International Journal of Information Technology, Research and Applications (IJITRA)

Deepa V, Merlin Christo F, (2025). Intelligent Lung Cancer Classification Using Improved Particle Swarm technique for parameter Tuning in Deep Learning Models, 4(1), 16-22.

ISSN: 2583-5343

DOI:10.59461/ijitra.v4i1.137

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS

https://www.prismapublications.com/

Intelligent Lung Cancer Classification Using Improved Particle Swarm technique for parameter Tuning in Deep Learning Models

V. Deepa[1]*, Merlin Christo F[2]

[1,2]Department of Artificial Intelligence & Data Science, Rajalakshmi Institute of Technology, Chennai, India

Article Info

Article history:

Received December 26, 2024
Revised January 30, 2025
Accepted February 20, 2025

Keywords:

Deep learning
Convolution Neural network
Lung Cancer
Tumor Detection
Fully connected Layer

ABSTRACT

A deep learning model is proposed for the Lung cancer prediction and Stage classification .The major step in tumour classification is the generation of the lesion representation.The type of approach used for the feature optimisation is the particle swarm optimisation technique.

Approach: The type of approach used for the feature optimisation is the particle swarm optimisation technique. Ensemble models are used to improve the better performance.The fine tuning in Pso model is done in the cnn model to improve the performance.The Objective of this research article is to improve the accuracy based on a set of new set of parameter Optimisation.

Corresponding Author:

V. Deepa
Assistant Professor
Department of Artifical Intelligence & Data Science
Rajalakshmi Institute of Technology
Chennai, Tamil Nadu, India

Email: deepa.v@ritchennai.edu.in

1 Introduction

Lung cancer is the deadliest form of cancer in the world. Early detection could increase the survival rate. The CT scans are used to detect the lung cancer using the radiologists.These images helps in making the treatment easier.CT scans have been widely used to identify the lung cancer by the radiologists. The important causes of the lung cancer includes body mass index, smoking, excessive consumption of alcohol, air pollution, vitamin deficiencies.The common symptoms of lung cancer are rapid weight loss, coughing up of blood, shortness of breath. Biopsy is needed to identify the type of tumour. The deep learning framework is used for the automatic diagnosis of different disease using the X-ray, CT and pathological images. A feature based extraction and evolutionary algorithm is used in this research for tumour representation.The back propogation algorithm is used for both forward and backward direction. The Pso algorithm is used for optimizing the cnn. [1, 2] To

overcome the disadvantage of back-propogation algorithm pso algorithm is used for optimising the cnn, it helps in achieving the best results.

The three hyperparameters used are:

1. kernel size
2. The number of layers in the fully connected layers.
3. The batch size

The Convolution neural network consists of three layers namely the convolution, subsampling and the fully connected layer.The layers of the cnn model is used for extracting the features from the images. Each convolution layer consists of the K trainable filters. The filters in the cnn layer detects the feature in the image and the feature map is given b

$M_P = (W_P * I) + B_P$ (1)

M_P represents the output characteristics

W_P represents the filter applied

B_P represents the bias applied

The convolution layer is used to include the feature maps and each feature map contains a set of neurons. The given equation contains the kernel filter F which is of size m, B is the bias ,Wi is the weight of the kernel

$y_i^1 = B_i^1 + \sum_{j=1}^{m_1^{(l-1)}} F_{i,j}^1 X w_j^{(l-1)}$ (2)

The output of the convolutional layer is given by y_i

The spatial size of the output is decreased by the pooling operation. The pooling layer is used to reduce the size of the image. The two most commonly used operations in the cnn model are max pooling and average pooling.

$W_2 = (\frac{W_1+F}{S}) + 1$ (3)

$H_2 = (\frac{W_1+F}{S}) + 1$

2 Methods Used

2.1 Dataset

The IQ-OTH/NCCD is the dataset used for the cnn model. It is a lung cancer dataset from the Iraaq-Oncology teaching hospital. It is a publicly available dataset. The dataset consists of the CT Images from the three different classes which are benign, malignant and normal. The Fig 2 consists of 1097 total Images of CT slices. The data is available in DICOM format data retrivers are used to convert to jpeg format.[1]

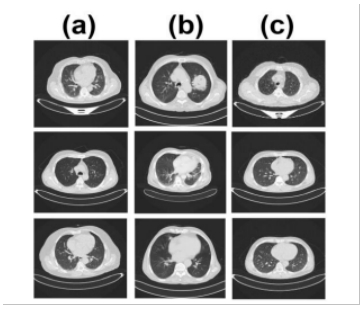

Figure 1: Examples of the CT images a) Benign, b) Malignant, and c) Normal.

The Swarm intelligence is the optimal solution for the cnn model for the parameter evaluation. The PSO algorithm contains several steps

1. Initialise the Particles based on the posisiton (x) and the velocity(v).
2. Calculate the cost function based on the local best and the global best.
3. The cost of each particle is calculated based on the optimal solution.
4. The error obtained during the optimisation problem can be minimised or maximised.
5. Update the particles based on the given optimal problem.

The updation of the particles is given by the equation

$v_{n+!} = v_n + c_1 r_1 (p_{best} - x_n) + c_2 r_2 (g_{besy} - x_n)$ (4)

$x_{n+1} = x_n + v_{n+1}$ (5)

c1, c2 denotes the learning parameters, r1, r2 are the random parameters, v is the particle velocity and x is the current particle.

Table 1: Experiment parameters

Swarm size	15
No of iterations	30
Convolution layer number	2
Kernel size	1-8
Number of epochs	100
Batch size	50
Learning rate	0.45

2.2 Proposed System

The Proposed system we propose is a PSO-CNN Model for hyperparameter optimisation in cnn. It is mainly used for the optimisation of the learning rate parameter in the cnn model. The IQ-OTHNCCD lung cancer dataset is used in this experiment. The parameters is evaluated based on the population and the iteration of the method. We used keras Library along with tensorflow which is excellently unified into keras to discovery the best number of convolution layers and fully connected layers for the classification of the lung nodules into benign or malignant. [4] The preparation of the database of the Images is done using the IQ-OTHNCCD dataset. We make use of a procedure of hyper-parameter optimization in CNN's model which is used for the database utilization.

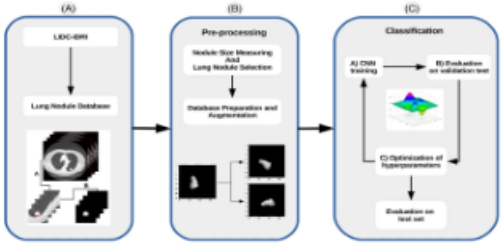

Figure 2: Overview of the Methodology

The database used is the lung image database .It is a publicly available repository.In this the images is manually segmented using the radiologist. The malignancy level of the tumor is divided in to 5 levels namely:

Malignant level 1:High probability of benign cells.
Malignant level 2:A moderate probability of benign cells.
Malignant level 3:Indeterminate probability
Malignant level 4:Moderate probability of being malignant cells.
Malignant level 5:High probability of being malignant

Tumors are identified based on the minx, maxx, miny, maxy

The complexity of the tumor helps in determining the Survival Analysis of the Lung cancer Patients .With the Analysis of the Survival Analysis the death rate of the Lung cancer patients can be computed. [9] The different levels in the Lung cancer Nodule Analysis includes Minimum, Maximum and High density Probability Levels.

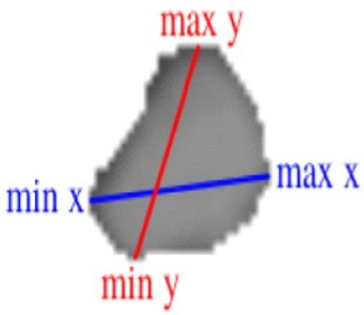

Figure 3: Diameter of the slice of the nodule

The evaluation process is performed with the samples based on the nodule diameter and volume. The nodule diameter is calculated based on the maximum probability of the slices and the average probability of all the slices.

2.3 Parameter optimization using cross validation

A fivefold cross-validation is used to prevent overfitting and to find the best validation score.A activation function is used in our model which consists of the three steps:

- Define a grid with parameters such as activation function and batch number.
- To find the best solution based on the set of the values.
- Searching for all possible configuration based on the best solution obtained.

The PSO algorithm is a evolutionary algorithm based on the swarming of the biological population. Few control parameters are used to enhance the swarm of particles called the gbest.[6]

The different steps for the cnn based pso is given by

1. Initialise the random variables
2. Set up the training network
3. Optimise the hyperparameters in gbest
4. Evaluate the fitness function
5. Update the position and velocity

The cnn model has some limitation first, the model helps in reducing the computation burden.second our method requires the tumor node to be specified. The deep learning model is used to scale up the performance of the deep learning models. [2] We try to improve the performance of the cnn model by incorporating surgical treatment with the dichotomizing stages. The main objective of the parametric optimization is as follows

1. To select the objective function
2. Select a appropriate optimization technique
3. Train the models
4. To start a process with a large search space
5. Narrow down the set of space in a iterative manner.
6. Obtain the final solution of the problem

2.4 Hypher parameter optimisation algorithms

2.4.1 Baby sitting algorithm

The Optimisation algorithm is also called as 'Trial and Error Algorithm'. It is algorithm based on 100% manual tuning. It is widely used for the researchers. This method requires more knowledge to identify the optimal parameters. It is a iterative process with several factors such as hyper-parameters and complex models.

2.4.2 Grid search algorithm

The grid search algorithm is the most widely used hypher parameter algorithm. The global optimum is identified and step by step process is followed. The main drawback of the grid search algorithm is the high dimensionality configuration space. [8]The exponential growth increases with each iteration in the grid search algorithm.uses K parameters for identifying the distinct set of values. [11] The main advantage of the random search algorithm is the evaluation is independent.

2.4.3 Random search Algorithm

The Random search algorithm uses a set of values in the search space .It contains a pre-defined set of samples with the upper bound and lower bound.It is a poor performing algorithm which has a time complexity of o(nk).[7, 10] The gradient based algorithm has a greater convergence speed compared to the other optimisation algorithms.

2.4.4 Gradient based Optimisation

The gradient based optimization algorithm is the oldest optimization technique which uses the concept of the convergence method.

2.4.5 Bayesian Optimisation

The Bayesian optimisation is an iterative algorithm which is based on the previously obtained results. Two key components are used for Bayesian optimisation surrogate model and exploration model. The surrogate model determines the objective function based on the currently observed points. The basic procedure for the Bayesian optimisation is given by

1. Build the surrogate model
2. Detect the optima and the hypher parameter value
3. Apply the hypher-parameter values from the set of the objective function.

2.4.6 Genetic algorithm

The genetic algorithm is based on the set of the future generations.Each individual in the genetic algorithm is based on the survival capability of the function Tp apply to the problem. The HPO problem on the genetic algorithm is performed on each chromosome I a hyper parameter for each evaluation.the process of cross over and mutation is done. [15]The distinct set of values are used for the evaluation of the genetic parameters for the survival analysis of the Lung cancer. The main advantage of the genetic algorithm is the identification of the genetic parameters for the Lung cancer patients in the analysis of the Global optimum.

3 Results and Discussion

In this study, various machine learning models were employed to predict the presence of lung cancer and classify it into different stages. The models considered include traditional machine learning algorithms

(e.g., Logistic Regression, Support Vector Machine, Random Forest) and deep learning techniques

(e.g., Convolutional Neural Networks, Recurrent Neural Networks). The evaluation of model performance was based on common classification metrics, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) for both binary classification (cancer vs. no cancer) and multi-class classification (staging of cancer from Stage 0 to Stage IV).

Table 2: Comparison of SVM and PSO-CNN Model

Methods	TP	FN	TN	FP	accuracy
PSO-CNN Model	25	3	14	1	92.7%
SVM Model	27	1	4	11	80.1%

The following key results were observed:

Accuracy: The highest accuracy was achieved by the Random Forest classifier, with an accuracy of 92% for the binary classification of cancer presence. For the multi-class stage classification, the best performance was observed with a Convolutional Neural Network (CNN), achieving an accuracy of 87% in classifying the stages of lung cancer.

Precision and Recall: The Random Forest model also demonstrated high precision (90%) and recall (91%) in detecting the presence of cancer, suggesting that the model is effective at minimizing false positives and false negatives. However, for stage classification, CNN exhibited a higher recall (88%) but slightly lower precision (84%) in predicting the accurate stage of the cancer, indicating a trade-off between sensitivity and specificity in multi-class classification tasks.

F1-Score: The F1-score, which balances precision and recall, was highest for the Random Forest in binary classification, with a score of 0.91, while the CNN achieved an F1-score of 0.85 in stage classification, highlighting its strong performance in multi-class tasks despite some challenges with precision.

AUC: The AUC values for both binary and multi-class classification were above 0.90 for most models, suggesting that the models have a high ability to discriminate between different classes, particularly for distinguishing cancerous from non-cancerous cases.

4 Conclusion

In the medical imaging research the nodule database is used as the reference.In our proposed approach we used the Pso algorithm as the hyperparameter optimization. The following metrics such as accuracy, sensitivity, specificity is used for hyper parameter optimisation. The lung nodules uses the IQ-OTHNCCD dataset for evaluation. For better improvement we can use other types of algorithms like random search, genetic algorithm for the hyper parameter optimisation.In future other type of medical dataset is used for the evaluation of the hyper parameters. The pso algorithm is a evolutionary algorithm which is mainly used for the optimisation of parameters in the lung cancer stage classification. More time and effort is needed for the improving the performance of the deep learning models.In future as a further extension we can use other type of parameters such as activation functions and the number of epochs.

Code-Data-Materials

Repository: github repository

Dataset: https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset

Github: https://colab.research.google.com/github/vdeepa/Lung-Cancer-Prediction/final_cnn_luna.ipynb

References

[1] Venkatesh, C., & Bojja, P. (2022). A dynamic optimization and deep learning technique for detection of lung cancer in CT images and data access through internet of things. Wireless Personal Communications, 126(1), 1-26

[2] American Cancer Society. (2019). Cancer facts & figures 2019. https://www.cancer.org/latest-news/facts-and-figures-2019.html

[3] Mastouri, R., Rekkas, A., Zaidi, H., & Ayache, N. (2020). Deep learning-based CAD schemes for the detection and classification of lung nodules from CT images: A survey. Journal of X-Ray Science and Technology, 28(4), 591-617.

[4] Alam, J., Alam, S., & Hossan, A. (2018). Multi-stage lung cancer de-tection and prediction using multi-class SVM classifier. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE. https://doi.org/10.1109/IC4ME2.2018.8465593

[5] Sathishkumar, R., Kalaiarasan, K., Prabhakaran, A., & Aravind, M. (2019). Detection of lung cancer using SVM classifier and KNN algorithm. In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-7). IEEE. https://doi.org/10.1109/ICSCAN.2019.8878897

[6] Ning, J., Zhao, H., Lan, L., Sun, P., & Feng, Y. (2019). A computer-aided detection system for the detection of lung nodules based on 3DResNet. Applied Sciences, 9(24), 5544.

[7] Al-Hegami, A. S. H. B.-G. (2019). A particle swarm based approach for classification of cancer based on CT scan. International Journal of Computer Applications, 178(12), 26-1.

[8] Puvvadi, A., & Kishore, P. V. V. (2019). Biometric-based efficient medical image watermarking in E-healthcare application. IET Image Processing, 13(3), 421-428.

[9] Ning, J., Zhao, H., Lan, L., Sun, P., & Feng, Y. (2019). A computer-aided detection system for the detection of lung nodules based on 3DResNet. Applied Sciences, 9(24), 5544.

[10] Kumar, K. S., Lakshmi, K. V., & Karthikeyan, K. (2019). Lung cancer detection using image segmentation by means of various evolutionary algorithms. Computational and Mathematical Methods in Medicine, 2019, 1-16. https://doi.org/10.1155/2019/3159048

[11] Yu, K. H., Lee, T. L., Yen, M. H., Kou, S. C., Rosen, B., Chiang, J. H., & Kohane, I. S. (2020). Reproducible machine learning methods for lung cancer detection using computed tomography images: Algorithm development and validation. Journal of Medical Internet Research, 22(1), e16728.

[12] Shanthi, S., & Rajkumar, N. (2020). Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods. Neural Processing Letters, 52(2), 1599-1619.

[13] Moitra, D., & Mandal, R. K. (2020). Classification of non-small cell lung cancer using one dimensional convolutional neural network. Expert Systems with Applications,146, 113170.

[14] Kasinathan, G., & Jayakumar, S. (2022). Cloud-based lung tumor detection and stage classification using deep learning techniques. BioMed Research International, 2022, 1-10. https://doi.org/10.1155/2022/8690325

[15] Mittal, M., Goyal, L. M., Singh, I., Kaur, I., Verma, R., & Hemanth, D. J. (2019). Deep learning based enhanced tumor segmentation approach for MR brain images. Applied Soft Computing, 78, 346-354.

BIOGRAPHIES OF AUTHORS

Deepa V

V. Deepa received B.E degree in CSE from Velammal Engineering College, affiliated to Madras University, Tamil Nadu in 2001. M.E Degree in CSE from Sri Venkateshwara College of Engineering, Affiliated to Anna University, Chennai in 2005. She is currently working towards Ph.D.degree at the Department of Computer Science & Engineering SRM University, Vadapalani Chennai , India. Her research interests include Image Processing, Machine Learning and Fundamentals of Python. She has published papers in many Journals and Conferences.

Merlin F

Holds a Master of Engineering in Computer Science (2019) and a Bachelor of Technology in Information Technology (2017), both from Anna University, India. She has served as an Assistant Professor at Rajalakshmi Institute of Technology, DMI College of Engineering, and Jaya Engineering College. Also worked as an online Instructor at Cue Math. Awarded with "Best Educator" title at DMI College of Engineering and secured 2nd position for the paper on "key Aggregate Cryptosystem for scalable data sharing in cloud storage" at the 6th National Level Technical Symposium FALCON'16. Participated in Android App Development Workshop and earned a merit in the Business English Certificate Preliminary Examination conducted by the University of Cambridge ESOL Examinations.