features, and an attribute class. The main purpose of our work is to evaluate and compare the performance of SVMs and others.

3.2. Data Preprocessing. Preprocessing is the first step in lung cancer diagnosis, which should fill a gap in the data and eliminate redundant data. Therefore, missing values are calculated using the three-neighbor K-Nearest Neighbors method to increase the reliability of all data. Training and sample testing are required.

3.3. Training and Testing Samples. The input data is first trained and then tested using a neural network, which is a type of artificial intelligence. The weight of the neural network is determined by the input data at the beginning of the process. A neural network is trained on the sample dataset and then evaluated on the same data used for training.During deployment, data is weighted to determine how much of the problem occurs or the error rate, and errors are corrected by re-weighting the dataset.

3.4. Feature Extraction. For a large number of features to be extracted from a dataset to lessen the complexity of detecting

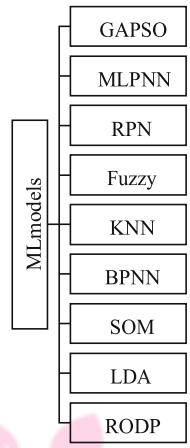

Fig 1:- ML models

Data for lung cancer should be segmented. This test will remove tumors from the lungs due to the growth of cancer cells (features). This feature extraction method uses the PSO programming language. Extracting features from the input data is an essential part of the pattern recognition algorithm used to store data to obtain more useful and trivial features and to predict patients' Cancer-related literature for translation and distribution of reference data. SVM.

3.5. Classification Using SVM.

Classification is the process of dividing data into logical groups. Decisions can be made based on both structured and raw data that the system can identify. Malware and Denial of Service (DoS) attacks can be prevented and mitigated using mining techniques.This means that the text is organized and categorized to ensure its integrity is always maintained.

In addition to the standard classification of tumors as malignant (M) or benign (B), a new classification called precancerous lesions has been added to category B. Patients who fall into this category will receive more care and treatment. When it comes to information security, categories such as copying, sending and storing information will always be under surveillance for suspicious activities.

The practice of classifying useful information to improve its usability and discovery is called taxonomy.It helps you save on storage and backup costs by reducing data backups. Therefore, the working time can be reduced in some cases. To distinguish between different objects, an SVM is an algorithm that teaches the machine to learn to distinguish between them by itself. Since the edges of the plane are determined from those closest to it, the best SVM model will have a hyperplane that divides the class as widely as possible.

The SVM primary goal is to maximize the profit margin as much as possible, Given the dataset ^{x} and a featureThe n-dimensional vector is given by the distribution yi∈ f–1.1g,

where i=1,⋯,N denotes the data N for the distribution. When constructing a classifier, formula (1) uses the normal hyperplane weight vector ω, defined as a rescaled hyperplane classifier satisfying formula (2), and the weight vector is determined by subtracting the weight vector from the weight vector. as the weight vector in equation (3). can solve Equation (3) using the Lagrangian function in Equation (4), which results in ω = ∑iαiyixi and b = yi −ωTxi as solutions. using for denote the Lagrange multipliers for each inequality constraint, which stands for inequality constraint multipliers.

$f(x) = \text{sign}(\omega x + b)$, **…(1)**

(1) $y_i\omega^T x_i + b \geq 1$, $i = 1, \cdots, N$, (2) minimize **..(1)**

$$\frac{1}{2}\omega^T\omega s.t. y_i\left(\omega^T x_i + b\right) \leq 1, i = 1, \cdots, N, (3) \quad \text{..(2)}$$

$$L(\omega, b, \alpha) = \frac{1}{2}\omega^T\omega + \sum_i \alpha_i\left(1 - y_i\left(\omega^T x_i + b\right)\right) \quad \text{..(3)}$$:

(4)

Since the data are inseparable, change equation (2) to equation (3) by including the loose variable ξ in equation (5). This is done by multiplying the 0.5 ωTω spark constant C to obtain equation (6). Misallocation of training instances may result in a permanent C penalty to your account. For a lower value of C, the optimization will choose a hyperplane

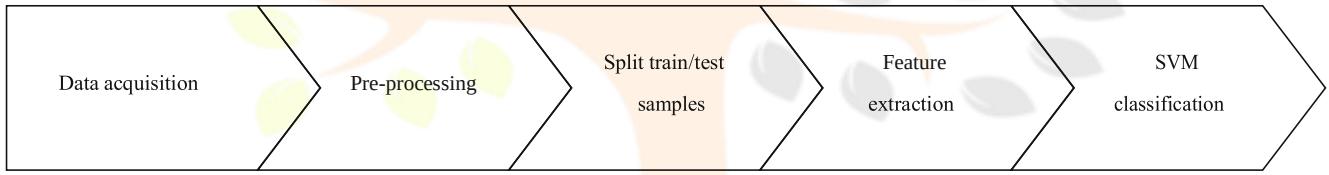

Fig 2:-SVM WORKING

with a large margin of safety, causing more points to be unclassified.

$y_i\omega^T x_i + b \geq 1 - \xi_i$, $i = 1, \cdots, N$, (5) ..(5)

minimize $\frac{1}{2}\omega^T\omega + C\xi_i s.t. y_i\left(\omega^T x_i + b\right) \leq 1 - \xi_i, i = 1, \cdots, N$:
..(6)

When combined with the kernel function, the linear hyperplane can be used to separate vectors that are not linearly separable. Nonlinear DVM classifiers that are not linear in nature often use the radial basis function (RBF) kernel as the kernel function. We can do this from the custom ϕ(x) map, ie. K(xi, xj) = ϕ(xi)Tϕ(xj). In equation (7) below, we see the RBF kernel. We use the formula γ = 1/2σ2 to reduce the term in the equation.The Lagrangian function can be used to construct Equation (8); where α is the solution of the dual problem in Equation (7), which is the dual solution in Equation (9). Now we can relax (10), we can finally arrive at the result predicted by equation (10).

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \quad \text{..(7)}$$

$$b = y_i - \sum_j \alpha_j y_j K(x_j, x_i) \forall i \cdot \alpha_i > 0, \quad \text{..(8)}$$

$$\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j K(x_i, x_j), \text{ s.t.: } \sum_i \alpha_i y_i = 0' \quad \text{..(9)}$$

$$\omega^T x + b = \sum \alpha_i y_i K(x_i, x) + b: \quad \text{..(10)}$$

As for binary classification problems, it is common practice to use SVM methods to solve them. A one-to-one strategy is proposed in which a binary model is created for each predicted class and predictions are made for this class based on the reliability of the binary model compared to all other classes.

4. Data Availability

Kaggle is a great platform containing real time data and also gives a balanced data format for the proper execution of the process. To make this project effective we use a dataset named as "Survey Lung Cancer". This dataset contains the Excel file which contains the survey of three hundred ten people including both male and female. In survey the male are fifty two percent and female is forty eight percent. The data set contains various columns which contains symptoms due to which a person can predicted that those person has lung cancer or not. The symptoms are smoking, yellow fingers, anxiety , peer pressure, chronic disease, fatigue, allergy, wheezing, alcohol consuming , coughing , shortness of breath , swallowing difficulty , chest pain .On the basis of that symptoms the SVM classify that the person have the lung cancer or not. In dataset 1 is denoted as no and 2 is denoted as yes.

1	GENDER	AGE	SMOKING	YELLOW_F	ANXIETY	PEER_PRE	CHRONIC	FATIGUE	ALLERGY	WHEEZIN	ALCOHOL	COUGHIN	SHORTNE	SWALLOW	CHEST PA	LUNG_CANCER
2	M	69	1	2	2	1	1	2	1	2	2	2	2	2	2	YES
3	M	74	2	1	1	1	2	2	2	1	1	1	2	2	2	YES
4	F	59	1	1	1	2	1	2	1	2	1	2	2	1	2	NO
5	M	63	2	2	2	1	1	1	1	1	2	1	1	2	2	NO
6	F	63	1	2	1	1	1	1	1	2	1	2	2	1	1	NO
7	F	75	1	2	1	1	2	2	2	2	1	2	2	1	1	YES
8	M	52	2	1	1	1	1	2	1	2	2	2	2	1	2	YES
9	F	51	2	2	2	2	1	2	2	1	1	1	2	2	1	YES
10	F	68	2	1	2	1	1	2	1	1	1	1	1	1	1	NO
11	M	53	2	2	2	2	2	1	2	1	2	1	1	2	2	YES
12	F	61	2	2	2	2	2	2	1	2	1	2	2	2	1	YES
13	M	72	1	1	1	1	2	2	2	2	2	2	2	1	2	YES
14	F	60	2	1	1	1	1	2	1	1	1	1	2	1	1	NO
15	M	58	2	1	1	1	1	2	2	2	2	2	2	1	2	YES
16	M	69	2	1	1	1	1	1	2	2	2	2	2	1	1	2 NO
17	F	48	1	2	2	2	2	2	2	2	1	2	2	2	1	YES
18	M	75	2	1	1	1	2	1	2	2	2	2	2	1	2	YES
19	M	57	2	2	2	2	2	1	1	1	2	1	1	2	2	YES
20	F	68	2	2	2	2	2	2	1	1	1	2	2	1	1	YES
21	F	61	1	1	1	1	2	2	1	1	1	1	2	1	1	NO

survey lung cancer (1)

Fig 3:-DATASET

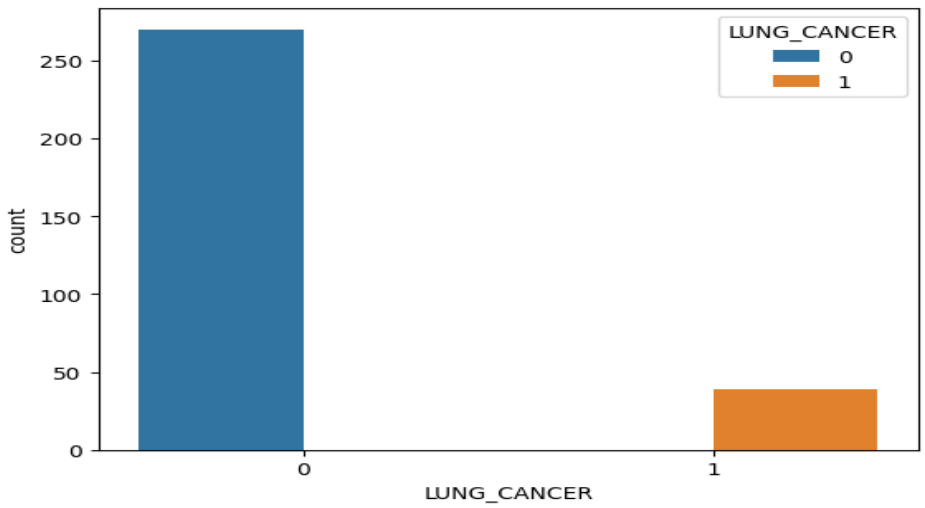

Fig 4:-Graph Representation of dataset

4. Conclusion

Lung cancer is the most generally analyzed kind of malignant growth. As the quantity of cases develops every year, powerful, fast, and early detection of cancer is important. If lung cancer is recognized in beginning phases, it very well might be dealt with without any problem. The removal of lung cancer in the final stages is costly, while in beginning phases lung cancer are simple and affordable to treat. To repeat, this project was fully intent on using support vector machine (SVM) to analyze and recognize lung cancer from symptoms which a person have. It moreover investigated the reliability of binary model as compared to other classes by the strength of SVM . The best model, to be specific our machine learning model, is working with an accuracy of 90.01%..

7. References

[1] P. Brocken, B. A. Kiers, M. G. Looijen-Salamon et al., "Timeliness of lung cancer diagnosis and treatment in a rapid outpatient diagnostic program with combined 18FDG-PET and contrast enhanced CT scanning," Lung Cancer, vol. 75, no. 3, pp. 336–341, 2012.

[2] P. Vivekanandan, "An efficient SVM based tumor classification with symmetry non-negative matrix factorization using gene expression data," in In 2013 International conference on information communication and embedded systems (Icices), pp. 761–768, IEEE, USA, February 2013. [3] J. D'Cruz, A. Jadhav, A. Dighe, V. Chavan, and J. Chaudhari, "Detection of lung cancer using backpropagation neural networks and genetic algorithm," Computing Technologies and Applications, vol. 6, pp. 823–827, 2016.

[4] J. Shen, J. Wu, M. Xu, D. Gan, B. An, and F. Liu, "A hybrid method to predict postoperative survival of lung cancer using improved SMOTE and adaptive SVM," Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 2213194, 15 pages, 2021.

[5] S. Mandal and I. Banerjee, "Cancer classification using neural network," International Journal, vol. 172, pp. 18–49, 2015.

[6] D. M. Abdullah, A. M. Abdulazeez, and A. B. Sallow, "Lung cancer prediction and classification based on correlation selection method using machine learning techniques," Qubahan Academic Journal, vol. 1, no. 2, pp. 141–149, 2021.

[7] F. Taher, N. Prakash, A. Shaffie, A. Soliman, and A. El-Baz,
"An overview of lung cancer classification algorithms and their performances," IAENG International Journal of Computer Science, vol. 48, no. 4, 2021.

[8] A. Jaweed and F. Siddiqui, "Implementation of machine learning in lung cancer prediction and prognosis: a review," in Cyber Intelligence and Information Retrieval, pp. 225–231, India, 2022.

[9] V. N. Jenipher and S. Radhika, "SVM kernel methods with data normalization for lung cancer survivability prediction application," in In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 1294–1299, IEEE, Canada, February 2021.

[10] V. A. Binson, M. Subramoniam, Y. Sunny, and L. Mathew, "Prediction of pulmonary diseases with electronic nose using SVM and XGBoost," IEEE Sensors Journal, vol. 21, no. 18, pp. 20886–20895, 2021.

[11] B. R. Manju, V. Athira, and A. Rajendran, "Efficient multi-level lung cancer prediction model using support vector machine classifier," in In IOP Conference Series: Materials Science and Engineering, vol. 1012, India, 2021no. 1, Article ID 012034IOP Publishing.

[12] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, and D. Rathee, "A hybrid algorithm for lung cancer classification using SVM and neural networks," ICT Express, vol. 7, no. 3, pp. 335–341, 2021.

[13] G. Yin, Y. Song, X. Li et al., "Prediction of mediastinal lymph node metastasis based on 18F-FDG PET/CT imaging using support vector machine in non-small cell lung cancer," European Radiology, vol. 31, no. 6, pp. 3983–3992, 2021.

[14] Y. Xie, W.-Y. Meng, R.-Z. Li et al., "Early lung cancer diagnostic biomarker discovery by machine learning methods," Translational Oncology, vol. 14, no. 1, article 100907, 2021.

[15] M. Yakar, D. Etiz, M. Metintas, G. Ak, and O. Celik, "Prediction of radiation pneumonitis with machine learning in stage III lung cancer: a pilot study," Technology in Cancer Research & Treatment, vol. 20, p. 153303382110163, 2021.

[16] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," Expert Systems with Applications, vol. 164, article 113981, 2021.

[17] S. Vijh, R. Sarma, and S. Kumar, "Lung tumor segmentation using marker-controlled watershed and support vector machine," International Journal of E-Health and Medical Communications (IJEHMC), vol. 12, no. 2, pp. 51–64, 2021.

[18] S. S. Ashwini, M. Z. Kurian, and M. Nagaraja, "Lung cancer detection and prediction using customized selective segmentation technique with SVM classifier," in Emerging Research in Computing, Information, Communication and Applications, pp. 37–44, Springer, Singapore, 2022.

[19] H. A. Miller, X. Yin, S. A. Smith et al., "Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data," Lung Cancer, vol. 156, pp. 20–30, 2021.

[20] X. Fang, M. Netzer, C. Baumgartner, C. Bai, and X. Wang, "Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer," Cancer Treatment Reviews, vol. 39, no. 1, pp. 77–88, 2013.

[21] N. L. Guo and Y. W. Wan, "Pathway-based identification of a smoking associated 6-gene signature predictive of lung cancer risk and survival," Artificial Intelligence in Medicine, vol. 55, no. 2, pp. 97–105, 2012.

[22] V. Krishnaiah, G. Narsimha, and N. S. Chandra, "Diagnosis of lung cancer prediction system using data mining classification techniques," International Journal of Computer Science and Information Technologies, vol. 4, no. 1, pp. 39–45, 2013.

[23] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," Computer Methods
and Programs in Biomedicine, vol. 113, no. 1, pp. 202–209, 2014.

[24] K. Polat and S. Güneş, "Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer," Expert Systems with Applications, vol. 34, no. 1, pp. 214–221, 2008.
[25] Y. Qian, Y. Zhang, X. Ma, H. Yu, and L. Peng, "EARS: emotion-aware recommender system based on hybrid information fusion," Information Fusion, vol. 46, pp. 141–146, 2019.
[26] Y. Shin and J. Ghosh, "Ridge polynomial networks," IEEE Transactions on Neural Networks, vol. 6, no. 3, pp. 610–622, 1995.
[27] F. Taher and R. Sammouda, "Lung cancer detection by using artificial neural network and fuzzy clustering methods," in In 2011 IEEE GCC conference and exhibition (GCC), pp. 295– 298, IEEE, United Arab Emirates, February 2011.
[28] Q. N. Tran, "A novel method for finding non-small cell lung cancer diagnosis biomarkers," BMC Medical Genomics, vol. 6, Supplement 1, p. S11, 2013.
[29] T. Kaur and N. Gupta, "Classification of lung diseases using particle swarm optimization," International Journal of Advanced Research in Electronics and Communication Engineering, vol. 4, no. 9, pp. 2440–2446, 2015.
[30] S. Xiao, H. Yu, Y. Wu, Z. Peng, and Y. Zhang, "Self-evolving trading strategy integrating internet of things and big data," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2518–2525, 2018.
[31] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, "PEA: parallel electrocardiogram-based authentication for smart healthcare systems," Journal of Network and Computer Applications, vol. 117, pp. 10–16, 2018.
[32] Y. Zhang, Z. Tu, and Q. Wang, "TempoRec: temporal-topic based recommender for social network services," Mobile Networks and Applications, vol. 22, no. 6, pp. 1182–1191.