Indian Journal of Natural Sciences *www.tnsroindia.org.in ©IJONS*

Vol.14 / Issue 78 / June / 2023 *International Bimonthly (Print) – Open Access* *ISSN: 0976 – 0997*

RESEARCH ARTICLE

Optimizing Convolution Neural Networks with Genetic Algorithm for Lung Cancer Classification

Ritu Nagila[1]* and Abhishek Kumar Mishra[2]

[1]Research Scholar, IFTM University, Moradabad, Uttar Pradesh, India
[2]Associate Professor, IFTM University, Moradabad, Uttar Pradesh, India

Received: 09 Apr 2023 Revised: 29 Apr 2023 Accepted: 15 May 2023

***Address for Correspondence**
Ritu Nagila
Research Scholar,
IFTM University,
Moradabad, Uttar Pradesh, India.

ABSTRACT

One of the leading causes of cancer-related death in both men and women is lung cancer. According to the World Cancer Research Fund International, 2.2million new instances of this disease were reported in 2020. Lung nodules are essential for cancer screening, because early detection enables therapy and speeds up patient's recovery. Even though a lot of effort is being done in this area, accuracy needs to be improved in order to boost the patient persistence rate. Traditional technologies, however, fail to accurately separate cancer cells of various kinds, and no system has gained greater dependability. In this work, an efficient method is suggested to not only quickly identify lung cancer lesions but also to improve accuracy. Medical personnel are prompted to treat patients in a safer and more efficient manner by early diagnosis and classification of the disease. The development of deep learning technology has made it possible to investigate the role of specific genes in diseases like lung cancer. In order to classify a patient's lung cancer status, this study used genetic algorithm (GA) as a feature (genes) selection method for the convolutional neural network (CNN). With noteworthy predictive ability, genetic algorithm successfully discovered genes that categorise patient lung cancer status. The proposed framework has a 98.97% accuracy rate.The suggested study shows that accuracy is increased, and execution time decreases also the data is compiled utilising genetic and convolutional neural network method.

Keywords: Cancer, Big data, Cancer Detection Technique, Feature Selection, Genetic Algorithm, Convolutional Neural Network, Image segmentation, Optimization.

INTRODUCTION

Cancer is a condition when the body's cells grow out of control. Lung cancer is a term used to describe cancer that first appears in the lungs. Lung cancer starts in the lungs and can spread to the lymph nodes or other body organs, like the brain. The lungs may potentially become infected with cancer from other organs. Metastases are the term used to describe the spread of cancer cells from one organ to another. Typically, there are two basic forms of lung cancer: small cell and non-small cell (including adenocarcinoma and squamous cell carcinoma). They develop differently and respond to treatment in different ways. The occurrence of non-small cell lung cancer exceeds that of small cell lung cancer. The type of tumour present determines how [2] lung cancer is treated. To achieve improved survival rates, it is crucial to classify distinct tumour kinds. However, it is difficult to classify lung cancers. For monitoring and making decisions regarding malignant lung growths, computed [3] tomography (CT) is the most important image mode. The early detection of lung cancer by CT allows physicians to recommend more effective therapy. For ordered and comforting treatment, estimating and healing elements for dispersed illness with definite malignancy stages are needed. The patient's continuation rate has a strong relationship to how quickly the lung cancer period ends. Clinical terminology refers to the condition as odd hyperplasia, and it affects people in ways that go well beyond the 200 types. Computer-Aided Diagnosis (CAD) technologies have become significant and indispensable in the categorization problem of lung cancer in order to simplify this diagnosis procedure and improve the odds of survival. With the use of many components like medical image processing, artificial intelligence, and computer vision, CAD is a technique that aims to lower the rate of false positive or negative diagnoses. Image pre-processing, feature extraction, and classification are the three primary processes of this technique, as shown in Figure

1. Convolutional neural networks (CNNs) are a deep learning subfield [1] that have been combined of the widely used techniques, particularly in the field of medical imaging. One of the important uses for this technology is to aid doctors in the early detection of lung cancer, which can lower mortality rates. However, a wide range of factors influence system diagnosis precision. The use of computer-aided technology for this aim has been increasingly attractive to scientists in recent years. In this study, lung cancer images are categorised using a meta-heuristic optimised CNN classifier that is applied to trained network models for visual datasets. The convolution layer in Fig. 2 analyses the output of the neurons that are input-connected to the local area. The calculation is done by multiplying the weights of each neuron by the region to which they are linked (the activation mass). The pooling layer's primary objective is to [1] subsample the input image in order to lessen the computational workload, memory requirements, and the number of parameters (over fitting). The neural network becomes less sensitive to picture displacement when the input image is smaller (independent of the position). There are, however, a variety of techniques for improving the convolutional neural network learning process, and there aren't many studies about deep learning-based neural networks and their uses. The weights and biases in the CNN models are optimised in the current work using a novel method based on the genetic algorithm. The new approach is then put to the test against 10 well-known classifiers using datasets for lung cancer, including LIDC (Lung Image Database Consortium) Database. According to experimental findings, this refined approach performs more accurately than other categorization methods.

LITERATURE REVIEW

Many research works have been conducted on different types of cancer detection and prediction. The research works applied different approaches for detecting cancer of different types. Different algorithms were developed and compared. Some of the previous research works on lung cancer detection are discussed in this section. In 2020, Long Zhang *et al* [1] proposed a method for optimizing convolution neural network with whale optimization algorithm. This approach was then put to the test against 10 well-known classifiers using two datasets for skin cancer, including DermI S Digital Database and Dermquest Database. According to experimental findings, this refined approach performed more accurately than other categorization methods.

Joey Mark Diaz *et al* proposed a method [2] for in 2018. In order to classify a patient's lung cancer status, this study used genetic algorithm as a feature (genes) selection method for the support vector machine and artificial neural network. With noteworthy predictive ability, genetic algorithm (GA) successfully discovered genes that categorise patient lung cancer status. Chapala Venkatesh *et al* investigate [3] a method for optimizing neural network in 2022. In this study, a screening method that was efficient and increases accuracy was suggested for identifying lung cancer lesions. In this process, the cuckoo search algorithm was used to establish the ideal criteria for partitioning cancer nodules, and Otsu thresholding segmentation was used to achieve perfect isolation of the chosen area. The pertinent lesion features were extracted using a local binary pattern. Based on the obtained features, the CNN classifier can determine if a lung lesion is malevolent or not. The proposed framework has a 96.97% accuracy rate. The suggested study shows that accuracy is increased, and the data is compiled utilising genetic and particle swarm optimization.
C Vankatesh *et al.* [4] proposed a hybrid method for classifying lung cancer in 2019. The identification of cancer in CT scans was addressed in this study using a hybrid strategy that combines a genetic optimization algorithm, SVM classification, and a unique feature selection method. With the use of this technique, clinicians can accurately identify lung nodules at an early stage. A feed-forward neural network was trained using GA by Montana et al. [5] in one of the earlier efforts. The authors demonstrated that by adjusting the weights during the neural network learning process, GA increased accuracy in comparison to back-propagation neural networks.

Chuang Zhu *et al* [6] proposed a categorization of breast cancer histopathology images using numerous small Convolutional Neural Networks (CNNs) in 2019. A hybrid CNN architecture with a local model branch and a global model branch was used in this study. This hybrid approach gains a stronger capacity for representation through local voting and two-branch information merging. Second, by incorporating the suggested Squeeze-Excitation-Pruning (SEP) block into this hybrid model, it was possible to learn the importance of each channel and eliminate the redundant ones. The suggested channel pruning strategy can generate improved accuracy with the same model size while lowering the danger of over fitting. In 2021 F. Taher *et al* [7] discussed the accuracy, sensitivity, and specificity of various machine learning techniques, including convolutional neural networks (CNN), support vector machines (SVM), artificial neural networks (ANN), multi-layer perceptrons (MLP), K-nearest neighbour (KNN), and the entropy degradation method (EDM). In comparison to other approaches, the CNN approach utilising a short dataset yields the greatest results (96% accuracy), whereas EDM yields the lowest accuracy (77.8%).

Samuel Zumbuka *et al* [8] showed a study in 2021.This study listed the predictive techniques and instruments for routinely detecting, predicting, contrasting, or classifying lung cancer. Thirty-four (34) papers relating to the applications of machine methods were chosen and taken into consideration using Boolean keyword searches in various journal databases and filters. The analysis's findings indicate that the majority of studied utilised combined classical models, with ANN and deep learning being used less frequently. Additionally, only a small portion of the data set—which also included personal characteristics, X-ray images, symptoms, etc.—used data sets other than CT scan images. The results of this study indicated several restrictions and future research possibilities for lung cancer and associated diseases.

METHODOLOGY

Physicians identify lung cancer after carefully analysing CT images, a process that takes a lot of time and is not always correct. Modern optimization methods and image processing techniques were necessary to produce imagery that was as accurate, functional, and efficient as feasible. The suggested technology will help physicians accurately detect lung nodules at an early stage and examine the internal structure. As a part of the contribution, certain errors with the identification of lung cancer are mentioned here. Using a cutting-edge segmentation technique called Otsu there sholding and cuckoo search optimization, the region of interest is located. With this suggested partitioning method, nodules of various sizes and shapes can be precisely separated using only a few parameters.

Convolutional Neural Network (CNN)-Fig 3 represents a CNN architecture. CNNs frequently use pooling layers when dealing with complex layers. They are mostly employed for reducing the tensor's size and accelerating estimations. These layers are all quite basic. Therefore, the image must be divided into smaller sections in the pooling layer, with the maximum value being chosen for each portion and each piece was then completed through some technique. It is positioned appropriately in the output after being portioned. Rectified linear units, or RELUs, are a type of hidden layer. The activation function is mostly utilised in CNNs, which are neural networks.

Genetic Algorithm- Fig 4 represent the concept of Genetic Algorithm (GA). Adaptive heuristic search algorithms based on genetics and natural selection theory are known as [9] genetic algorithms (GAs). As a result, [6] they offer an innovative utilisation of a random search approach to optimization issues. GAs are not random, despite the fact that they are randomised; rather, they use previous data to focus the search on areas of the search space where it will perform the best. In particular, those who adhere to the "survival of the fittest" ideas first established by Charles Darwin, the fundamental GA techniques are created to imitate processes in natural systems necessary for evolution.
After a randomly generated initial population, the method evolves through three operators in accordance with Natural Selection:
1.Selection-this corresponds to the survival of the fittest;
2.Crossover- this depicts mating between individuals;
3.Mutation- this generates random modifications.

Selection Operator
a) **Principal concept**: prefer superior people in order to pass on their genes to the upcoming generation.
b) Each individual's goodness is based on their level of fitness.
c) Fitness can be assessed using a subjective evaluation or an objective function.

Crossover Operator
a) The selection operator selects two people from the population.
b) There is a random selection of a crossover location along the bit strings.
c) Up until this point, the values of the two strings are switched.
If S1=000000 and s2=111111 and the crossover point is 3 then S1'=111000 and s2'=000111
d) The two new offspring produced by this pairing are added to the population's next generation. This procedure likely results in the creation of even better people by recombining pieces of good people.

Mutation Operator
a) With some low probability, a portion of the new individuals will have some of their bits flipped.
b) Its goal is to prevent early convergence and preserve population variety.
c) A random stroll through the search space is brought on by mutation alone.

Although genetic algorithms have been employed for feature selection, parameter optimization, and rule reduction, they are best for finding the best solution to a problem. But there is still a lot that can be done to enhance GAs, both generally and specifically for cancer research. In order to identify topics for future work, this section elaborates on areas that could use improvement.

Design: Combining CNN and GA
In this section, we will go through how we evolved the CNN weights for the histopathology lung image classification challenge using GA's global search feature. An effective CNN architecture for categorising binary histological lung image data is planned. The block diagram of our model is shown in Fig 6. Below is a description of the layers that the model uses.

Input Layer
The convolutional layer receives the loaded input images from this layer and processes them. Lung images from the LIDC dataset are used as inputs in our approach.

Convolutional Layers
In order to create the feature map, convolutional layers apply the convolution operation to the input images using kernels (filters). The network has eight convolutional layers and numerous 3-by-3 filters. Each layer has eight filters for the first two levels, sixteen for the third and fourth layers, twenty-four for the following two tiers, and ultimately thirty-two for the final two layers. With either a uniform or normal distribution, these weight filters are started at random. All of the weights of the network, including the weights of the convolutional filters and fully connected layers, are stored in a matrix for subsequent computation because employing matrix form simplifies the operation of CNN. However, 1-dimensional vectors are used to hold the initial [10] population of GA.

Pooling Layer- Convolution layers stage is followed by this step. The average pooling function is used to minimise the features map dimension. The average of each filter in a convolutional layer is determined by this function.

Activation Layer
To increase the rate at which the learning process converges, all convolutional and pooling layers are subsequently followed by a activation function.

Fully-Connected Layer
Following convolution and pooling processes, a process known as flattening is used to convert the entire features matrix into a single vector. The network's fully connected layer then receives this vector as its input neurons. To put it another way, the first layer is a vector of features from the convolutional and pooling layers, and the last layer is a vector of size 2, which is equal to the number of classes in our binary classification, i.e., the Benign and Malignant classes.

Training Process
The entire system is taught following the creation of the CNN blocks. Training is the process of iteratively changing the network's weights to provide the most accurate results. Three alternative optimization methods are used in our network, and their classification accuracy is contrasted: Adam optimization, mini batch gradient descent, and GA.

Data Set-The performance of our suggested model for binary classification of lung CT images is evaluated using the LIDC dataset, a public accessible dataset. To assess the performance and demonstrate the effectiveness of their proposed classifiers, researchers need a distinctive dataset. As a result, most of the researchers used the LIDC dataset, which contains images of lung cancer, as a well-known database addressing the problem of classifying lung cancer. Examples of CT-scan images of lung cancer patients from the LIDC dataset are shown in Fig-7. To train our classifier, we employed 70% of the available images in this study, and the remaining 30% were used to test the proposed model. We did this by randomly dividing the LIDC dataset into two sets, a training set and a testing set. Both sets are Patients utilised to form the training set and the test set were divided patient-wise, therefore they are not the same patients. Additionally, we are categorising the CT-images independently of their sub-classes and magnification factors because our focus is on the binary categorization of CT images into Benign and Malignant cases.
Pseudo Code -In the following, the general pseudo code for Optimizing CNN through GA is given as follows.
Algorithm 1
Create the CNN architecture;
Generate the initial weights of the model randomly;
Store the weights of the model in a matrix;
Set the parameters of GA;
Convert the weight matrix to the vector of initial population;

while Termination condition is not true **do**
Start the CNN process;
Predict the output using CNN;
Calculate the fitness function;
Sort individuals based on fitness value to select fitter individuals for producing
the offspring of the next generation;
Applying a single-point crossover operation;
Applying mutation operation to add a random value to a randomly selected
gene with probability of 0.1.
end while

The CNN using GA optimization approach is described in above Algorithm. In our implementation of the suggested GA optimizer, we make use of the following variables:

(i) The number of solutions per population varies from 20 to 60;
(ii) The number of generations varies from 100 to 1000;
(iii) The mutation rate varies from 0.1 to 0.01; and
(iv) The network weights are initially distributed using both normal and uniform distributions.

RESULTS AND DISCUSSION

Based on many measures, including classification accuracy, execution time we report the findings of the lung cancer CT image classification provided by the LIDC dataset. The binary classification of the input images is the output. Every image is sent to the network and given a Benign or Malignant case. This study combines convolutional neural network and genetic algorithm to propose an effective and automated method for detecting lung cancer. The process starts with image pre-processing, followed by image segmentation, feature extraction, and classification. The procedure is evaluated using lung CT-images. Figure 8 revealed that the proposed model is able to classify benign and malignant lung cancer patients with accuracy rates of 98.97%.Additional it also reduces the exception time which can be shown in figure 9. This result is consistent with the fact that the classification accuracy of lung cancer is frequently higher when optimising a convolution neural network with a genetic algorithm.

CONCLUSION

Because of the ability to classify data, machine learning techniques like Convolution Neural Networks (CNN) have been frequently used in CAD systems. The network weights, including the weights of all convolutional filters and the weights of connecting edges in the fully-connected layer, have a significant impact on the accuracy of extracted features in CNN. Therefore, these weights are very important for classification accuracy. In this study, we suggested employing the Genetic Algorithm (GA) to optimise the CNN weights for lung cancer classification problem. The five levels of the proposed algorithm are input layer, convolution layer, pooling layer, activation layer, fully-connected layer, and training process. The weights of these layers have a significant impact on the classification accuracy. Utilizing selection, crossover, and mutation operators, the weights are evolved. Using the LIDC dataset, we compare our suggested method with using CNN approach. To investigate metrics like accuracy, execution time, and the impact of genetic parameters on the proposed method, we carried out numerous trials with varied batch sizes and iteration counts.

REFERENCES

1. Long Zhang, Hong Jie Gao, Jianhua Zhang, Benjamin Badami, “Optimization of the Convolutional Neural Networks for Automatic Detection of Skin Cancer”, 2020.

2. Joey Mark Diaz, Raymond Christopher Pinon, Geoffrey Solano, "Lung Cancer Classification Using Genetic Algorithm to Optimize Prediction Models" ,2014.
3. Chapala Venkatesh, Kadiyala Ramana, Siva Yamini Lakkisetty, Shahab S. Band,Shweta Agarwal and Amir Mosavi, "A Neural Network and Optimization Based Lung Cancer Detection System in CT Images", Frontiers in Public Health, www.frontiersin.org, June 2022 | Volume 10 | Article 769692.
4. CVenkatesh, Polaiah Bojja, "A Novel Approach for Lung Lesion Segmentation Using Optimization Technique",Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495, DOI 10.29042/2019-4832-4837.
5. D. J. Montana and L. Davis, "Training feedforward neural networks using genetic algorithms." in IJCAI, vol. 89, 1989, pp. 762–767.
6. C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancerhistopathology image classification through assembling multiple compact CNNs,"BMC medical informatics and decision making, vol. 19, no. 1, p. 198, 2019
7. F. Taher, N. Prakash, A. Shaffie, A. Soliman, A. El-Baz, "An Overview of Lung Cancer Classification Algorithms and their Performances",IAENG International Journal of Computer Science, 48:4, IJCS_48_4_19, Volume 48, Issue 4: December 2021.
8. Samuel Zumbuka, Yusufu Gambo and O. Sarjiyus, "Application of Machine Learning in LungCancer",IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 9, September 2021, ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72.
9. Ritu Nagila, Abhishek Kumar Mishra, "A Comprehensive Analysis of Big Data Analytics Applications in Cancer Detection Methods",Industrial Engineering Journal, Volume 15 Issue 12 * December 2022, ISSN: 0970-2555.
10. KatsuoUsuda, Shun Iwai *et al,* "Novel Insights of T2-Weighted Imaging: Significance forDiscriminating Lung Cancer from Benign Pulmonary Nodulesand Masses", Cancers 2021, 13, 3713. https://doi.org/10.3390/cancers13153713.
11. Ibrahim M. Nasser, Samy S. Abu-Naser, Palestine, "Lung Cancer Detection Using Artificial Neural Network", International Journal of Engineering and Information Systems (IJEAIS), ISSN: 2000-000X ,Vol. 3 Issue 3, March – 2019, Pages: 17-23.
12. Lal Hussain et al, "Lung Cancer Prediction Using Robust Machine Learning andImage Enhancement Methods on Extracted Gray-LevelCo-Occurrence Matrix Features",Appl. Sci. 2022, 12, 6517. https://doi.org/10.3390/app12136517.
13. Antony Judice A, Dr. K. Parimala Geetha, "A Novel Assessment of Various Bio-Imaging Methods for Lung Tumor Detection and Treatment by using 4-D and 2-D CT Images", International journal of Biomedical science, June 2013 Vol. 9 No. 2 Int J Biomed Sci www.ijbs.org.
14. Parthasarathy G *et al,* "PREDICTION OF LUNG CANCER USING DEEP LEARNING ALGORITHM", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 4 April 2020 | ISSN: 2320-2882.

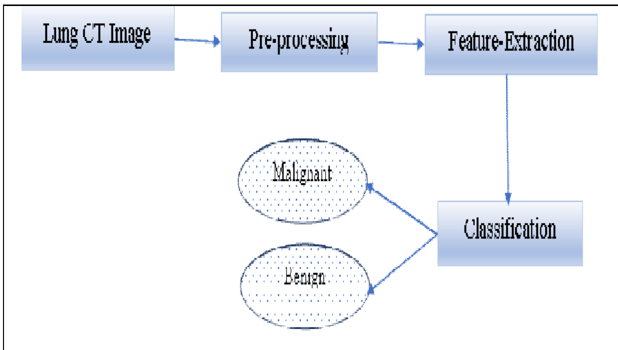

Fig 1: An overview of a simple CAD system for lung cancer classification task.

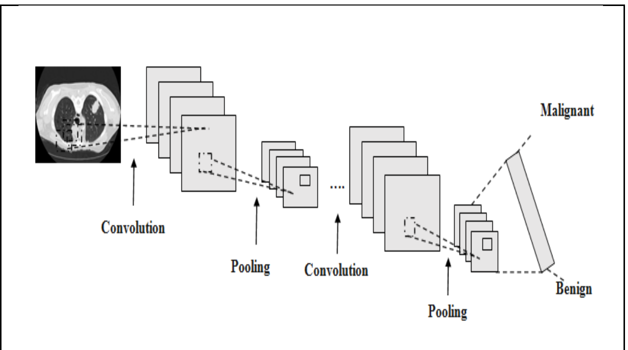

Fig 2: A simple lung cancer detection using ordinary CNN

Ritu Nagila and Abhishek Kumar Mishra

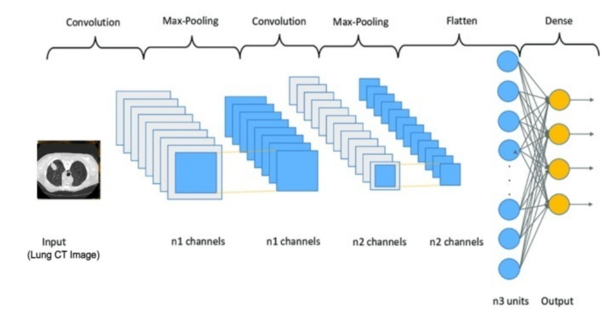

Fig.3- Architecture of CNN

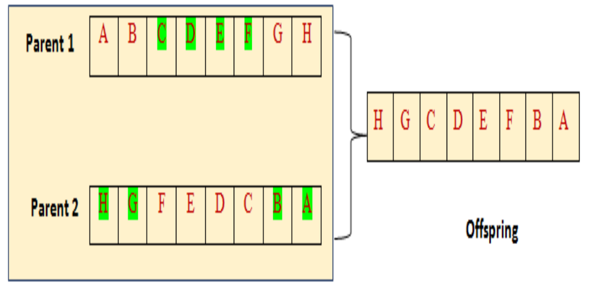

Fig.4 Concept of GA

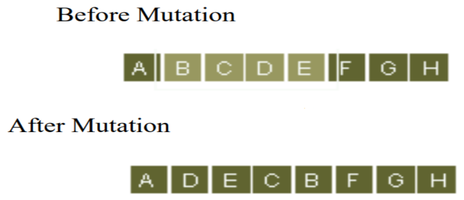

Fig.5 Before and After Mutation

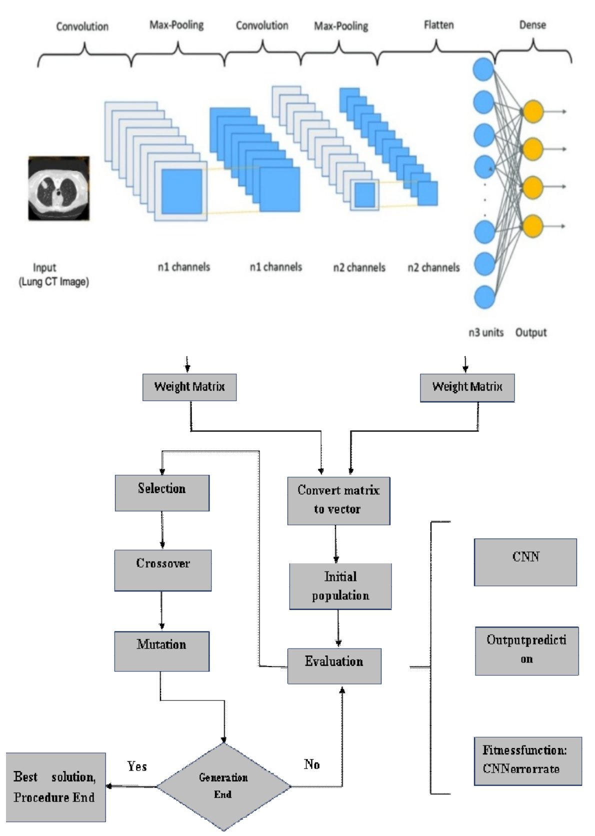

Fig. 6 Block diagram for utilising GA to optimise the CNN's parameters.

Fig.-7 CT-images of lung cancer

Ritu Nagila and Abhishek Kumar Mishra

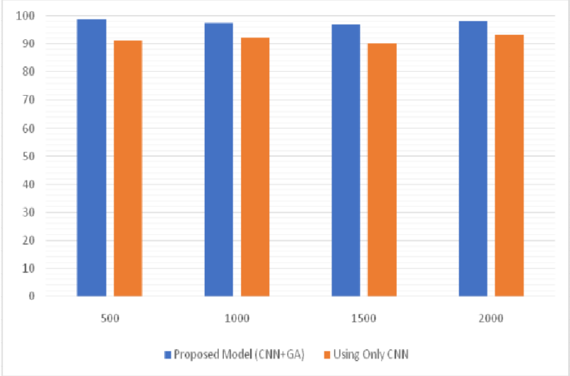

Fig.-8 Illustrates the accuracy of the proposed model and the only CNN method in percentage

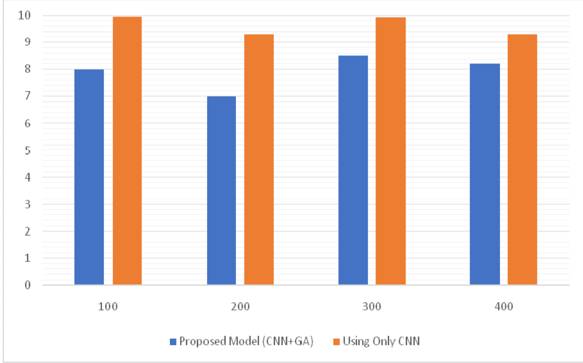

Fig.-9 Illustrates the execution time of the proposed model and the only CNN method