MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Identifying Lung Cancer Using CT Scan Images Based on Artificial Intelligence

MD. Ismail Hossain Sadhin[1], Methila Farzana Woishe*[1], Nila Sultana[1], and Tamanna Zaman Bristy[1]

[1]*Dept. of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh*

KEYWORDS

Lung cancer
CT image segmentation
Image processing
X-ray.

ARTICLE HISTORY

Received 10 February 2022
Received in revised form 7 March 2022
Accepted 15 March 2022
Available online 16 March 2022

ABSTRACT

Lung cancer is among the leading cause of death among men and women. Early detection of lung cancers can increase the possibility of survival amongst patients. The preferred 5-years survival rate for lung most cancers sufferers will increase from 16% to 50% if the disease is detected on time. Computerized tomography (CT) is frequently used for diagnosis and is more efficient than X-ray. However, the images need to be reviewed by a qualified physician who specializes in interpreting the CT scan. This may lead to misinterpretation and conflicting reports among physicians. Therefore, a lung cancer detection system that uses image processing methods to categorize lung cancer in CT images will be more consistent and precise. This paper presents a lung cancer detection system using the Artificial Intelligence (AI) method. The study uses Median, Gaussian, and Watershed segments to reduce noisy and shredded CT images. Then, the Weight Optimization Neural Network method was used to improve accuracy and reduce the computational time. The results were compared with previous works and shows higher accuracy and shorter computational time.

© 2022 The Authors. Published by Penteract Technology.
This is an open access article under the CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/).

1. INTRODUCTION

Lung cancer is one of the most sterling cancers worldwide, with the lowest survival rate after the determination, showing a non-stop increment every year. The previous detection is the better the probabilities of fruitful treatment. But the detection has a few issues moreover. Here our focus is how to extend the quality of early cancer detection. The most common problem is nodule size. It is extensive in the human lung. Generally, a nodule diameter can take any value between couples of millimeters up to several centimeters [1]. Nodule shows an enormous variety in thickness and can be deceiving on a radiograph. As nodules can seem everywhere inside the lung field. Another problem is the complexity of time. Early detection requires complexity to reduce time. Apart from this, accuracy is essential. Moreover, image pre-processing comparable noise removal image smoothing can assist with increasing the recognition of nodules. These are the critical recognition elements in this study. Lung cancer is one of the reasons for cancer demises. It is hard to stumble on as it arises, and well-known signs and symptoms are inside the terminal phase [2]. The main often analyzed cancer is lung cancer, that's 11.6%. However, early detection and treatment of the disease can reduce the chance of mortality. The high-quality imaging approach CT imaging is stable for lung cancer dedication considering that it can unveil every suspected and unpredicted lung cancer nodule [3]. Misinterpretation among experts and radiologists would possibly reason problems in detecting the cancerous cell [4]. There were many systems advanced and research occurring recognition of lung cancer. However, these systems do not have a high accuracy rate and require a long detection time.

Image processing strategies and machine getting are among the current methods used to detect and classify lung cancers. Additionally, Artificial Intelligence (AI) strategies were exploited for extensive data. This paper also reviews the modern-day advanced systems for cancers recognition using CT scan images of lungs. Then, the best methods are chosen for the detection.

1.1. Research Background

Lung cancer is one of the bases of cancer demise. It is strenuous to see as it occurs and has incurable prodromes. Cancer most frequently analyzed is lung cancer with 11.6%.

*Corresponding author:
E-mail address: Methila Farzana Woishe < woishe.methila97@gmail.com >.
2785-8901/ © 2022 The Authors. Published by Penteract Technology.
This is an open access article under the CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/).

However, fast detection and treatment of the disease can reduce the likelihood of death. Best Imaging Method Computed tomography is robust for determining lung cancer because it can reveal any suspected and unsuspected lung cancer nodule. In both cases, the escalating changes in the CT images and the misinterpretation of the anatomical structure by specialists and radiologists could lead to problems when stamping the cancer cell. Many systems evolved, and studies have been accomplished to locate lung cancers. However, a few systems do now no longer have pleasant detection accuracy, and a few systems nonetheless want to be advanced to gain the most accuracy of 100%. Image processing and machine studying strategies had been updated to become aware of and classify lung cancers. Artificial intelligence strategies had been used to resolve the prediction and selection of large data. We tested contemporary systems that have evolved especially in cancer recognition based on computed tomography images of the lungs to select modern outstanding systems. The assessment became carried out on them and the new edition changed proposed.

1.2. Research Motivation and Objective

Health maintenance is one of the elemental sources of enormous information. Meticulous analysis of health maintenance information is exceedingly in a request for early diagnosis of the illness. According to the new situation in medical science, Lung cancer is one of the unsafe and execute maladies within the world. Nevertheless, early conclusion and medicament can spare life. Most cancers begin within the lung and one of the reasons is smoking. After that square of our body from battling it. Harms in cigarette smoke can debilitate the body's resistant framework, making it harder to murder cancer cells. Recently, many research works have been designed to identify patterns in massive amounts of information with higher quality. However, there can be a demand for a completely unique class technique to increase the evaluation accuracy with time. Moreover, ML algorithms are designed to increase the prediction accuracy of massive amounts of information. However, the error rate is still not exploited to its full potential. Therefore, these paintings motivate optimized system studying algorithms to enhance the evaluation accuracy with reduced time and error. Therefore, these matters are motivated us to optimize the accuracy and take some steps to improve this field.

1.2. Problem Statement

Our main task here is to improve the quality of early cancer detection. Usually, the diameter of a nodule can be anything from little millimeters to several centimeters. The nodules have a wide variety of thicknesses and therefore stand out on x-rays. Because nodules can seem everywhere inside the lung field complexity of time. Early detection requires complexity to shorten the time. Beyond that, accuracy is also important. Again, no pre-processing such as noise removal, image smoothing, seems to help improve nodule recognition. It is our important understanding factor in this research work. It is one in all the foremost real cancers at intervals in the world, with the littlest durability rate once the determination, with endless increment within the number of passing every year. The previous of detection is, the upper the probabilities of fruitful treatment are. However, detection incorporates a few problems moreover. Here our focus is the way to extend the standard of early cancer detection. Those are the point that is frequently interrupted completely over the research:

- Nodule size
- Identify the affected Nodule
- Can't reduce the time complexity
- Noisy image

2. LITERATURE REVIEW

A few analysts have proposed and carried out lung cancer using distinct image processing techniques and machine learning. The study presented in [5] offers classification among nodules and ordinary lung anatomy structures. LDA is applied as a classifier and perfect thresholding for the division. The framework has 84.06% accuracy, 97.14% affectability, and 53.33% specificity. Even though the machine detects the maximum cancers nodule, its accuracy remains unacceptable. No, machine learning techniques had been applied to classify. Authors [6] used a contortion neural network as a classifier in his CAD framework to pick out lung cancer. The framework has 83.8% accuracy, 82.6% sensitivity, and 86.8% specificity. The advantage of this version is that it utilizes around channel inside the Region of interest (ROI) extraction stage, which decreases the rate of preparing and acknowledgment steps [7]. Although the implementation value is reduced, it has however unacceptable accuracy.

The research in [8] created a framework utilizing watershed segmentation. Pre-dealing with it makes Gabor clear out to enhance the image quality. It contrasts the precision and neural fluffy version and area developing technique. The accuracy of the proposed is 90.09%, which is notably higher than the show with the division. The advantage of this version is that it uses marker-controlled watershed segmentation, which looks after over-division issues. As an impediment, it does now no longer set up the sickness as beneficial or dangerous and exactness is excessive but currently no longer acceptable. In [9], the Nonparallel Plane Proximal Classifier (NPPC) become stated for cancer type in a Computer-Aided Diagnosis (CAD) system to assure excessive type accuracy and to minimize the computation time. Artificial intelligence-primarily based computer diagnostics (CAD) is a non-invasive, goal-oriented solution that facilitates radiologists' diagnosis of lung nodules [10]. However, Valvular coronary heart problems had been one of the toughest elegance troubles. Author in [11] used three effective and popular team-learning agents for early detection of valvular coronary heart disease in bagging, promotion, and random subspaces. However, the type time decreases the usage of the methods, the rate at which accuracy is no longer achieved. In this research, has [12], three Generalized Mixing (GM) abilities have been carried out by the usage of dynamic weights to enhance the typing accuracy of the sorting system. Although the statistic handles the single label type, more than one label's troubles are not solved.

Tumor tissue primarily based totally on neurotic assessment is considered one of the fundamental pressings for early dedication in most cancers' patients. A Basic review of ANN was presented in [13], which comes approximately in an increment inside the demonstrative strategies' adequacy and specificity. However, it comes up quickly to minimize the computational complexity. The author proposed a method that

was modified to study using the thoracic surgery dataset to verify the accuracy of their proposed method in distinguishing the multiple strategies used in current strategies that include a weight-optimized maximum likelihood boosted neural network (WONN-MLB) for the core. Based on study and selection of function and most lung cancers (LCD) [14]. Automated image analysis strategies enhance the diagnostic accuracy of the disorder and decrease human error. In [15], the authors proposed distinct computational approaches for using convolutional neural networks (CNN).

3. RESEARCH METHODOLOGY

In the proposed method, a CT image was used as an input which is noisy and over shredded. This inconvenient problem has a significant effect on identifying the cancer nodule. To reduce this problem, methods include median filter, Gaussian filter, and were used to clear the images as shown in Figure 1.

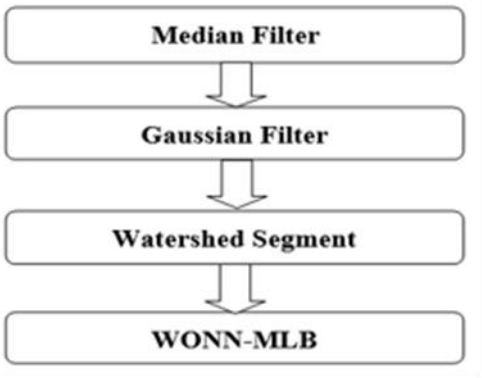

Fig. 1. Proposed Model.

Median filters are beneficial for decreasing random noise, particularly while the density of the sound amplitude possibility has huge tails and periodic patterns. All pulses within the input signal are eliminated with enough median filter passes simultaneously as all root capabilities of the input signal are preserved. The signal of the finite period is filtered after a limited variety of passes via the median filter with a set window to the basis signal, which ends up in signal convergence. This article analyzes the basis signal and its properties for a one-dimensional signal. An adaptive period median filter, a weighted median filter, a hybrid median FIR filter, and a linear mixture of a weighted median filter had been taken, and their roots had been obtained. Their properties are analyzed by figuring out the spectral strength density, the basis manner rectangular error, and the signal-to-noise ratio. The median filtering technique is executed via way of means of sliding the window over the image.

The median filter is one of the best-acknowledged filters for order information as it plays correctly for sure sorts of noise, including Gaussian, random, and salt and pepper sounds. We use this filter to dispose of ultrasonic pixels on protein crystal images earlier than the binary technique. Median filters are usually used to reduce image noise withinside the identical manner as median filters. However, it frequently works higher than a mean filter to get beneficial info withinside the picture. The significant filter often gets rid of salt and pepper noise from CT images.

Gaussian filter is a linear filter. It has been extensively studied in image processing and computer vision. By using the gossip filter to suppress the noise, the noise is accelerated while the signal is distorted. Using Gaussian filters as processing for edge identification also results in edge position displacement, fading edges, and past edges. Here the authors first examine the different techniques of these problems. They then propose an adaptive goose-filtering algorithm in which the filter varies according to both the noise characteristics and the local variation of the signal. The linear Gaussian filter could be a famous in-floor feature, is extensively utilized by researchers, and has grown to be the usual industrial cleaning method. It is usually used to dazzle the image or reduce noise. This facilitates the image and removes stained noise from the image. Only a Gaussian filter will blur the edges and reduce the contrast. This can be applied to the input surface by wrapping the surface measured by the Gaussian weight function.

Watershed segmentation is any other technique that has its origins in mathematical morphology. Vincent made a step forward in applicability and offered a set of rules in order of importance quicker and greater unique than the preceding ones. The watershed segmentation treats the image as a topographic landscape with ridges and valleys. The terrain heights are normally decided through the gray values of the corresponding pixels or their gradient amplitude. Based on this 3-D representation, the watershed transformation divides the image into watersheds. Watersheds separate the basins from every other. The watershed transform decomposes the image and assigns every pixel to the area or watershed. There are numerous small areas when there may be noisy scientific imaging data. This is known as over-segmentation. Watershed segmentation is a field-primarily based totally method that uses image morphology. It calls for selecting at least one marker interior for every for-budget of the image, inclusive of heritage as a separate budget.

To apprehend watersheds, one can consider an image a surface on which bright pixels constitute mountain peaks and valleys of darker pixels. The surface is pierced into some valleys, after which slowly submerged in a water bath. Water flows into every puncture and starts to fill the valleys. However, water from specific punctures must now no longer be mixed. That's why the demo has to be made at the primary touchpoints. These dams are the bounds of the water basin and the objects' bounds within the picture. A conventional set of rules is used for splitting, that is, to split specific things in an image. Image pre-processing makes use of a gabber filter to decorate the image and makes use of a marker-primarily based watershed technique to split and stumble on cancerous nodules. In many cases, the icons are decided on because of the neighborhood minima of the image from which the bridge is filled. This version best capabilities with the area, perimeter, and eccentricity of cancerous nodules.

After using median, Gaussian, and watershed segmentation, the CT image was prepared to read and identify the cancer nodule in the lung. However, the accuracy time that means identifying cancer in the Nodule was much delayed. That

is why the accuracy was slightly affected and occurred the time delay. To solve this problem, we use WONN-MLB. This technique is used with a weight-optimized neural network to have the most probability of lung cancer disease. Since WONN-MLB considered the beneficial functions primarily based totally on likelihood, the informative and large functions have now been removed, compromising disease diagnosis accuracy. This method reduces the time delay and helps to improve the accuracy. For 1000 patient data: Diagnosing Accuracy 92%, False Positive Rate 8.5%, Classification Time 8.3 ms,F1-score 92%.

Machine learning (ML) algorithms were discussed in Section 1.2. Here in the proposed method, two ML algorithm is used. Those are Support Vector Machine (SVM) and Random Forest. The correct result accuracy increased for using Random Forest because it got the maximum number of similar results. With Random Forest, the model is slow, and time complexity becomes higher. To reduce this problem SVM method has been used. Because the SVM method works in time and minimizes unnecessary processing time. The main goal of the proposed system is to reach close to this performance. The proposed CAD system pre-processes the 3D CT scans using watershed segmentation, normalization, down sampling, and zero-centering. The preliminary technique changed into the pre-processed 3D CT scans into 3D CNNs. However, the consequences have been poor. So, an extra pre-processing was transformed into finished to input the best areas of interest into the 3D CNNs. Then input areas around nodule applicants detected through the U-Net have been fed into 3D CNNs to classify the CT scans as tremendous or terrible for lung cancer in the long run.

4. RESULTS AND ANALYSIS

4.1. Results

Median, Gaussian, and watershed segmentation were used for identification-affected Nodule for image pre-processing. WONN-MLB was used in this work.

1000 patient records taken into consideration for the proposed WONN-MLB. The number of records effectively-recognized because the disease is '920', the diagnosing accuracy is considered as follows:

DA= (The number of data currently diagnosed as disease / Total counted data) *100
DA = (920 /1000) ∗ 100 = 92%

For false positive rate, '1000' wide variety of patient data taken into consideration as samples and '85' wide variety of patient data incorrectly diagnosed with lung cancer disease, the false positive rate is set as follows:

FPR= (Incorrectly diagnosed data / Number of total counted data)*100

FPR= (85 / 1000) * 100= 8.5%

Calculating one patient classification the WONN-MLB take 0.0083 ms and for 1000 times the calculation is-

CT =1000 * 0.0083 ms =8.3 ms

F1-score = 2 × (Precision × Recall) / (Precision+ Recall)
= 2 * (92*91) / (92+91)
= 91.49 %

F1-score is a single measure of performance test for the positive class.

Table 1. Summary of Result

Method	Test Set	Accuracy (%)	Error (%)
Median Filter	1000	45.87	54.13
Gaussian Filter	1000	62.34	37.66
Watershed Segmentation	1000	86.6	13.4
WONN-MLB	1000	92	8

Table 1 shows the results of the methods used. After using Median Filter, the model accuracy was just 45.87%, and it increased 62.34% when a Gaussian filter was used. However, the model accuracy was very low. To improve the accuracy, Watershed Segmentation is utilized. Although the accuracy was increased but it was high. For the case, WONN-MLB is worn, and the result reached 92%.

4.2. Evolution and Implementation

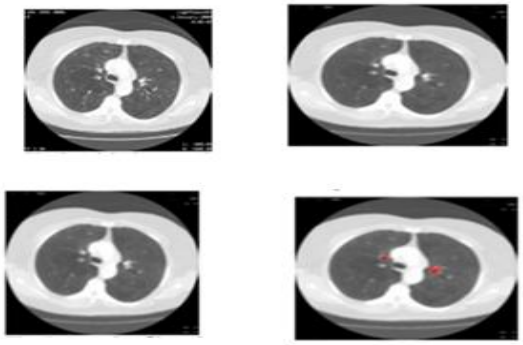

Fig. 2. Median and Gaussian Filter Application and Cancer Identification.

Those pictures are filtered by Median and Gaussian filtration, and the red mark place visualize the cancer-affected area.

4.2. Comparison with Previous Research

The previous method used only WONN-MLB to classify and predict their model. However, the proposed model uses filters and machine learning algorithms. In summary, the proposed WONN-MLB able to improve the accuracy and decrease the computational time.

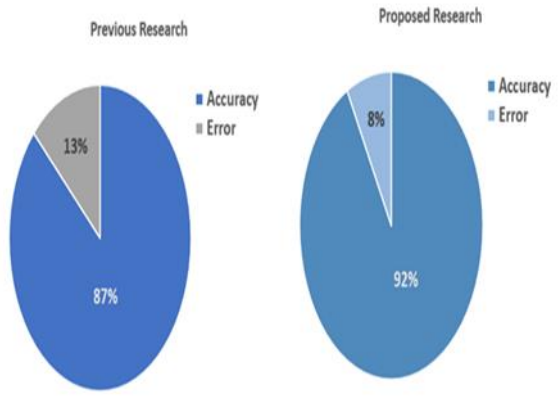

Fig.3. Comparison between the Previous and Proposed Research.

5. CONCLUSIONS AND FUTURE WORKS

The present method of lung cancer detection has lower accuracy and takes a long time to classify the degree of cancer detected in nodules. Therefore, a cutting-edge system is proposed. The presented method identifies the cancerous Nodule from the lung CT experiment image using watershed segmentation for detection. The proposed version detects cancer with 92% accuracy, and the classifier has an accuracy of 86.6%. However, this proposed method is unable to classify stages of cancer. Future works will cover the categorization of the cancer stages.

This research work is often expanded to broader sectors for generalization or sector-specific observations. The principal goal of this work is to recognize and predict the affected nodules of lung cancer. But no publicly available dataset mainly focused on affected nodules of lung cancer prediction. Lung cancer prediction forestalls future actions based on past actions. For that reason, the action and activity-based datasets were relied on to evaluate the model. The main limitations of CT screening are the high nodule detection rate. More than 50% of participants have at least one unaccounted Nodule. CT scan of the results associated with additional costs. Cost of biopsy and removal of patient or benign non-calcified nodules. The risk of cancer associated with multiple follow-up CT scans is small but difficult to quantify. Below are the key points for improvement of the current proposed method:

- Able to predict and give accurate results with high computation resources from big data.
- Using another pre-processor filter.
- The use of other methods or Layers for higher accuracy.
- More computational resources.

REFERENCES

[1] Sharma, S. (2018). A two-stage hybrid ensemble classifier-based diagnostic tool for chronic kidney disease diagnosis using optimally selected reduced feature set. International Journal of Intelligent Systems and Applications in Engineering, 6(2), 113-122.

[2] Podolsky, M. D., Barchuk, A. A., Kuznetcov, V. I., Gusarova, N. F., Gaidukov, V. S., & Tarakanov, S. A. (2016). Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels. Asian Pacific journal of cancer prevention, 17(2),835-838.

[3] Gindi, A., Attiatalla, T. A., & Sami, M. M. (2014). A comparative study for comparing two feature extraction methods and two classifiers in the classification of early-stage lung cancer diagnosis of chest x-ray images. Journal of American Science, 10(6), 13-22.

[4] Suzuki, K., Kusumoto, M., Watanabe, S. I., Tsuchiya, R., & Asamura, H. (2006). Radiologic classification of small adenocarcinoma of the lung: radiologic-pathologic correlation and its prognostic impact. The Annals of thoracic surgery, 81(2), 413-419.

[5] Aggarwal, T., Furqan, A., & Kalra, K. (2015, August). Feature extraction and LDA-based classification of lung nodules in chest CT scan images. In 2015 International Conference on Advances in Computing, Communications, and Informatics (ICACCI) (pp. 1189-1193). IEEE.

[6] Jin, X. Y., Zhang, Y. C., & Jin, Q. L. (2016, December). Pulmonary nodule detection based on CT images using convolution neural network. In 2016 9th International symposium on computational intelligence and design (ISCID) (Vol. 1, pp. 202-204). IEEE.

[7] Maurer, A. (2021). An Early Prediction of Lung Cancer using CT Scan Images. Journal of Computing and Natural Science, 39-44.

[8] Ignatious, S., & Joseph, R. (2015, April). Computer-aided lung cancer detection system. In 2015 Global Conference on Communication Technologies (GCCT) (pp. 555-558). IEEE.

[9] Ghorai, S., Mukherjee, A., Sengupta, S., & Dutta, P. K. (2010). Cancer classification from gene expression data by NPPC ensemble. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8(3), 659-671.

[10] Li, K., Liu, K., Zhong, Y., Liang, M., Qin, P., Li, H., & Liu, X. (2021). Assessing the predictive accuracy of lung cancer, metastases, and benign lesions using an artificial intelligence-driven computer aided diagnosis system. Quantitative Imaging in Medicine and Surgery, 11(8), 3629.

[11] Das, R., & Sengur, A. (2010). Evaluation of ensemble methods for diagnosing valvular heart disease. Expert Systems with Applications, 37(7), 5110-5115.

[12] Costa, V. S., Farias, A. D. S., Bedregal, B., Santiago, R. H., & Canuto, A. M. D. P. (2018). Combining multiple algorithms in classifier ensembles using generalized mixture functions. Neurocomputing, 313, 402-414.

[13] Dande, P., & Samant, P. (2018). Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: a review. Tuberculosis, 108, 1-9.

[14] Obulesu, O., Kallam, S., Dhiman, G., Patan, R., Kadiyala, R., Raparthi, Y., & Kautish, S. (2021). Adaptive diagnosis of lung cancer by deep learning Classification Using Wilcoxon gain and generator. Journal of Healthcare Engineering, 2021.

[15] Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O., & Hajirasouliha, I. (2018). Deep convolutional neural networks enable the discrimination of heterogeneous digital pathology images. EBioMedicine, 27, 317-328.

.

.