FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 4 | Pages 888-894 | e-ISSN: 2347-470X

Optimized Feature Selection and Image Processing Based Machine Learning Technique for Lung Cancer Detection

Dr. P. Nancy[1], S Ravi Kishan[2], Kantilal Pitambar Rane[3], Dr. Karthikeyan Kaliyaperumal[4*], Dr. Meenakshi[5] and I Kadek Suartama[6]

[1]*Assistant Professor, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur Campus, India, nancysundar09@gmail.com*
[2]*Associate professor, Department of CSE, V R Siddhartha Engineering College, Vijayawada, India, suraki@vrsiddhartha.ac.in*
[3]*Professor, Department of Electronics and Communication, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India, kanthilal@klh.edu.in*
[4]*Associate Professor, IT @ IoT - HH Campus, Ambo University, Oromia Regional State, AMBO, Ethiopia, karthikeyan@ambou.edu.et*
[5]*Assistant Professor, GD Goenka University Sohna, Haryana, India, mt6458@gmail.com*
[6]*Universitas Pendidikan Ganesha, Indonesia, ik-suartama@undiksha.ac.id*

*__Correspondence:__ Dr. Karthikeyan Kaliyaperumal; karthikeyan@ambou.edu.et

ABSTRACT- The primary contributor to lung cancer is an abnormal proliferation of lung cells. Tobacco usage and smoking cigarettes are the primary contributors to the development of lung cancer. The most common forms of lung cancer fall into two distinct types. Non-small-cell lung cancers and small-cell lung cancers are the two primary subtypes of lung cancer. A computed tomography, or CT, scan is an essential diagnostic technique that may determine the kind of cancer a patient has, its stage, the location of any metastases, and the degree to which it has spread to other organs. Other diagnostic tools include biopsies and pathology tests. The creation of algorithms that allow computers to gain information and abilities by seeing and interacting with the world around them is the core emphasis of the field of machine learning. This article demonstrates how to detect lung cancer via the use of machine learning by using improved feature selection and image processing. Image quality may be improved with the help of the CLAHE algorithm. The K Means technique is used in order to segment a picture into its component components. In order to determine which traits are beneficial, the PSO algorithm is utilised. The photos are then categorised using the SVM, ANN, and KNN algorithms respectively. It uses images obtained from a CT scan. When it comes to detecting lung cancer, PSO SVM provides more accurate results.

Keywords: Lung Cancer detection, Support Vector Machine, Particle Swarm Optimization, CLAHE, K Means algorithm.

ARTICLE INFORMATION
Author(s): Dr. P. Nancy, S Ravi Kishan, Kantilal Pitambar Rane, Dr. Karthikeyan Kaliyaperumal, Dr. Meenakshi and I Kadek Suartama;
Received: 23/05/2022; **Accepted**: 30/08/2022; **Published**: 30/10/2022;
e-ISSN: 2347-470X;
Paper Id: 0822SI-IJEER-2022-05;
Citation: 10.37391/IJEER.100423
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-10/ijeer-100423.html
Crossref member CROSSREF.ORG THE CITATION LINKING BACKBONE

This article belongs to the Special Issue on **Complexity and Uncertainty in Big Data Analytics, and Machine Learning in Real-World Applications**

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The uncontrolled growth of cells in the lung tissues is what leads to the development of lung cancer. The extended use of tobacco products or the smoking of cigarettes is the risk factor that increases the likelihood of acquiring lung cancer more than any other. In addition to the primary classification, there is also the possibility of further subdividing patient instances of lung cancer into two subgroups. It is shown in *figure 1*.

There are two primary subtypes of lung cancer: small-cell and non-small-cell lung cancer. Squamous cell carcinoma, adeno carcinoma, and large cell carcinoma are all subtypes of non-small cell lung cancer. The non-small cell carcinoma subtype of lung cancer is the most frequent form of the disease. Cancer of the oat cells is a kind of lung cancer that develops in the airway cells of the lungs' tiny airways. This particular kind of cancer has a very high risk of metastasizing, or spreading, to other organs in the body. Additionally, there is a potential of cancerous growths. It is possible to determine whether or not a tumour is benign by determining whether or not it has spread to other parts of the body and whether or not it can be removed surgically. Cancerous tumours have the ability to infect the tissue around them and metastasis, which is another word for spreading to other areas of the body. When cancer cells move to other organs in the body, this process is referred to as metastasis [1] [2].

A computed tomography [3] [4], or CT, scan is an essential diagnostic technique that is used to determine the existence of cancer, its extent, and the possible influence that cancer may have on other organs in the body, in addition to providing a conclusive diagnosis of cancer itself. It does this by creating a three-dimensional depiction of the inside of a human body with the use of an X-ray machine.

Review Article | Volume 10, Issue 4 | Pages 888-894 | e-ISSN: 2347-470X

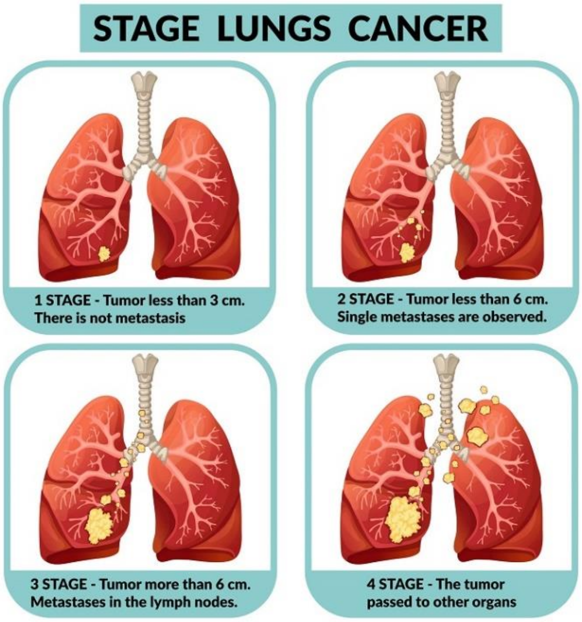

Figure 1: Lung cancer stages

The images are then assembled by the computer into a cross-sectional view, in which any abnormalities or tumours are brought to the user's attention. After that, the medical team will be able to determine which treatment method will be most effective for each specific patient.

Researchers' primary goal in the subfield of computer science known as "machine learning" is the development of algorithms that, when applied to massive data sets, may impart new knowledge to computing devices [6] [7]. The primary objective of machine learning is to automate the process by which these systems acquire the knowledge they need to recognize intricate patterns in data and to respond intelligently in response to those patterns. This goal is accomplished by teaching computers to learn from their own mistakes. The difficulty, however, is that the available training data cannot adequately reflect the variety of inputs since there are just too many of them. The most important goal for a student is to be able to generalize their knowledge from particular examples. While the training data come from some unknown probability distribution, the learner has to identify something typical, something about the distribution, that enables it to deliver meaningful answers for future occurrences.

Cancers may be difficult to diagnose due to the complex and multidimensional anatomy of the human lung, which can make the procedure very difficult. The processing of the data that is produced from computed tomography (CT) scans on a regular basis in vast numbers demands a technology that has been highly developed [8] [9]. In the case of images of the lungs, radiologists expend a large amount of work trying to identify and categories cancers, despite the fact that the results are prone to variance. They do this despite the fact that the findings are not consistent. After the scan has been completed, using a CAD system to locate and outline the tumour in the image is a method that is both effective and accurate in its execution. The manual segmentation of lung cancer is a difficult and time-consuming operation, particularly when it includes the study of several pictures.

Section 2 contains literature survey; *Section 3* contains optimized feature selection and image processing-based machine learning technique for Lung Cancer Detection. Images are enhanced using CLAHE algorithm. Image segmentation is performed by K Means algorithm. Useful features are selected using the PSO algorithm. Then classification of images is performed by SVM, ANN and KNN algorithm. *Section 4* contains result analysis and discussion. Performance of various techniques is compared on the basis of certain parameters. *Section 5* contains conclusion.

2. LITERATURE SURVEY

The earlier lung nodules are detected and removed, the lower the risk that the patient may develop to a later, more deadly stage of lung cancer. A machine learning classifier known as Random Forest, or RF, is used to differentiate between nodules and other forms of soft tissues seen in the lungs. A novel categorization method that was presented by El-Askary and colleagues [10] with the purpose of improving the early diagnosis of lung nodules. A model with five steps was developed and then verified with the use of data from 165 different LIDC cases. During the first phase, we were responsible for some basic processing as well as the gathering of pictures. After that, we will demonstrate how we extracted 119 attributes from the CT scans for your review. Organization Responsible for the Archiving and Storage of Medical Images of the Lungs (LIDC) The feature vectors may then be fine-tuned by modifying the RF parameters after the elimination of duplicate occurrences and the reduction in the sample size of the non-nodule class. Analyzing a large number of collections that were formed from the recovered feature sets allows for the selection of the classifications with the greatest possible scores. In terms of accuracy, RF fared better than KNN, DT, and SVM when it came to classifying data. In order to accomplish this objective, the suggested approach was to examine and choose those traits that produced the most effective classification conclusion. The feature sets that were based on pixels as well as those that were based on wavelets fared much better than the rest. When we were trying to fine-tune the RF, we employed 170 trees and an in-bag percentage of 0.007. The proposed model was found to have a sensitivity of 90.67 percent, a specificity of 90.81 percent, and an accuracy of 90.733 percent, according to the data.

After receiving a diagnosis of non-small cell lung cancer (NSCLC), He et al. [11] made an effort to estimate patients' chances of surviving the disease by using phenotypic radiomics attributes that were gleaned from CT scans. Pyradiomics was used to analyse the CT scans of 186 individuals diagnosed with NSCLC in order to extract characteristics from those images. In order to do this, we made use of a technique known as the synthetic minority oversampling technique (SMOTE). Additionally, we divided the final dataset into training and validation sets with a ratio of 3:1. The training of a large number of RF models was performed by using a hyper-parameter grid

Review Article | Volume 10, Issue 4 | Pages 888-894 | e-ISSN: 2347-470X

search in conjunction with ten-fold cross-validation. The accuracy or recall of the models was used as the criteria for evaluation. After that, the decision threshold for the chosen model was figured out. When evaluating the final model, predictive effectiveness and ROC were the metrics that were used. There were a total of 186 individuals whose scans were segmented, which resulted in the retrieval of 1218 different features. It was determined that a decision threshold of 0.56 was ideal, and the preferred model was chosen with recall serving as the criteria. The AUC score for Mode's ability to forecast was 0.9296, and the accuracy of its predictions was 89.33%. The increased survival prediction offered by an automated classifier equipped with hyper-parameter tuning RF has great potential for patients diagnosed with NSCLC.

Raweh et al. [12] created a hybrid method to the prediction of cancer, which combines the processes of feature selection and extraction. The F-score, a filter features selection technique, is used in the methodology that is presented in order to solve the issue of high dimensionality that is inherent in the data on DNA methylation. Naive Base, RF, and Support Vector Machine (SVM) are used to forecast breast, kidney, colon, lung, uterine, and other cancers, with and without a hybrid technique, to assess the method's accuracy. New features extraction methods for accurate cancer classification include peaks in the mean methylation density, the Fast Fourier Transform Algorithm, and the symmetry between a sample's methylation density and the mean methylation density of both sample types (normal and cancer). According to the findings, the accuracy of categorization is improved almost everywhere. Additionally, inferred proof of honesty and reliability is offered.

Authors [13] developed a novel IoT-based predictive model for the prediction of lung cancer by continuous monitoring and for the enhancement of healthcare via the supply of medical education. They did this by using fuzzy cluster-based augmentation and classification. The transition region extraction serves as the foundation for the fuzzy clustering method that is employed for successful image segmentation. This procedure begins with the extraction of the transition area. In addition, the Fuzzy C-Means Clustering Algorithm was used in order to differentiate between the features of the pictures of lung cancer and the transition zones. In the course of this study, the Otsu Thresholding technique was used in order to single out the transition zone in photographs of lung cancer. In addition, a morphological thinning method as well as the right edge image were used in order to improve the accuracy of the segmentation. In order to generate object areas, lung cancer edge pictures must first go through morphological cleaning and then have their empty spaces filled in. In this article, we provide a novel incremental Classification Algorithm (ICA) that makes use of temporal variables by combining ARM, DT, and a CNN in order to take into account time-related factors. In order to carry out the experiment, we made use of both reference photographs that we gleaned from databases as well as real-time health data that was gathered from patients by means of IoT devices. The results of the experiments indicated that the suggested model was more accurate than the models that were already in use.

A helpful algorithm for detecting and predicting lung cancer was suggested by Alam et al. [14]. The algorithm made use of support vector machine (SVM) classifiers. The procedure of determining whether or not someone had cancer consisted of a number of steps. In addition to that, it was able to calculate the risk of developing lung cancer. Image enhancements and segmentations were performed at each successive level of classification. Scaling, contrast improvement, and colour space transformation were the techniques that were used so that the photographs might be enhanced. In the segmentations, both threshold-controlled and marker-controlled watershed-based approaches were applied. We used the support vector machine classifier in order to sort the data into the appropriate categories. The suggested method was more effective in both detecting potential cases of lung cancer and making accurate diagnoses.

Using information taken from electronic medical records, Pradeep and Naveen [15] were able to provide an accurate prediction of the lung cancer survival rate. The study made use of machine learning methods to make predictions on patient survival probabilities, which made it possible for more patients to get treatment for their disease. Support vector machines (SVM), Naive Bayes (NB), and classification trees were used for the purpose of doing research on the patterns of lung cancer risk factors (CT). A previously acquired lung cancer dataset from the same institution and a freshly gathered patient dataset were used in the evaluation of the ensemble. Both datasets were obtained from the same institution. According to the findings of the experiments, C 4.5 was able to enhance its accuracy in the prediction of lung cancer as the quantity of the training dataset increased. This was indicated by the increase in the area under the receiver operating characteristic curve (ROC).

Bhuvaneswari and Therese's research looked on methods for the early diagnosis of lung cancer [16]. In this paper, an approach known as genetic KNN, which is nonparametric, is proposed for diagnosing cancer. The use of an algorithm enables medical professionals to detect nodules in CT lung pictures before they become unmanageable. The use of Genetic KNN provides a solution to this issue, which is important given the protracted nature of human CT scan interpretation. It is hoped that the methodology would appropriately classify the photos in a timely manner. An implementation that is based on the MATLAB image processing toolbox is used for the purpose of classifying the CT lung images. Just two of the many performance measures that have been looked at are the accuracy of classification and the percentage of false positives. In order to categories the data, classic KNN first calculates the distance between the test set and the training set, and then gives preference to neighbors whose distances are less. During each iteration of Genetic KNN, a fresh set of K samples is chosen with the intention of optimizing fitness. Fitness in this context refers to the level of classification accuracy that is reached. Results that are consistently trustworthy and reach a high degree of accuracy may be obtained.

Saritas and Yasar [17] used ANN with Naive Bayes classifiers in order to conduct their analysis of the sickness. There was only one destination for the data, yet there were nine different sources. A comparison of the two algorithms' performance is

International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 4 | Pages 888-894 | e-ISSN: 2347-470X

offered, with the comparison being based on instances of data classifications that were either successful or failed. The findings of the trials indicated that the anthropometric information and the parameters derived from the usual blood analysis are both extremely useful in making a correct diagnosis of breast cancer. Faisal et al. [18] examined the capacity of a large number of predictors to discriminate between patients with and without lung cancer. This was done with the goal of making the process of detecting lung cancer symptoms more effective. A variety of different classifiers, including SVM, C 4.5, Decision Tree, MLP, and NN, are evaluated with the use of UCI's benchmark dataset. Their effectiveness is evaluated in comparison to that of both proportional representation voting and majority voting. The findings of the studies make it abundantly evident that the gradient-boosted tree surpassed the approaches that are now considered to be state-of-the-art and attained 90% accuracy.

3. METHODOLOGY

In the next part, an improved method for detecting lung cancer using a machine learning algorithm that is based on image processing is shown. The CLAHE algorithm is used to improve the quality of the images. The K Means method is used to do the segmentation of the image. The PSO algorithm is used to determine which traits are beneficial. After that, an SVM, ANN, and KNN algorithm are used in order to classify the photos. *Figure 2* illustrates it for us.

In order to effectively recognize a picture, the background extraction method used must be adaptable enough to take into account the specific characteristics of the image. When constructing a histogram using CLAHE, the pixels to be included in the histogram need to be in close proximity to the pixel that is now being analyzed. By establishing a maximum on the height of the local histogram, which is often referred to as a "clip level," CLAHE is able to place a cap on the greatest amount of contrast that may be adjusted. The maximum contrast enhancement factor is thus decreased as a result. As a direct consequence of this, the image will have far less noise in the background. CLAHE is unmatched in its ability to increase the visibility of tiny features when applied to mammograms [19]. When seen against a white background, these lesions immediately stand out to the observer. Even while the method makes it easy to distinguish between information and noise, there is still a noticeable amount of grain in the image.

This research divides the ROI into a number of different sections by using a technique known as regional segmentation. Each of these sections has the ability to display a different collection of patterns and textures. The local mean is used as the clustering pattern, and each observation is then allocated to one of the k groups that are formed by the k-means algorithm. In this particular setting, the variable k stands for the total number of groups and is consulted in order to identify data clusters. The squared Euclidean distance enables an immediate evaluation of how near data points are to one another. Elements from the data set are partitioned into one of k categories that have been established in advance. We are able to organize the points in our data set into meaningful groups by using a similarity measure [20].

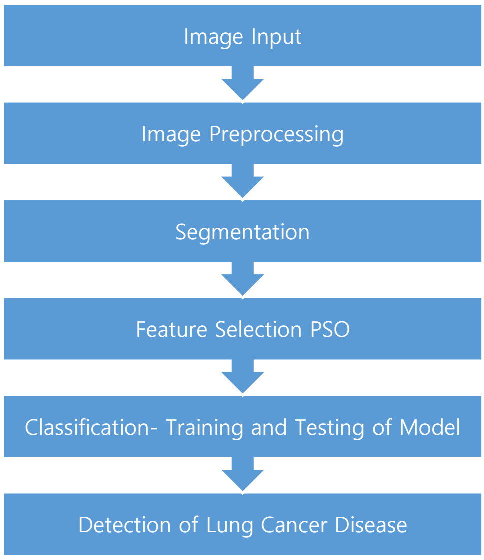

Figure 2: Optimized feature selection and image processing-based machine learning technique for Lung Cancer Detection

We utilize either a white or black background for this presentation so that we can show how effective clustering may be. This is because it is difficult to accurately split the return on investment (ROI) while filming an object against a live backdrop. It is challenging to use clustering to detect the ROI ill region in a picture while the backdrop is moving, which is what K claims to be able to achieve.

The K-nearest neighbor technique [21] produces the most consistent results when it comes to the discovery of patterns and the classification of data. Distance functions and similarity metrics are used by k-nearest neighbor algorithms in order to locate the individuals who are geographically located the most closely to one another. The similarity measure assists in the classification of newly occurring cases while simultaneously tracking the current state of all previously existing examples. Because of the use of instance-based learning, straightforward classification is now a practical possibility. When classifying a new instance of a dataset, the majority of votes from classes that are geographically close to the class into which the new instance is being categorized are taken into consideration. The distance is figured out by taking information from both the training set and the test set into consideration. The first thing that has to be done is to calculate the distance, or k, that separates one occurrence from the next, and only after that can we get things moving.

Mathematically speaking, this model depicts both the structure and the functioning of a biological neural network. Neural networks of living organisms carry out their functions in a biological manner. To put it another way, its functions are comparable to those that are performed by the human brain. The basic computing units that may be found in a human's brain and nervous system are represented as "neurons" in an artificial neural network (ANN). These are the fundamental building blocks of the signaling architecture of the nervous system. Each neuron has its own unique cell body, and each neuron has the capacity to carry out a wide range of distinct tasks. Neurons are found in the brain. Components of theoretical circuits that are capable of executing computations have been modelled after the architectural structures of biological networks. In order to have any meaningful conversation about how the brain works, one must first have a solid grasp of how individual neurons operate. An artificial neural network, often known as an ANN, is a method of data processing that makes use of a connectionist approach. Synthetic neurons serve as the fundamental building blocks of this system. Similarities may be drawn between this idea and the axons and dendrites that can be found in the nervous system. The robustness of the model, its adaptability, and its capacity for collaborative computing are all advantages, but the model's capacity to self-organize and change is what makes it really useful. Adaptability and robustness go hand in hand. There are three major components that have to be in place before an artificial model of a biological neuron can function properly. Weights, much as in actual neurons, are utilized to represent connections in this model. In the brain, synapses may either be formed spontaneously or by the process of artificial neurogenesis. Synapses of a biological kind are far more likely to be encountered than those of any other sort. A quantitative value that may be conceptualized as a "number" is called a weight, and it is given to each individual synapse that makes up an artificial neuron. Weights whose values are in the negative refer to an inhibitory relationship, while weights whose values are in the positive signal to an excitatory association. First, all of the inputs are summed together, and then they are weighted. This is what the professionals refer to as a linear combination. Simply using an activation function is all that is necessary to effect changes to the output volume. As an example, the acceptable range for output values might be anywhere from 0 to 1 (or -1 to 1) [22].

Within the realm of machine learning, classification and regression work are two of the most common applications of support vector machines, sometimes known as SVMs for short. Through the use of training examples, the Support Vector Machine (SVM) approach is able to develop a non-probabilistic binary linear classifier. This classifier can then be used to the classification of new situations. Using the biggest gap that is currently available, the data points that make up an SVM model have the potential to be partitioned into as many meaningful categories as are practically possible. LS-SVM is an upgrade of the SVM approach that may be used to solve linear equations and generate a training model for classification. LS-SVM is an acronym that stands for "linear system support for support vector machine." Equations of the linear and quadratic varieties are used in the construction of support vector machines (SVMs). It will save you both money and time if you use an LS-SVM classifier. Since the LS-SVM only requires one linear equation to describe how it works, it is possible that it is easier to understand than the traditional SVM. Before it can be used, LS-SVM needs minor adjustments to be made to just a few of its characteristics. Radial basis functions, sometimes known as RBFs, are used by the SVM's RBF kernel [23].

4. RESULT ANALYSIS AND DISCUSSION

This experiment is performed on 250 images selected from [24]. 175 images are used to train the classification model and remaining 75 images are used to test the classification model. Images are enhanced using CLAHE algorithm. Image segmentation is performed by K Means algorithm. Useful features are selected using the PSO algorithm. Then classification of images is performed by SVM, ANN and KNN algorithm. Performance and results are shown in *figure 3, figure 4* and *figure 5.*

Accuracy= (TP + TN) / (TP + TN + FP + FN)
Sensitivity = TP/ (TP + FN)
Specificity = TN/ (TN + FP)

Where
TP= True Positive
TN= True Negative
FP= False Positive
FN= False Negative

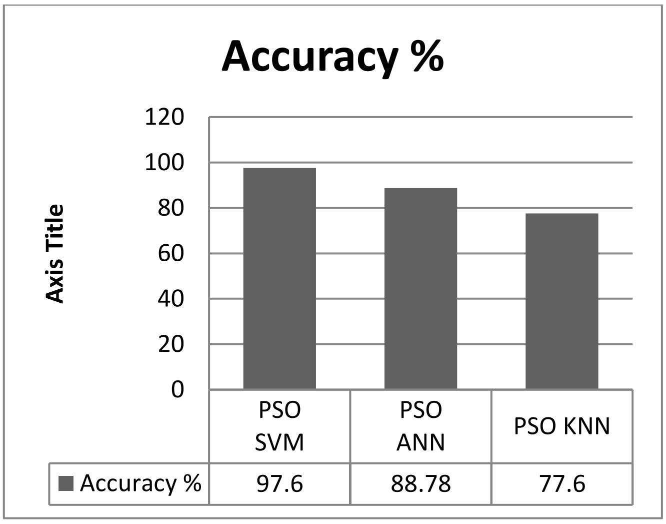

Figure 3: Accuracy of Classifiers for Lung Cancer Detection by CT-Scan examination

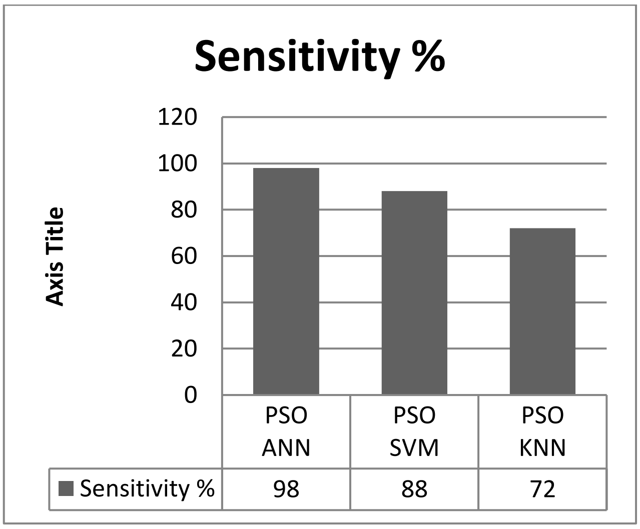

Figure 4: Sensitivity of Classifiers for Lung Cancer Detection by CT-Scan examination

Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 4 | Pages 888-894 | e-ISSN: 2347-470X

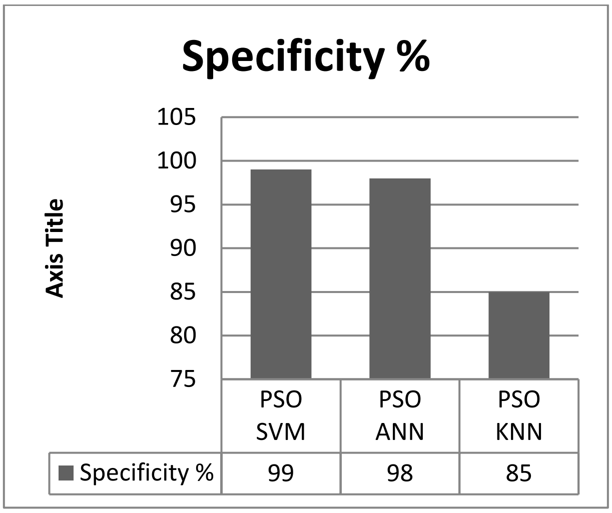

Figure 5: Specificity of Classifiers for Lung Cancer Detection by CT-Scan examination

5. CONCLUSION

The uncontrolled proliferation of lung cells is one of the most important factors that might lead to the development of lung cancer. The use of tobacco products and the smoking of cigarettes are the two primary risk factors that contribute to the development of lung cancer. There are two main subtypes that are considered to be the most frequent kinds of lung cancer. Non-small cell lung cancer and small cell lung cancer. A computed tomography, or CT, scan is an essential diagnostic tool that is used to determine the type of cancer a patient is suffering from, as well as its stage, the location of any metastases, and the degree to which the disease has spread to other organs. Other diagnostic tools include a magnetic resonance imaging (MRI) scan and an ultrasound. The subject of computer science known as machine learning, sometimes abbreviated as ML, focuses on the creation of algorithms that can teach computers new behaviors based on examples collected from the real world. Machine learning, also abbreviated as ML, is frequently referred to as "ML." This article provides an explanation of how machine learning may be used in the process of diagnosing lung cancer. To be more specific, this is achieved by making use of optimized feature selection and picture processing. The CLAHE algorithm is used in order to accomplish improving the overall picture quality. When there is a need to segment a picture, the K Means method is the one that is used. In this case, feature selection is accomplished by the use of the PSO approach. After that, the SVM, ANN, and KNN classification algorithms are used on the images in order to sort them into categories. The computed tomography (CT) scan is used to generate the images. For compared to other techniques, PSO SVM performs very well when identifying cases of lung cancer.

REFERENCES

[1] Siegel, RL, Miller, KD, Jemal, A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70:7–30. https://doi.org/10.3322/caac.21590.

[2] Sung, H, Ferlay, J, Siegel, RL, Laversanne, M, Soerjomataram, I, Jemal, A, et al.. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71:209–49. https://doi.org/10.3322/caac.21660.

[3] Raghuvanshi, A., Singh, U., & Joshi, C. (2022). A Review of Various Security and Privacy Innovations for IoT Applications in Healthcare. Advanced Healthcare Systems, 43-58. doi: 10.1002/9781119769293.ch4

[4] Zhao, W, Yang, J, Sun, Y, Li, C, Wu, W, Jin, L, et al.. 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. Cancer Res 2018; 78:6881–9.

[5] V. Durga Prasad Jasti, Abu Sarwar Zamani, K. Arumugam, Mohd Naved, Harikumar Pallathadka, F. Sammy, Abhishek Raghuvanshi, Karthikeyan Kaliyaperumal, "Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis", Security and Communication Networks, vol. 2022, Article ID 1918379, 7 pages, 2022. https://doi.org/10.1155/2022/1918379

[6] Pennathur, A, Brunelli, A, Criner, GJ, Keshavarz, H, Mazzone, P, Walsh, G, et al.. Definition and assessment of high risk in patients considered for lobectomy for stage I non-small cell lung cancer: the American Association for Thoracic Surgery expert panel consensus document. J Thorac Cardiovasc Surg 2021; 162:1605–18.e6. https://doi.org/10.1016/j.jtcvs.2021.07.030.

[7] Sushovan Chaudhury, Alla Naveen Krishna, Suneet Gupta, K. Sakthidasan Sankaran, Samiullah Khan, Kartik Sau, Abhishek Raghuvanshi, F. Sammy, "Effective Image Processing and Segmentation-Based Machine Learning Techniques for Diagnosis of Breast Cancer", Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 6841334, 6 pages, 2022. https://doi.org/10.1155/2022/6841334

[8] Wang, J, Wu, L. An evaluation of aumolertinib for the treatment of EGFR T790M mutation-positive non-small cell lung cancer. Expert Opin Pharmacother 2022; 23:647–52.

[9] Hemamalini, V., Rajarajeswari, S., Nachiyappan, S., Sambath, M., Devi, T., Singh, B., & Raghuvanshi, A. (2022). Food Quality Inspection and Grading Using Efficient Image Segmentation and Machine Learning-Based System. Journal Of Food Quality, 2022, 1-6. doi: 10.1155/2022/5262294

[10] El-Askary, NS, Salem, MAM & Roushdy, MI 2019, „Feature Extraction and Analysis for Lung Nodule Classification using Random Forest". In Proceedings of the 2019 8th International Conference on Software and Information Engineering. pp. 248-252.

[11] He, B, Zhao, W, Pi, JY, Han, D, Jiang, YM & Zhang, ZG 2018, „A biomarker basing on radiomics for the prediction of overall survival in non–small cell lung cancer patients". Respiratory research, vol. 19, no. 1, pp. 1-8.

[12] Raweh, AA, Nassef, M & Badr, A 2018, „A hybridized feature selection and extraction approach for enhancing cancer prediction based on DNA methylation". IEEE Access, vol. 6, pp. 15212-15223.

[13] Senthil Kumar, K, Venkatalakshmi, K & Karthikeyan, K 2019, „Lung cancer detection using image segmentation by means of various evolutionary Algorithms". Computational and mathematical methods in medicine.

[14] Alam, J, Alam, S & Hossan, A 2018, „Multi-stage lung cancer detection and prediction using multi-class svmclassifie". In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) IEEE. pp. 1-4.

[15] Pradeep, KR & Naveen, NC 2018, „Lung cancer survivability prediction based on performance using classification techniques of support vector machines, C4". 5 and Naive Bayes Algorithms for healthcare analytics.Procedia computer science, vol. 132, pp. 412-420.

[16] Bhuvaneswari, P & Therese, AB 2015, „Detection of cancer in lung with k-nn classification using genetic Algorithm". Procedia Materials Science, vol. 10, pp. 433-440.

[17] Saritas, MM & Yasar, A 2019, „Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification". International Journal of Intelligent Systems and Applications in Engineering, vol. 7, no. 2, pp. 88-91.

[18] Faisal, MI, Bashir, S, Khan, ZS & Khan, FH 2018, „An Evaluation of Machine Learning Classifiers and Ensembles for Early Stage Prediction of Lung Cancer". In 2018 3rd International Conference on Emerging

Trends in Engineering, Sciences and Technology (ICEEST) IEEE. pp. 1-4.

[19] Hardikkumar S. Jayswal and Dr. Jitendra P. Chaudhari (2022), Lime Diseases Detection and Classification Using Spectroscopy and Computer Vision. IJEER 10(3), 677-683. DOI: 10.37391/IJEER.100343.

[20] Mittal, H., Pandey, A.C., Saraswat, M. et al. A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. Multimed Tools Appl (2021). https://doi.org/10.1007/s11042-021-10594-9

[21] Prasanthi Rathnala, M.S. Pradeep Kumar Patnaik, Srinivasa Rao Sura, Bolla Prasad, N Siva Mallikarjuna Rao and Delione N Rayan (2022), Design of an Efficient Face Recognition system using Deep Learning Technique. IJEER 10(3), 689-693. DOI: 10.37391/IJEER.100345.

[22] N. Darapaneni et al., "Handwritten Form Recognition Using Artificial Neural Network," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 420-424, doi: 10.1109/ICIIS51140.2020.9342638.

[23] Vijaya Shetty, S., Karan, R., Devadiga, K., Ullal, S., Sharvani, G.S., Shetty, J. (2023). Kannada Handwritten Character Recognition Techniques: A Review. In: Rajakumar, G., Du, KL., Vuppalapati, C., Beligiannis, G.N. (eds) Intelligent Communication Technologies and Virtual Mobile Networks. Lecture Notes on Data Engineering and Communications Technologies, vol 131. Springer, Singapore. https://doi.org/10.1007/978-981-19-1844-5_56

[24] https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images