EFFICIENT CONVOLUTIONAL NEURAL NETWORK FOR THE EARLY PREDICTION OF LUNG CANCER

Dr. Manju C C[1,*], P. Veena[2]

[1]*Lecturer in Computer Engineering, Govt Polytechnic College, Punalur. manjucc@gmail.com*
[2]*Lecturer in Computer Engineering, NSS Polytechnic College, Pandalam, Kerala. veenapadma@gmail.com*

ABSTRACT

One of the most common illnesses in the world's population and the main cause of the increasing mortality rate is lung cancer. Medical experts believe computed tomography (CT) screening for lung cancer early detection can reduce mortality in patients with lung nodules. Lung cancer killed 40,000 people annually in the United States in 1965; by 2000, that number had risen to almost 200,000. Overall, the 5-year survival rate is still around 15%. It is only recently that the incidence of lung cancer has started to reduce, primarily due to a decrease in smoking. Tobacco use is thought to be a risk factor for a number of malignancies in addition to lung cancer, such as leukemia and cancers of the mouth, throat, bladder, ureter, kidney, esophagus, stomach, pancreas, and throat. This study presents an effective CNN for the early detection of lung cancer, together with Mayfly optimization and CNN classification techniques for lung cancer prediction. Lung tissue must be taken beforehand to decrease the detection area, save computing time, and increase accuracy. A better maximal inter-class variance (OTSU) technique that incorporates morphological processes is put forward. CNN are a type of artificial neural network that are mostly used for image processing and identification since they are able to recognize patterns in pictures. The architecture of the CNN model is built utilizing the parameters derived from the Grey Level Co-occurrence Matrix (GLCM) technique for feature extraction. CNN has an accuracy of 90.7% and a specificity comparison of 91%.

Keywords- *CT, OTSU, CNN, Lung Cancer and GLCM*

I. INTRODUCTION

Lung cancer, often referred to as lung carcinoma, is a kind of malignant tumor that is typified by unchecked cell proliferation in the lung tissues. Treating this is essential to prevent its growth from metastasizing to other bodily parts. Carcinomas account for the majority of lung malignancies that begin [1]. The two primary types of lung cancer are small-cell and non-small-cell lung tumors. For 85% of instances, lung cancer may be attributed mostly to long-term tobacco smoking. Ten to fifteen percent of cases include people who have never smoked but are affected by air pollution, asbestos, radon gas, and second-hand smoke. Radiographs and computer tomography (CT) scans are the standard techniques used to identify lung cancer [2, 3].

A biopsy, which is often carried out by bronchoscopy or CT scan, confirms the diagnosis. Therefore, developing a novel, reliable technique to identify lung cancer early on is crucial. It has been demonstrated that the iterative median filter, OTSU segmentation, GLCM feature extraction, Mayfly optimization, and CNN-based algorithm together provide results that are more accurate than any other combination [4, 5].

The body's tissues and organs are composed of little units known as cells. Although the appearance and functions of cells in different bodily areas may vary, most cells replicate in the same way. New cells are created to replace the aging and dying cells that are present in the body [6]. Cell development and division are normally orderly processes, but if they spiral out of control for whatever reason, the cells will keep dividing and eventually form a bulge known as a tumor [7]. There are two types of tumors: benign and malignant. A tumor that is malignant is called cancer.

A malignant (cancerous) tumor is made up of cancer cells with the capacity to spread outside of its initial location. They have the potential to infect and kill nearby tissues if left untreated. Sometimes cancer cells separate from their initial, main source and move through the circulation or lymphatic system to other organs. The term "secondary" or "metastasis" refers to the possibility that these cells may continue to divide and create a new tumor when they reach a new location inside the body. Understanding that there is more than one kind of cancer and that there is no one cure for it is crucial. More than 200 distinct types of cancer exist, each with a unique name and course of therapy [8, 9]. When diagnosing lung illnesses, computed tomography (CT) scans can yield important information. Finding any malignant lung nodules from the input lung picture and classifying lung cancer according to severity are the primary goals of this study [10].

One of the main causes of mortality in the globe is cancer. Lung cancer is one kind of cancer that is regarded as one of the main causes of death globally. Because of the nature of cancer cells, early identification of lung cancer is challenging. Lung cancer is among the most common cancers [11]. The chance of survival can be greatly increased by prompt detection and identification. Lung cancer, which is the most frequent and deadly illness in both men and women, is characterized by unregulated cell development in the lung tissues. Lung cancer is believed to be the most common cause of cancer-related fatalities worldwide since symptoms don't appear until the disease progresses. Lung cancer thus has the greatest death rate among all cancer forms [12].

Lung cancer claims the lives of more people than all other malignancies combined, including breast, colon, and prostate cancer. Considerable data proposes that detecting lung cancer at an early stage will lower the death rate Classification tests are a part of the diagnostic process from a statistical perspective [13]. The prognosis, which determines the expected course of an illness, can only be determined following a condition's diagnosis. A person with cancer may have a dismal prognosis that indicates they won't live for more than a few short months or weeks. A precise forecast cannot be made in the absence of complete understanding about a certain illness [14].

The World Health Organization's most current statistics indicates that this type of cancer causes almost 7.6 million releases worldwide each year. Moreover, it is expected that the global cancer death toll would increase further, reaching around 17 million by 2030. The phrases "diagnosis" and "prognosis" are crucial when discussing disorders. Health care the term "diagnosis" describes both the process of trying to discover or determine a potential illness or ailment as well as the conclusion that comes from this procedure [15].

The Contribution of the paper is Iterative Mean filter preprocessing and OTSU segmentation are used to forecast lung cancer. The GLCM approach is used for identification, and it is dependable in extracting both texture and color information. Lung cancer is categorized using the CNN Classification.

II. PROPOSED SYSTEM

The OTSU segmentation approach with selectively GLCM feature collecting is used in an innovative way to identify and classify Mayfly optimized lung cancer. Figure 1 indicates the intended method's precise structure.

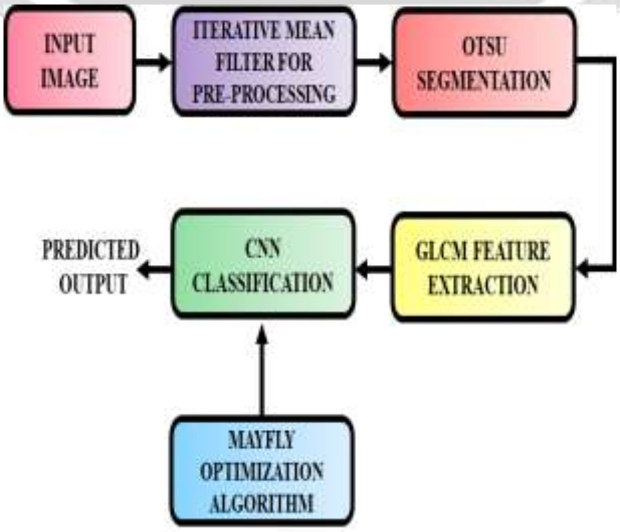


Figure 1. Block diagram for proposed system

By calculating the mean of the gray values of noise-free pixels inside a fixed-size frame, an iterative mean filter (IMF) may eliminate salt-and-pepper noise. IMF does not expand the window, which might result in a decline in the accuracy of noise reduction. One well-liked method for picture thresholding is Otsu's segmentation algorithm. Pixels are automatically clustered into two groups: foreground and background. The technique functions by examining the image's histogram, which depicts the distribution of pixel intensities. Digital image processing uses the GLCM texture analysis technique. CNN classification is a machine learning system that classifies pictures into predetermined classes using a CNN.

III. PROPOSED SYSTEM DESCRIPTION

A) ITERATIVE MEAN FILTER

The average gray value of noise-free pixels inside a fixed-size frame is utilized by IMF. IMF doesn't increase the window size as other nonlinear filters do. The accuracy of noise reduction decreases with size. IMF thus only employs a 3 × 3 window. With this capability, IMF can assess a new gray value for the central pixel more accurately.

Algorithm 1

```
        **Input:** A Noisy identical $B := [b_{ij}]_{m*n}$
        **Output:** A reinstated identical $A := [a_{ij}]_{m*n}$
Prepare $r := 1, k := 0, A^{[0]} := B, \varepsilon.$
Compute: $\delta_{max} := \underset{l \le i \le n}{\overset{max}{l \le i \le m\{b_{ij}\}}}, \delta_{min} := \underset{l \le i \le n}{\overset{min}{l \le i \le m\{b_{ij}\}}}$
        **Repeat**
                **If** $(a_{ij}^{[k]} \ge \delta_{max} || a_{ij}^{[k]} \le \delta_{min})$
                    Define: $W_{ij}^{*}(A^{[k]}, r).$
                    Update: $a_{ij}^{[k+1]} := \bar{W}_{ij}^{mean}(A^{[k]}, r)$
        **Else**
        Assign: $a_{ij}^{[k+1]} := a_{ij}^{[k]}$
        **End**
  **End**
**Until** $|A^{[k+1]} - A^{[k]}|_{\le \varepsilon}$
```

We provide an iterative method for IMF that successfully handles high-density noise. The following measures are operationalized in the research to evaluate the quality of the images: picture enhancement factor, visual information fidelity, and multistage structure similarity. We also compare the denoising results obtained using IMF with other cutting edge techniques.

B) OTSU SEGMENTATION

The biggest inter-class variance technique (OTSU), often referred to as the Otsu method or the Great Law, was introduced in the field of computer vision and image processing in 1979 by the Japanese researcher Otsu. It seeks to determine the global threshold's ideal value. It is a segmentation technique with an adjustable threshold that turns a grayscale image into a binary image. The algorithm first assumes that the image is split into two categories (foreground/target pixels and background pixels, respectively, based on the bimodal histogram distribution). Next, it calculates an ideal threshold to split the image into two categories so that between them the biggest variance is present. Otsu is Fisher's discrete judgment analysis shown in one dimension. It is generated using the least squares approach or decision analysis.

Its fundamental concept is to split the image's data into two groups using a threshold. The image's grayscale in one category's pixels is smaller than this threshold, whereas the image's grayscale in the other category's pixels is larger or closer to the threshold. The optimal threshold value is reached if the variance of the gray level pixels in these two classes is greater. Variance is a metric used to quantify how uniform the gray distribution is. The greater the difference between the foreground and backdrop in each class.

It demonstrates that the divergence between the two sections of the picture will decrease with increasing differences between the two parts, as determined by the partial variance circumscribed by the surroundings and foreground. Consequently, the classification with the lowest chance of misunderstanding is the one that optimizes the variation across classes. After that, the threshold may be adjusted to separate the background from the foreground. To be clear, the grayscale histogram is considered a distribution of likelihood and is normalized.

$$pi = \frac{ni}{n} + pi > 0 \sum_{i=1}^{L} P_i = 1 \quad (1)$$

A threshold at level k divides the pixels into classes C_0 and $C1$, which represent objects and background, or the inverse; C_0 stands for pixels with levels $[1,,k]$, whereas $C1$ stands for pixels with levels $[k + 1,..,L]$.

$$w_o = \Pr(Co) = \sum_{i=1}^{K} P_i = w(K) \quad (2)$$

$$w_1 = \Pr(C1) = \sum_{i=1}^{L} P_i = w(K) \quad (3)$$

and

$$\mu_0 = \sum_{i=1}^{K} iPr(i|C_0) = \frac{\mu(k)}{w(k)} \quad (4)$$

$$\mu_0 = \sum_{i=k+1}^{L} iPr(i|C_0) = \mu_r - \frac{\mu(k)}{1} - w(k) \quad (5)$$

where

$$w(k) \sum_{i=1}^{k} P_i \quad (6)$$

and

$$\mu(k) = \sum_{i=1}^{k} ip_i \quad (7)$$

Which, up to the kth level of the histogram, are the zeroth and first-order cumulative moments. Otsu thresholding is utilized in this paper because it is crucial to select an appropriate threshold for removing objects—in this example, the lung—from their backdrop.

C) GLCM FEATURE EXTRACTION

In digital image processing, the GLCM is a texture analysis technique. It depicts the connection between two adjacent grayscale pixels with respect to their distance, angle, and intensity. The second order statistical texture characteristics may be extracted using GLCM. It is frequently applied to picture categorization issues. Statistical texture characteristics like contrast, correlation, energy, entropy, and homogeneity may be extracted using GLCM.

Formulations to compute surface topographies from GLCM

Contrast $$\sum_{i,j=0}^{N-1} P_{\ i,j} \, (i-j)^2$$

Dissimilarity $$\sum_{i,j=0}^{N-1} P_{\ i,j} \, (i-j)$$

Homogeneity $$\sum_{i,j=0}^{N-1} \frac{P_{\ i,j}}{1+(i-j)^2}$$

Energy $$\sum_{i,j=0}^{N-1} P_{i,j}^{\,2}$$

Correlation $$\sum_{i,j=0}^{N-1} P_{i,j}\,[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)}(\sigma_j^2)}$$

ASM $$\sum_{i=1}^{G}\sum_{j=1}^{G}\{p(i,j)\}^2$$

D) CNN CLASSIFICATION

CNNs are used in the proposed process to identify and categorize patients' lung cancer CT images that are gathered from hospitals. It all comes down to combining deep learning with computer vision. A form of deep learning paradigm called convolutional neural networks is utilized to handle data with a grid layout like that of photos. See a neural network architecture in action and how it handles visual tasks like photos and movies to get an idea of this technique. Additionally, convolutional neural networks are an essential component of technology for object, face, and self-driving auto identification. A CNN typically consists of three distinct types of operation layers: the convolutional layer (CONV), the pooling layer (POOL), and the classifier layer (FC), as seen in the image below.

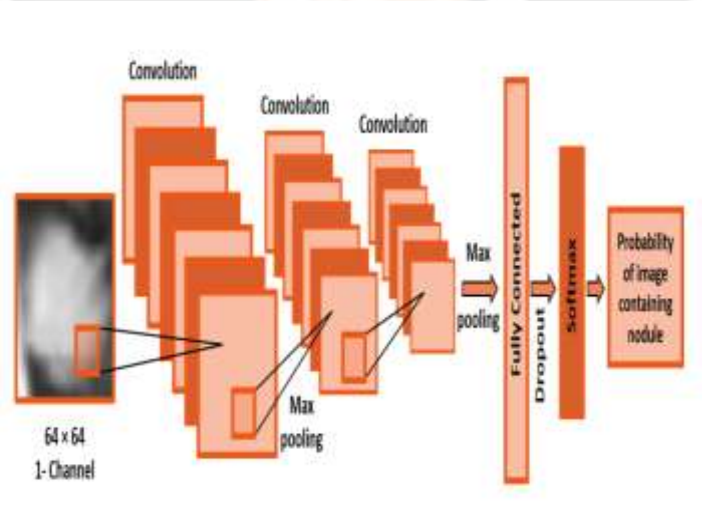


Fig.2. CNN Classification

Applications for convolutional neural networks (CNNs), also known as ConvNets, include object identification, face recognition, image processing, and more. The primary use of this kind of neural network is the detection of malignant or non-malignant lung cancers. Within the field of pattern recognition and computer vision research, convolutional neural network (CNN) models have gained popularity due to their promising results in producing high-quality picture representations.

E) MAYFLY OPTIMIZATION ALGORITHM

The mating ritual and flying habits of mayflies serve as the model for the population-based mayfly optimization algorithm (MA). Optimization issues, both single- and multi-objective, are solved using it. Inspired by mayfly mating behavior and flying, the mayfly optimization algorithm (MA) is a population-based method. Both single- and multi-objective optimization issues are solved using it.

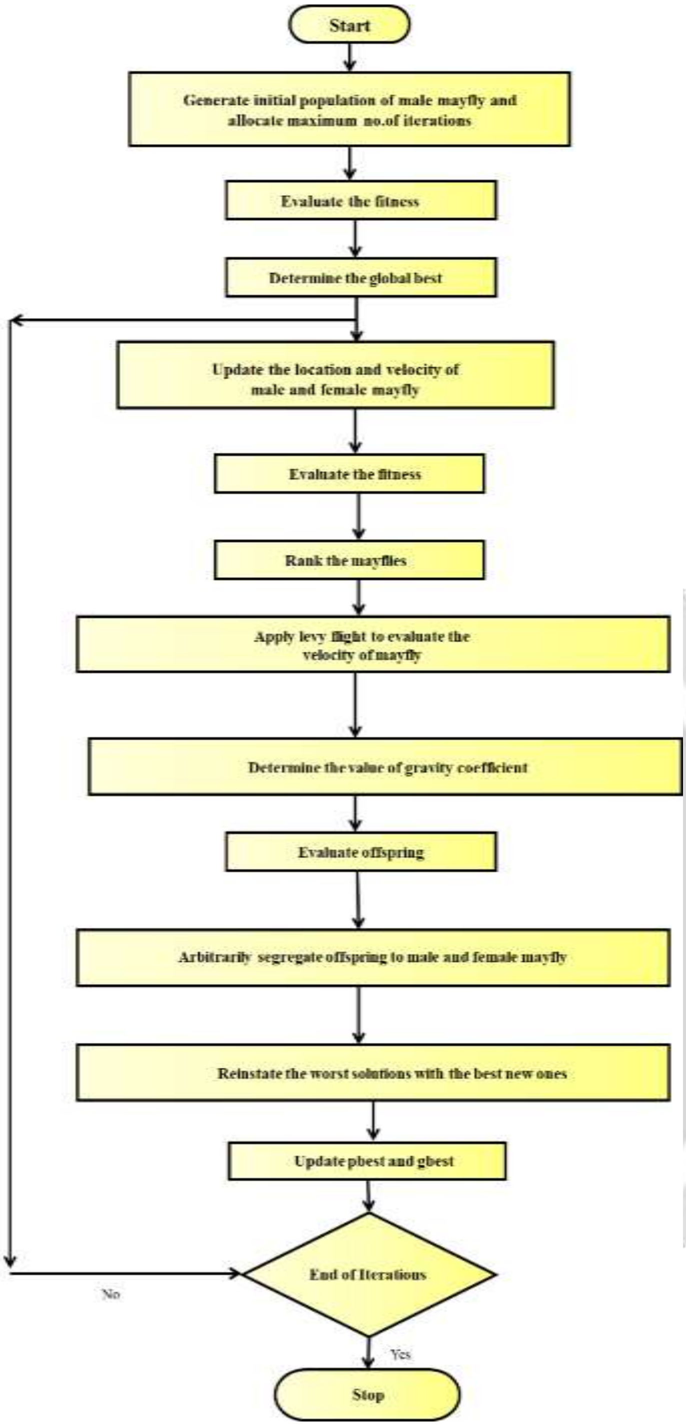

Start
Generate initial population of male mayfly and allocate maximum no.of iterations
Evaluate the fitness
Determine the global best
Update the location and velocity of male and female mayfly
Evaluate the fitness
Rank the mayflies
Apply levy flight to evaluate the velocity of mayfly
Determine the value of gravity coefficient
Evaluate offspring
Arbitrarily segregate offspring to male and female mayfly
Reinstate the worst solutions with the best new ones
Update pbest and gbest
End of Iterations
No
Yes
Stop

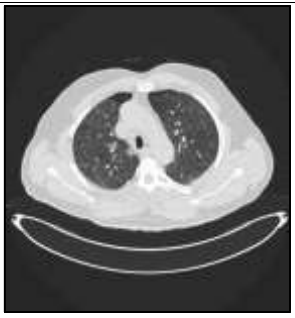
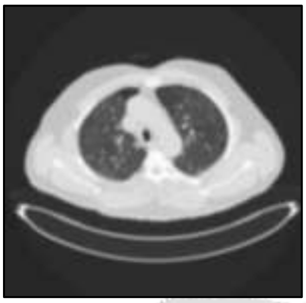
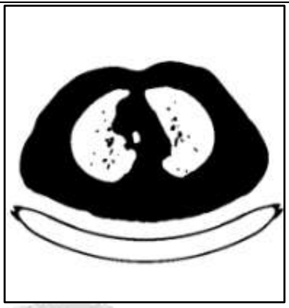
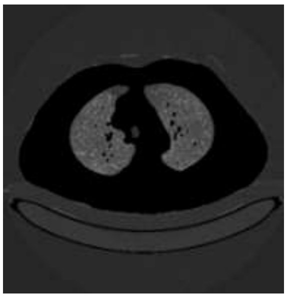
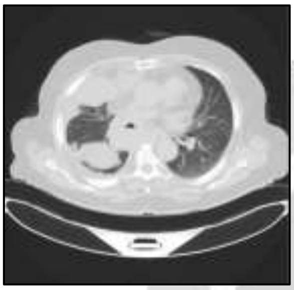
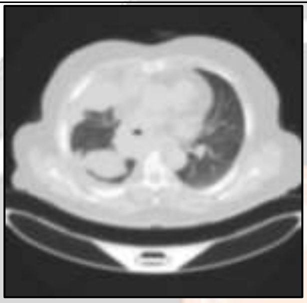
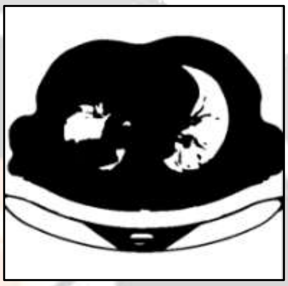
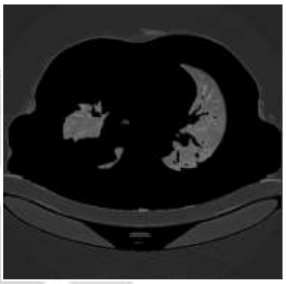
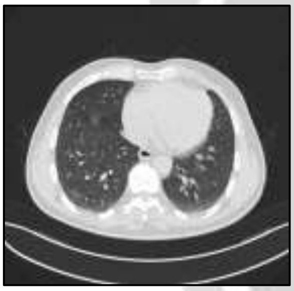
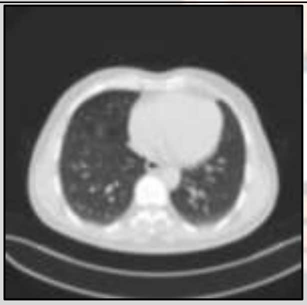
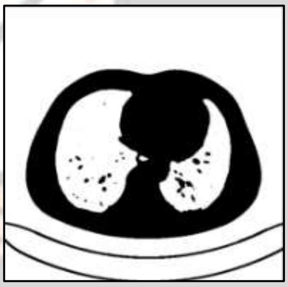
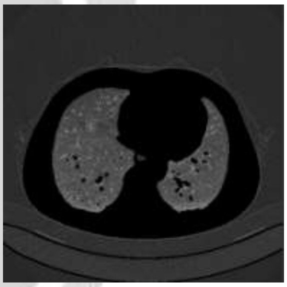
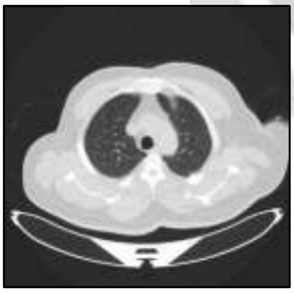
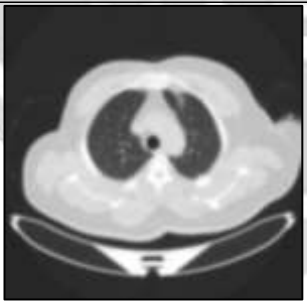
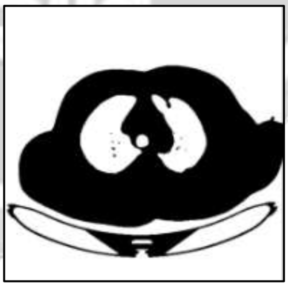
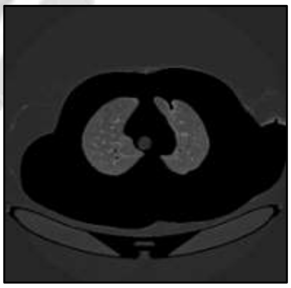
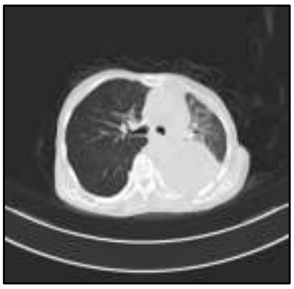
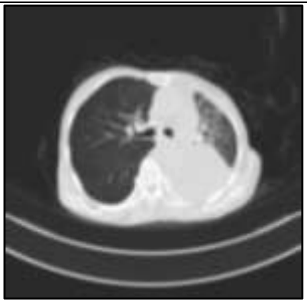
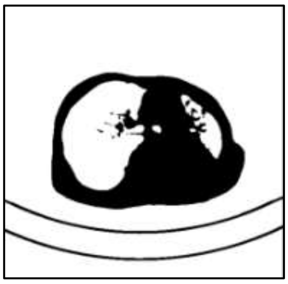
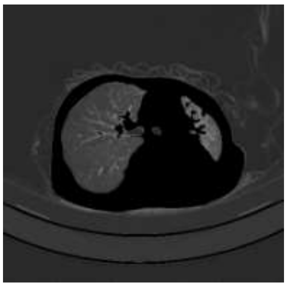
Fig.3. Flowchart of Mayfly Optimization Algorithm

The Mayfly Algorithm (MA), a population-based bio inspired algorithm that was introduced recently, has proven effective in solving several engineering issues. Its excessive number of parameters, however, makes it challenging to select and modify a set of suitable settings for various issues. These approaches have the following advantages: they can be used with ease to both discrete and continuous functions; they don't require the use of any additional complicated mathematical procedures, such derivatives; and they don't frequently become stuck at local optimal locations.

IV RESULTS AND DISCUSSION

This research compares and studies the effectiveness of Iterative Filter Preprocessing, OTSU Segmentation, GLCM Feature Extraction, and CNN Classification in improving the diagnosis of lung cancer

Table. 1. Image Processing in Lung Cancer Identification

ORIGINAL IMAGE	ITERATIVE MEAN	OTSU SEGMENTATION	
		THRESHOLD	MASKED REGIONS
			
			
			
			
			

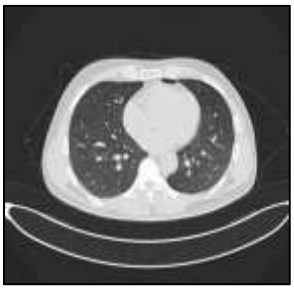
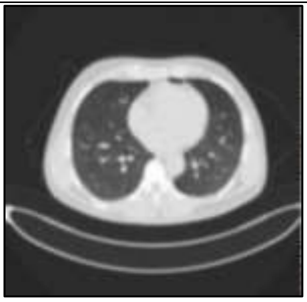
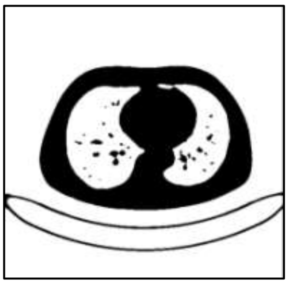
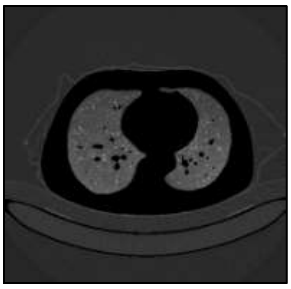
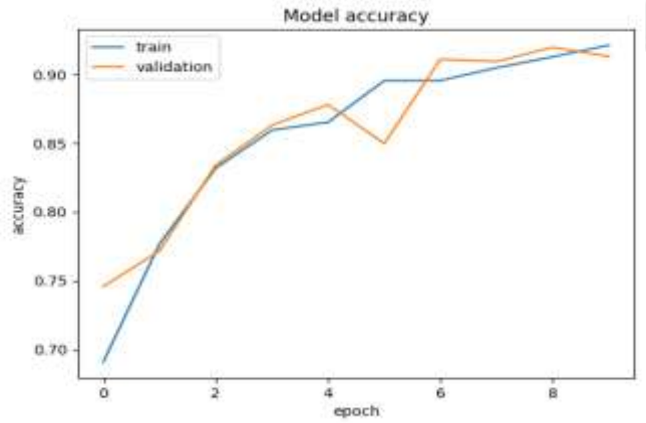
			

Table.2. GLCM Feature

	Contrast	Dissimilarity	Homogeneity	Energy	Correlation	ASM
Image 1	88.04525058	4.8097748	0.34498464	0.1531637	0.99221285	0.02345912
Image 2	44.5034667	3.0409086	0.42598365	0.09390329	0.99651503	0.00881783
Image 3	88.10267857	5.2017108	0.26284842	0.04893511	0.99221079	0.00239465
Image 4	81.64814702	5.05662916	0.312751	0.13936265	0.99357405	0.01942195
Image 5	47.65698768	2.66652015	0.50510058	0.12479949	0.99481601	0.01557491
Image 6	90.69703247	4.98893866	0.2990587	0.08347586	0.99073881	0.00696822

Figure 4 Describes the Model Accuracy and Model loss for Lung Cancer Identification.



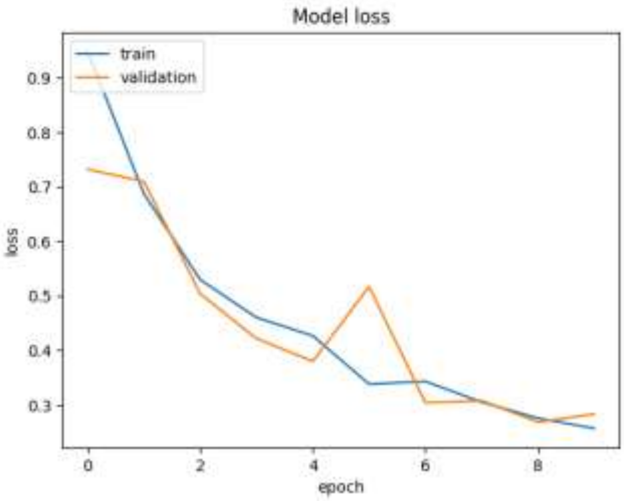


Fig.4. Model Accuracy and Model loss

Figure 5 presents the Accuracy Comparison of the Classification values. The results for the SVM, ANN, and CNN Classifier are, respectively, 86.2%, 89%, and 90.7%.

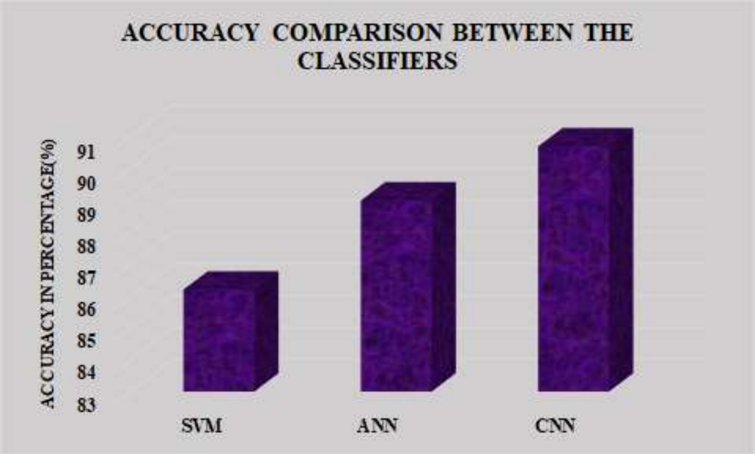


Fig.5. Accuracy Comparison between the Classifier

Fig. 6 presents the Specificity Comparison of the Classification values, which are 86.7, 88.2%, and 91% for the SVM, ANN, and CNN Classifiers, respectively.

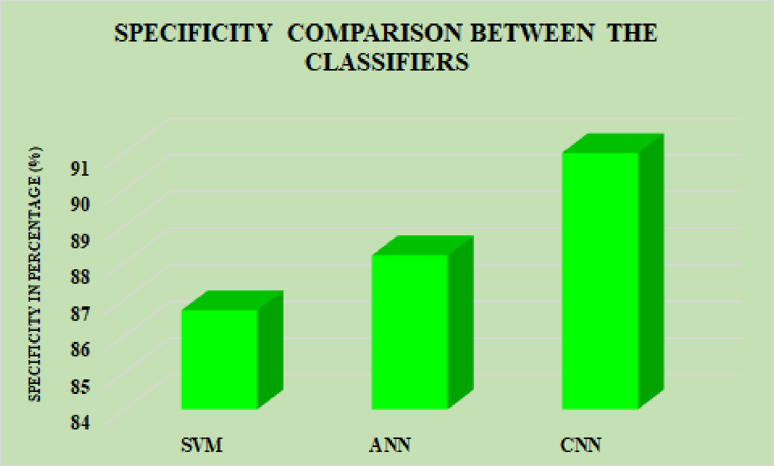


Fig.6. Specificity Comparison between the classifiers

V CONCLUSION

Due to late diagnosis, when the illness has a lower chance of being cured, lung cancer has one of the lowest five-year survival rates. There has been a 22% improvement in the national average of persons living five years after being diagnosed with lung cancer, to 26.6%. Unchecked cell development is cancer. Gene mutations have the potential to accelerate the pace of cell division or prevent the body from regulating regular processes like cell cycle arrest or programmed cell death, which can lead to cancer. A clump of malignant cells has the potential to grow into a tumor. The findings of the efficacy evaluation of the proposed technique show that neural networks may be used to help doctors in the latter stages of lung tumors. Doctors might be able to prepare better drugs and diagnose illnesses early with the help of the prediction. The Accuracy of the CNN is 90.7% and Specificity comparison of CNN is 91%.

REFERENCE

1. Krishna Chaitanya, Thandra, Adam Barsouk, Kalyan Saginala, John Sukumar Aluru, and Alexander Barsouk. "Epidemiology of lung cancer." Contemporary Oncology/Współczesna Onkologia, Vol. 25, no. 1, pp. 45-52, 2021.
2. Reem, Nooreldeen, and Horacio Bach. "Current and future development in lung cancer diagnosis." International journal of molecular sciences, Vol. 22, no. 16, pp. 8661, 2021.
3. Nadia, Gonçalo Forjaz, Howlader Meghan J. Mooradian, Rafael Meza, Chung Yin Kong, Kathleen A. Cronin, Angela B. Mariotto, Douglas R. Lowy, and Eric J. Feuer. "The effect of advances in lung-cancer treatment on population mortality." New England Journal of Medicine, Vol. 383, no. 7, pp. 640-649, 2020.
4. Matthijs, ShiYuan Liu, Oudkerk, Marjolein A. Heuvelmans, Joan E. Walter, and John K. Field. "Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives." Nature reviews Clinical oncology, Vol. 18, no. 3, pp. 135-151, 2021.
5. Amanda, Leiter, Rajwanth R. Veluswamy, and Juan P. Wisnivesky. "The global burden of lung cancer: current status and future trends." Nature Reviews Clinical Oncology, Vol. 20, no. 9, pp. 624-639, 2023.
6. Min, Yuan, Li-Li Huang, Jian-Hua Chen, Jie Wu, and Qing Xu. "The emerging treatment landscape of targeted therapy in non-small-cell lung cancer." Signal transduction and targeted therapy, Vol. 4, no. 1, pp. 61, 2019.
7. Jacob J., Chabon, Emily G. Hamilton, David M. Kurtz, Mohammad S. Esfahani, Everett J. Moding, Henning Stehr, Joseph Schroers-Martin et al. "Integrating genomic features for non-invasive early lung cancer detection." Nature, Vol. 580, no. 7802, pp. 245-251, 2020.
8. Mariano, Provencio,Ernest Nadal, José L. González-Larriba, Alex Martínez-Martí, Reyes Bernabé, Joaquim Bosch-Barrera, Joaquín Casal-Rubio et al. "Perioperative nivolumab and chemotherapy in stage III non–small-cell lung cancer." New England Journal of Medicine, Vol. 389, no. 6, pp. 504-513, 2023.
9. Aaron C., Tan, and Daniel SW Tan. "Targeted therapies for lung cancer patients with oncogenic driver molecular alterations." Journal of Clinical Oncology, Vol. 40, no. 6, pp. 611-625, 2022.
10. Aritraa, Lahiri, Avik Maji, Pravin D. Potdar, Navneet Singh, Purvish Parikh, Bharti Bisht, Anubhab Mukherjee, and Manash K. Paul. "Lung cancer immunotherapy: progress, pitfalls, and promises." Molecular Cancer, Vol. 22, no. 1, pp. 1-37, 2023.
11. M., P. Muthu Kannan, Lavanya, and M. Arivalagan. "Lung cancer diagnosis and staging using firefly algorithm fuzzy C-means segmentation and support vector machine classification of lung nodules." International Journal of Biomedical Engineering and Technology, Vol. 37, no. 2, pp. 185-200, 2021.
12. Rupak, Bhakta, and ABM Aowlad Hossain. "Lung tumor segmentation and staging from ct images using fast and robust fuzzy C-Means clustering." International Journal of Image, Graphics and Signal Processing, Vol. 12, no. 1. Pp. 38, 2020.
13. Leilei, Zhao, Junhui Qian, Fengchun Tian, Ran Liu, Bei Liu, Shuya Zhang, and Mengchen Lu. "A weighted discriminative extreme learning machine design for lung cancer detection by an electronic nose system." IEEE Transactions on Instrumentation and Measurement, Vol. 70, pp. 1-9, 2021.
14. K. Vijila, Rani, S. Albert Jerome, P. Josephin Shermila, L. K. Shoba, and M. Eugine Prince. "Automatic segmentation of lung tumor from X-ray images using advance novel semantic approach." IETE Journal of Research, Vol. 69, no. 7, pp. 4087-4098, 2023.
15. Hanan AR, Akkar, and Suhad Qasim G. Haddad. "Diagnosis of lung cancer disease based on back-propagation artificial neural network algorithm." Engineering and Technology Journal, Vol. 38, no. 3B, pp. 184-196, 2020.