International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 03, Issue 01, Maret 2022
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

Identifying Lung Cancer Using CT Scan Images Based On Artificial Intelligence

1stMD. Ismail Hossain Sadhin, 2ndMethila Farzana Woishe, 3rd Nila Sultana, 4thTamanna Zaman Bristy
[1]D3 Automotive Technology, [2]information Systems
[1,2,3,4]Dept. of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
[1,2,3,4]408/1, Kuratoli, Khilkhet, Dhaka 1229, Bangladesh.
[1]ismailhossain.sadhin17@gmail.com, [2]woishe.methila97@gmail.com,
[3]nilasultanagodhulee@yahoo.com, [4]tamannazb@gmail.com

Abstract—Lung cancer appears to be the common reason behind the death of human beings at some stage on the planet. Early detection of lung cancers can growth the possibility of survival amongst human beings. The preferred 5-years survival rate for lung most cancers sufferers will increase from 16% to 50% if the disease is detected in time. Although computerized tomography (CT) is frequently more efficient than X-ray. However, the problem regarded to merge way to time constraints in detecting this lung cancer concerning the numerous diagnosing strategies used. Hence, a lung cancer detection system that usage of image processing is hired to categorize lung cancer in CT images. In image processing procedures, procedures like image pre-processing, segmentation, and have extraction are mentioned intimately. This paper is pointing to set off the extra precise comes approximately through making use of distinctive improve and department procedures. In this proposal paper, the proposed method is built in some filter and segmentation that pre-process the data and classify the trained data. After the classification and trained WONN-MLB method is used to reduce the time complexity of finding result. Therefore, our research goal is to get the maximum result of lung cancer detection.

Keywords— Lung cancer, CT image, Segmentation, Image processing, X-ray.

I. INTRODUCTION

Lung cancer is one of the most sterling cancers inside the world, with the smallest survival rate after the determination, with a non-stop increment with inside the variety of passing every year. The previous detection is, the better the probabilities of fruitful treatment are. But detection has a few issues moreover. Here our main focus is how to extend the quality of early cancer detection. The most common problem is nodule size. It is basically very wide in human lung. Generally, a nodule diameter can take any value between couples of millimeters up to several centimeters [1]. Nodule show an enormous variety in thickness and hence deceivability 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 likewise important. Then again, no preprocessing comparable noise removal, image smoothing can assist with increasing the recognition of nodules. These are our important recognition elements in this exploration paper.

Lung cancer is one of the reasons for cancer demises. It is hard to stumble on as it arises and well-known shows signs and symptoms 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]. Be that because it may, the alternate of escalated CT scan images and anatomical shape misinterpretation through experts and radiologists would possibly reason problems in stamping the cancerous cell [4]. There were many systems advanced and research occurring recognition of lung cancer. However, a few systems do not have the best accuracy of recognition and more or less systems nonetheless have been given to be progressed to recognize the very best accuracy inclining to 100%. Image processing strategies and machine getting to know strategies were actualized to become aware of and classify lung cancers. Besides, Artificial Intelligence strategies were exploit to clear up the estimate and selection for large data. The paper studied modern-day systems advanced main cancers recognition based mostly on CT scan images of lungs to select out the current super systems and assessment modified into executed on them and new edition modified into proposed.

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%. However, primary 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 had been 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 which have evolved especially in cancer recognition based especially on computed tomography images of the lungs to select modern outstanding systems, and 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 diagnosing the illness at early organize. 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.

1.3 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 preprocessing 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

II. LITERATURE REVIEW

A few analysts have proposed and carried out the area of lung cancer making use of distinct techniques of image processing and machine learning. The study proposed [5] proposed that offers classification among nodules and ordinary lung anatomy structure. LDA is applied as a classifier and perfect thresholding for the division. The framework has 84.06% accuracy, 97.14%fectability, and round 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% of accuracy, 82.6% of sensitivity, and 86.8% of specificity. The advantage of this version is that it utilizes around channel with 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. This research [8] created a framework utilizing watershed segmentation. In pre-dealing with it makes use of 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 makes use of 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 now no longer acceptable. In this paper [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 totally 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 considered to be one of the toughest elegance troubles. Author [11] used 3 effective and popular team-learning agents for early detection of valvular coronary heart disease in bagging, promotion, and random subspaces. However, the type time becomes decreased the usage of the methods, the rate at which accuracy become now no longer achieved. In this research has [12], three Generalized Mixing (GM) abilities had been carried out 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. A Basic assessment of ANN becomes achieved in this paper [13], which comes approximately in an increment inside the adequacy, and specificity of the demonstrative strategies, 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]. Tumor tissue primarily based totally on neurotic assessment is taken into consideration as one of the fundamental pressings for early dedication in most cancers patients. Automated image analysis strategies enhance the diagnostic accuracy of the disorder and decrease human error. In this study [15] authors proposed distinct computational strategies for using convolutional neural networks (CNN).

S

III. RESEARCH METHODOLOGY

In the proposed method, we take a CT image as an output that is why initially our CT image is noisy and over shredded. This inconvenient problem makes a huge effect to identify the cancer nodule and effective part. To reduce this problem, we organized our paper in this method such as; median filter, Gaussian filter, watershed segmentation.

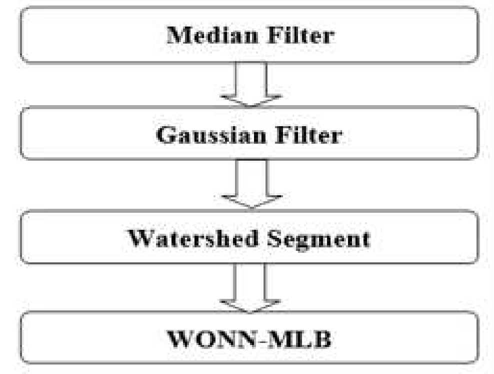

Figure 1. Proposed Model.

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 side the input signal are eliminated with enough median filter passes at the same time as all root capabilities of the input signal are preserved. The signal of the finite period is filtered after a finite variety of passes via the median filter with a set window to the basis signal, which ends up in signal convergence. In this article, the basis signal and its properties are analyzed 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 via way of means of figuring out the strength spectral 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 properly for sure sorts of noise which include 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, at the same time, 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 very famous in-floor feature, is extensively utilized by researchers, and has grown to be the usual of industrial cleaning. 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 nearby technique, which has its origins in mathematical morphology. The trendy idea became introduced. Vincent made a step forward in applicability and offered a set of rules in an 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 heights of the terrain 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 absolutely decomposes the image and hence assigns every pixel to the area or watershed. When there may be noisy scientific imaging data, there are numerous small areas. This is known as over-segmentation. Watershed segmentation is a field-primarily based totally method that makes use of image morphology. It calls for the selection of 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 as 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 offevolved 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 bounds of the objects withinside the picture. A conventional set of rules is used for splitting, that is, to split specific objects 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 together with the area, perimeter, and eccentricity of cancerous nodules.

After using median filter, Gaussian filter, and watershed segmentation the CT image was prepared for reading and identifying the cancer nodule in the lung. However, the accuracy time that means the identification of cancer in the nodule was much delayed. That is why the accuracy was a little bit affected and occur 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 boosting for lung cancer disease. Since WONN-MLB taken into consideration 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%.

In section 1.2, discussed about machine learning (ML) algorithm. Here in proposed method two ML algorithm is used. Those are Support Vector Machine (SVM) and Random Forest. For using Random Forest, the correct result accuracy increased because it getting the maximum number of similar result and represent as a final result. With Random Forest, the model being slow and time complexity become higher. To reduce this problem SVM method has been used. Because SVM method work in time and reduce unnecessary processing time.

The main goal of the proposed system is to reach close to this performance. The proposed CAD system starts with preprocessing the 3D CT scans using watershed segmentation, normalization, down sampling, and zero-centering. The preliminary technique changed into actually inputting the preprocessed 3D CT scans into 3D CNNs, however, the consequences have been poor. So an extra preprocessing changed into finished to input the best areas of interest into the 3D CNNs. Then input areas round nodule applicants detected through the U-Net have been fed into 3D CNNs to in the long run classify the CT scans as tremendous or terrible for lung cancer.

IV. RESULT AND ANALYSIS

4.1 Results

To image preprocessing, we used median filter, Gaussian filter, and watershed segmentation for identification-affected nodule. A method named WONN-MLB was also used in our research.

Proposed WONN-MLB: With ‘1000’ patient records taken into consideration for experimentation and 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,

With ‘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 as set as follows:

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

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

The Classification Time,

For 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

In table 1, all the usable method is displayed. After using Median Filter, the model accuracy was just 45.87% and it increased 62.34% when Gaussian filter is used. However, the model accuracy was very low. For the purpose to increase the accuracy, Watershed Segmentation is utilized. Although the accuracy being increased but it was not top of the mark. For the case, WONN-MLB is worn and the result reached on 92%.

4.1.1 Evolution of Implementation

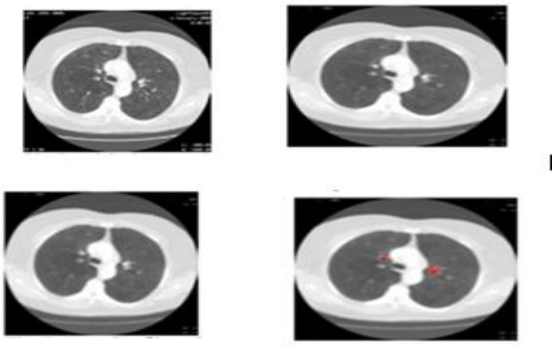

Figure 2. After Median and Gaussian Filter Application and Cancer Identification.

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 research has proposed a new method named WONN-MLB to decrease the time complexity and expand the accuracy.

- The Accuracy result of the proposed approach is higher than preceding research.

- The Error Rate is also very low compared to previous research.

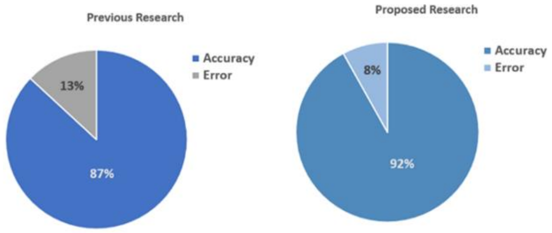

Figure 3. Comparison between the Previous and Proposed Research.

In this previous method, they used only WONN-MLB to classify and predict their model. However, in proposed model filters and machine learning algorithms are utilized to find out the higher accuracy in the result.

V. CONCLUSION AND FUTURE WORK

The present high-quality model has no exceptional results in accuracy and might no longer classify the degree of cancer detected in nodules. Therefore, a cutting-edge system is proposed. The proposed system is employed to come across the cancerous nodule from the lung CT experiment image using watershed segmentation for detection. The proposed version detects cancer with 92% accuracy that is above the contemporary version and the classifier has an accuracy of 86.6%. Overall, we can see development inside the proposed system compared to the contemporary high-quality version However, this proposed would not classify into different stages as stage I, II, III, IV of cancer. Therefore, future scope development for the duration of that is frequently finished through implementing classification in numerous levels. Also, similarly, accuracy is frequently elevated through right pre-processing and removal of false objects.

Our research work is often expanded to a broader range of sectors for generalization or sector-specific observations. The principal goal of our research was 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, we relied on the action and activity-based datasets to evaluate our model with a limited data sequence. 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. Those are the points, which will be working in the future:

- Can predict and give accurate results with high computation resources and big data.

- Try another pre-processor filter.
- Use other methods or Layers for higher accuracy.
- Use more frames per second and improve with 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] [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.