CSBPNN VS BPNN Machine Learning Network Models Significant Assessment to Lung Cancer Recognition System

OLUWAFEMI J. AYANGBEKUN[1], WILSON SAKPERE[2], OLADUNNI A. AKANNI[3]
[1]*Department of Computer Engineering, Federal Polytechnic Ayede, Oyo State. Nigeria*
[2]*Department of Computer Science, Lead City University Ibadan, Oyo State. Nigeria*
[3]*Department of Computer Engineering, Federal Polytechnic Ayede, Oyo State. Nigeria*

Abstract- Cancer of the lung have been regarded as one of the most dangerous widespread type of cancer that needed to be addressed clinically using computational intelligence. In this paper, total datasets of 893 CT lung images were collected which comprises of 200 Benign, 393 of Malignant and 300 of Normal lung dataset. The analytical performance evaluation of Back Propagation Neural Network (BPNN) technique and Chameleon Swarm (CS) Optimization technique on Lung cancer prevalence was applied and yielded significant results. CSBPNN shows better performance and significant improvement than the BPNN with 98.20% accuracy level at precision time of 38.86secs compared with BPNN with 97.40% and 59.89secs accuracy and recognition time respectively. Furthermore, the evaluation results in terms of FPR, Specificity and Sensitivity metrics produces 0.33%, 99.67% and 96.95% respectively for CSBPNN indicating better performance; compared with 1.00% FPR, 99.00% Specificity and 95.42% Sensitivity values. With the results obtained, the Chameleon Swarm Optimization technique outperformed the Back Propagation Neural Network technique in predicting and classifying lung cancer images. This was evident in the higher accuracy and precision scores of the Chameleon Swarm Optimization technique. Additionally, the Chameleon Swarm Optimization technique also showed better performance in terms of sensitivity and specificity compared to the BPNN technique. These results emphasise even more possibilities of the Chameleon Swarm Optimization technique to raise the lung cancer diagnostic and treatment accuracy. Thus, the study contributes to the existing literatures on the application of swarm optimization techniques in the field of medical research.

Index Terms- BPNN, CSBPNN, Lung Cancer, Optimization, Metrics, Performance

I. INTRODUCTION

Cancer can be described as uncontrolled proliferation of cells in any part of the body by causative agents (such as radiation ionization, human genetics system etc.). Cancer is the result of a cluster of cells undergoing uncontrolled and uneven development, leading to the formation of malignant tumors that infiltrate nearby tissue [1]. Cancer may be classified into many categories depending on the specific place of the body where it originates. In the field of Cancer Diagnosis, Histopathologists play a crucial role in examining and analyzing tissues and cells taken from odd growths. They define the type of the abnormality; if it is cancerous, provide detailed information to the clinician regarding the specific type of cancer, its severity, and its susceptibility to particular treatments [2].

Meanwhile, considered as the most fatal forms of cancer is lung cancer diseases in human across the globe [3]. Daily record of patient's influx in this category across the medical centres and hospitals kept on increasing thereby pushing the mortality rate to threshold level. This makes lung cancer to become the main reason of cancer-related mortality among men and women worldwide as shown (Table I).

Table I: Cancer Mortality Rate

Types of Cancer	Mortality Rate
Cancer of the Lung	1.80 million death
Cancer of the Breast	685, 000 death
Cancer of the Liver	830, 000 death
Cancer of the Colon and	916, 000 death

Rectum	
Cancer of the Stomach	769, 000 death

Typical Lung Cancer Symptoms include fatigue, chest pain, coughing, and coughing up blood, sore throat, shortness of breath, weariness, weight loss, and chest infection [4,5]. Early cancer identification is crucial in inhibiting the proliferation and dissemination of cancer cells. Inotherwords, Lung cancer is best detected at an early stage by the application of several image processing techniques as shown in Figure 1; such as Sputum cytology, MRIs, computed tomographies, and X-rays for the chest [6].

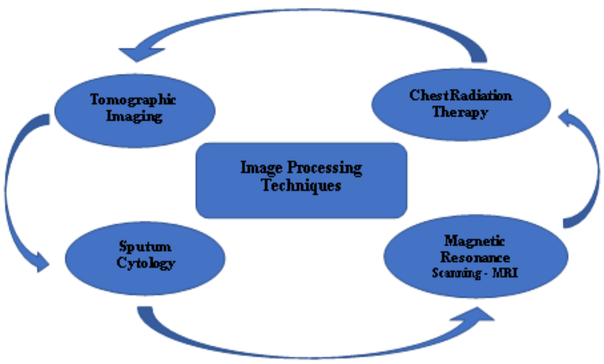

Fig. 1. Image Processing Techniques

Despite the presence of several approaches for detecting and classifying lung cancer, some limitations persist that might elevate the likelihood of treatment failure and incorrect care [8]. The complex structure of lung cancer cells is a challenge for early identification in the first stage, since the cells tend to overlap with one another. One of the neural network models – Back Propagation neural Network (BPNN) can be adopted to ameliorate some of these limitations. BPNN are computational models based on how human brain functions and structured; making use of backpropagation algorithm which is a supervised learning technique that facilitates the acquisition of knowledge by neural networks through the utilization of labeled training data. Backpropagation enables neural networks to acquire intricate patterns and correlations within data, therefore enabling activities including recognizing images.

Meanwhile, BPNN being a powerful algorithm for training neural networks, also comes with its own challenges and limitations such as vanishing and exploding gradients, local Minima and saddle Points, overfitting and choice of hyperparameters. To prevent the BPNN model from these aforementioned challenges, there is need to select optimal parameters with the help of an optimization algorithm / optimization techniques such as chameleon swarm optimization. The Chameleon Swarm Optimization (CSO) uses the Chameleon Swarm Algorithm (CSA) as Nature Inspired Algorithms can improve the performance of BPNN. This algorithm (meta-heuristic algorithm) draws inspiration from the hunting behavior of chameleons and their efficient hunting techniques setting a careful balance between exploration and exploitation, which are both essential components to achieve success [7]. Chameleons systematically explore every possible region within their search zone, utilizing their spherical eyes to survey a broad range.

II. REVIEW OF LITERATURES

A. Epidemiology of Lung Cancer

Global epidemiological patterns indicate the need for specialized criteria for lung cancer screening and algorithms for managing lung nodules, in order to consider the variations in lung cancer biology. The primary objective is to prevent lung cancer, and it is envisaged that domestic efforts to reduce smoking and air pollution will be implemented in both countries [9]. Globally, lung cancer is the most common type of cancer; it also accounts for most cancer-related mortality [5]. The prognosis is quite poor, as 75% of patients obtain severe level diagnosis. Existing medical diagnosis techniques tends to be inadequate for early detection, so it is crucial to develop novel ways for prompt and precise course of therapy. Cancer of the lung arises from a process called multistage carcinogenesis, which involves the accumulation of genetic and epigenetic alterations. Examining distinctive inherited traits can assist in the early detection of a condition. Only 25% of individuals diagnosed with advanced-stage lung cancer have a wide range of therapy choices available to them.

B. Subtypes of Lung Cancer

Non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) are two of the several subtypes of lung cancer. While the general death rate for lung cancer in the United States has decreased, there exist dearth of evidence about mortality trends for specific forms of lung cancer on demographic threshold. This is due to death certificates lacking specific details about subcategories [10].

1) Small-Cell Lung Cancer (SCLC)

Of all the cases of lung cancer, small-cell lung cancer (SCLC) makes about 15%. It is distinguished from other forms of lung cancer by its exceptionally rapid growth rate, pronounced inclination towards promptly disseminated to additional bodily areas, and very poor outlook [11, 12]. Common lung cancer, SCLC, is frequently asymptomatic until it has progressed and frequently spreads to the brain, liver, or bone. SCLC is the most deadly kind, with a 5-year mortality rate of 90% or higher due to its rapid development and early metastasis [13]. This fast spreading, highly metastatic disease [14] is directly associated with tobacco carcinogens.

2) Non-Small-Cell Lung Cancer (NSCLC)

Over 90% of lung cancer cases in the US are categorized into four primary histologic types: squamous cell carcinoma, adenocarcinoma, big cell carcinoma, or small cell carcinoma. The occurrence of NSCLC has decreased among men in North America, northwest Europe, Australia, and New Zealand from the mid-1970s to 1980s; however, the age-adjusted rate is still rising among these nations' female populations as well as among men and women in southern and Eastern Europe. The occurrence rates of different types of lung cancer have undergone considerable changes throughout time, with adenocarcinoma surpassing squamous cell carcinoma as the prevailing form [15, 16].

C. Machine learning with Statistical Inference

Within artificial intelligence, machine learning is a specialised area of study using several statistical, probabilistic, and optimisation methods to acquire knowledge from previous instances. Subsequently, this can be utilized for categorizing unfamiliar data, recognizing unexplored patterns, and forecasting unprecedented developments [17]. One possible inference from this is that machine learning, akin to statistics, can be employed for the analysis and interpretation of data. Machine learning relies on statistical and probability theorems; however, it is more powerful than traditional statistical methods. Meanwhile, the analysis and interpretation of data. Machine learning methods differ from statistical techniques by incorporating absolute conditionality (IF, THEN, ELSE), Boolean logic (AND, OR, NOT), conditional predications (probability of P given Q), and novel optimization algorithms for data modeling or detect variations.

D. CNN Architectures and Applications

Convolutional Neural Networks, are widely used mathematical models that can accurately estimate complex functions. They are particularly effective at identifying and extracting specific characteristics from images, such as obstacles and traffic signs. Facial recognition technology has been utilized to verify the identities of individuals based on physical characteristics of their face, like the span between the eyes, the form that defines their nose, and their lips curve [18]. Convolutional neural networks (CNNs) are powerful neural networks specifically designed for tasks related to image recognition. The predictions are generated by fully connected layers that utilize these features, following the extraction of relevant information from the input image by convolutional and pooling layers. To utilize a Convolutional Neural Network (CNN), it is necessary to train it with a substantial dataset consisting of labeled images. By employing optimization and back propagation techniques, the Convolutional Neural Network (CNN) acquires the ability to establish meaningful connections between extracted features and their corresponding labels. Once trained, the model may choose the label with the highest predicted probability from a collection of new, unseen photos and make predictions based on those images [19]. Throughout history, CNNs have achieved important advances in pattern recognition disciplines including voice and picture / image processing [20].

E. Deep Learning Approach

Deep learning-based models offer a comprehensive learning system capable of simultaneously acquiring feature representation and performing categorization. One accomplished this using multi-layer neural network, which is occasionally referred to as a Deep Neural Network (DNN). This network learns numerous layers of images matching varying degrees of abstraction in order to more effectively reveal the underlying patterns in the data [21]. Meanwhile, deep models have the benefit of learning a hierarchy of ideas in brain tumour and cancer predictions. Convolutional neural networks (CNNs), recurrent-neural-networks (RNNs), long short-term memory networks (LSTM), Auto-encoders, and Generative Adversarial Networks (GANs) were among the several deep learning architectures for image analysis and classifications.

III. METHODOLOGY

The research adopted quantitative approach because of its great importance and prominent adaptability in the domain of Computer Science, Engineering and Statistics. The, software building method which is a quantitative method is being adopted using Python programming language with Matlab R2022a embedded library for model implementation and performance metrics evaluation phase on a total of 893 CT scanned lung images. The dataset is being divided into two - training and testing datasets. 30% of the dataset is allocated for testing and 70% for training using Random Sub-Sampling Cross Validation Method.

The basic methodological structure in each phase regarding this investigation is presented as follows and as depicted in Figure 3.

A. Data Collection

large and diverse dataset of lung tissue images for developing and training the neural network models are obtained. This involve accessing existing data repositories or collecting new data using various imaging techniques. In this research, the first step is to collect a large dataset of digitized images of lung tissue samples obtained from Mendeley Data Publication Online - www.kaggle.com (Figure 2). This dataset includes both cancerous and non-cancerous samples for the purpose of training and evaluating the neural network.

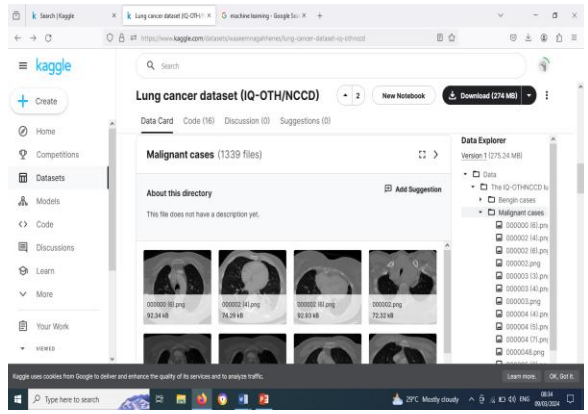

Fig. 2. Lung Cancer Datasets Acquisition

B. Preprocessing

Raw data are pre-processed to ensure that it is standardized and ready for use in the machine learning models. This involves image normalization, resizing, and feature extraction. Inotherword, collected data are pre-processed to remove noise, artifacts, and other factors that may interfere with the neural network precision. It includes resizing, normalization, and other techniques to standardize dataset images.

C. Model selection and training

An appropriate neural network model based on data and research question are selected. Subsequently, the pre-processed data was utilized for training purposes. This entails optimising the model's parameters to enhance its effectiveness on the data used for training.

D. Model evaluation

The trained model must undergo evaluation employing an independent dataset for validation to determine its accuracy, precision, recall, and other performance parameters. This might include employing cross-validation techniques to guarantee that model’s performance remains resilient and transferable. Procedures of training neural network using the pre-processed data include choosing the suitable neural network architecture, configuring hyperparameters, and tuning the network to enhance its accuracy.

E. Interpretation and validation

This is the final step involves interpreting the results of the model and validating them against clinical standards. It involves comparing the model's performance to that of human pathologists, and assessing its usefulness in real-world. Inotherword after training the neural network, validation is required to evaluate its performance. This entails evaluating the neural network using a distinct dataset that was not utilized during the training process, and then comparing the results with the validated clinical data.

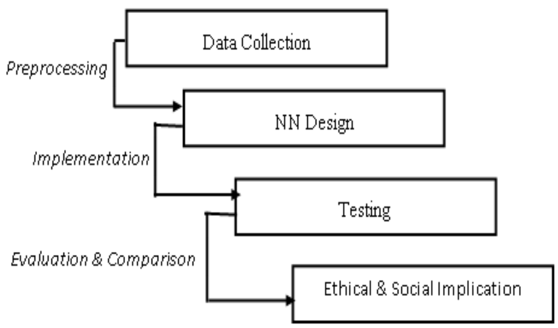

Fig. 3. Research Paradigm

To apply the Chameleon Swarm Optimization (CSO) technique for BPNN Parameter Selection, this investigation must initially comprehend the problem the machine trying proffering solution to. The key issue in this scenario is BPNN parameter selection. The Backpropagation Neural Network (BPNN) is a commonly utilized machine learning technique to perform both regression as well as classification tasks. It requires identification of suitable criteria such as learning rate, penalty parameter, and gamma parameter. The purpose of using CSO technique is to figure out the most ideal outcomes for the parameters of BPNN (Figure 4). There is a need to define the optimization problem. The purpose of optimization evaluation in this scenario is to determine the most favorable settings for the BPNN parameters that result in the lowest possible classification error. This study will formulate a mathematical optimization problem that can be resolved with an optimization procedure.

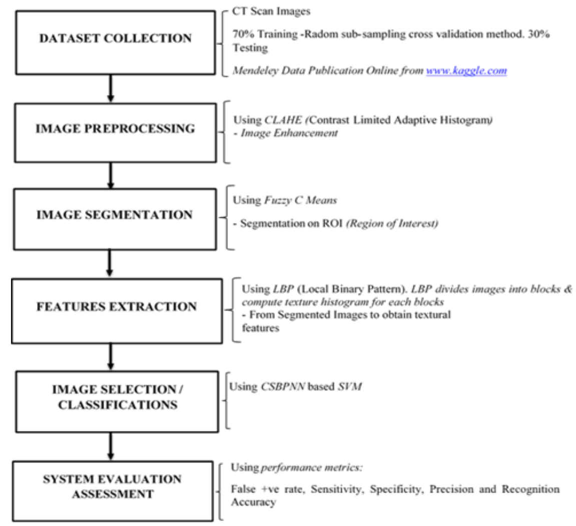

Fig. 4. Block Diagram Structure of the Developed Technique

IV. RESULTS EVALUATION

The acquired lung datasets for the simulation were annotated with labels indicating the presence of Malignant, Benign, and Normal/All; with the composition of 200 Benign, 393 of Malignant and 300 Normal/All lung images.

A. Model Simulation

To offer a comprehensive perspective on the classification, confusion matrices were adopted using different threshold values for both BPNN and CS-BPNN simulation (Figure 5). The threshold utilized in this research ranges from 0-0.25, 0.26-0.36, 0.37-0.50 and 0.51-0.85, respectively. However, the performance of the technique was affected by the threshold value and best results were obtained at a threshold value of 0.85 for every technique.

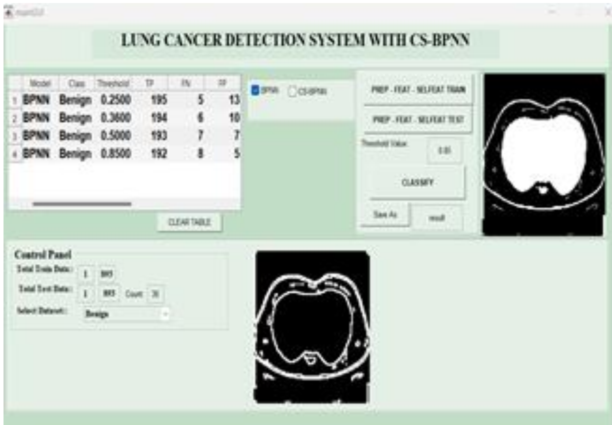

Fig. 5. Model Simulation Phase

The values obtained from each of the models (BPNN and CSBPNN) using evaluation metrics (False Positive Rate (FPR),

TABLE II: BPNN MODEL EVALUATION

Specificity (SPEC), Sensitivity (SEN), Accuracy (ACC) and recognition Time are tabulated below (Table II and Table III) respectively.

TABLE III: CSBPNN MODEL EVALUATION

Class	Threshold	TP	FN	FP	TN	FPR (%)	SPEC (%)	SEN (%)	ACC (%)	Time (Sec)
Normal/All	0 - 0.25	550	43	51	249	17.00	83.00	92.75	89.47	177.57
Benign		195	5	13	287	4.33	95.67	97.50	96.40	60.00
Malignant		382	11	19	281	6.33	93.67	97.20	95.67	59.89
Normal/All	0.26 - 0.36	548	45	45	255	15.00	85.00	92.41	89.92	177.27
Benign		193	7	7	293	2.33	97.67	96.50	97.20	59.64
Malignant		380	13	13	287	4.33	95.67	96.69	96.25	60.25
Normal/All	0.37 - 0.5	547	46	43	257	14.33	85.67	92.24	90.03	177.27
Benign		192	8	5	295	1.67	98.33	96.00	97.40	59.92
Malignant		379	14	11	289	3.67	96.33	96.44	96.39	60.15
Normal/All	0.51- 0.85	545	48	41	259	13.67	86.33	91.91	90.03	178.25
Benign		190	10	3	297	1.00	99.00	95.00	97.40	59.92
Malignant		375	18	7	293	2.33	97.67	95.42	96.39	59.89

B. Performance Assessment of both BPNN and CSBPNNN

The analytical performance evaluation for both BPNN and CS-BPNN at the optimal performance threshold of 0.85 across all evaluation metrics is summarily indicated below (Table 4) and graphically represented as shown in Figure 6, Figure 7 and Figure 8 respectively for the Normal, Benign and Malignant dataset using confusion evaluation matrix parameters.

TABLE IV: BPNN AND CS-BPNN RESULTS ON NORMAL, BENIGN AND MALIGNANT DATASETS.

Class	Threshold	TP	FN	FP	TN	FPR (%)	SPEC (%)	SEN (%)	ACC (%)	Time (Sec)
Normal/All	0 - 0.25	570	23	31	269	10.33	89.67	96.12	93.95	116.12
Benign		197	3	11	289	3.67	96.33	98.50	97.20	40.43
Malignant		386	7	15	285	5.00	95.00	98.22	96.83	39.94
Normal/All	0.26 - 0.36	568	25	25	275	8.33	91.67	95.78	94.40	120.03
Benign		195	5	5	295	1.67	98.33	97.50	98.00	40.40

Malignant		384	9	9	291	3.00	97.00	97.71	97.40	39.93
Normal/All	0.37 - 0.5	567	26	23	277	7.67	92.33	95.62	94.51	118.70
Benign		194	6	3	297	1.00	99.00	97.00	98.20	39.84
Malignant		383	10	7	293	2.33	97.67	97.46	97.55	40.00
Normal/All	0.51 - 0.85	565	28	21	279	7.00	93.00	95.28	94.51	117.67
Benign		192	8	1	299	0.33	99.67	96.00	98.20	39.86
Malignant		381	12	5	295	1.67	98.33	96.95	97.55	39.87

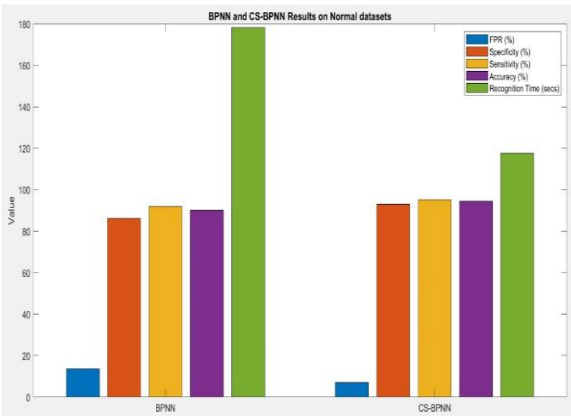

Fig. 6. BPNN and CS-BPNN Representation for Normal Dataset Metrics

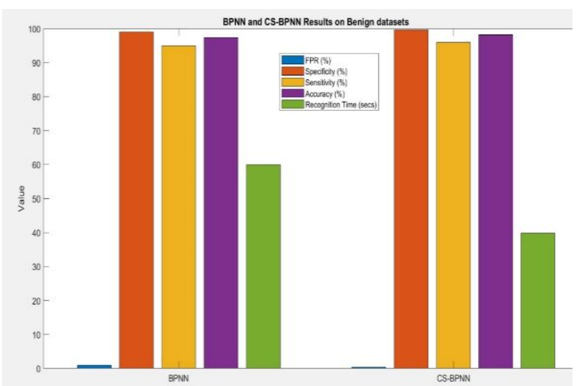

Fig. 7. BPNN and CS-BPNN Representation for Benign Dataset Metrics

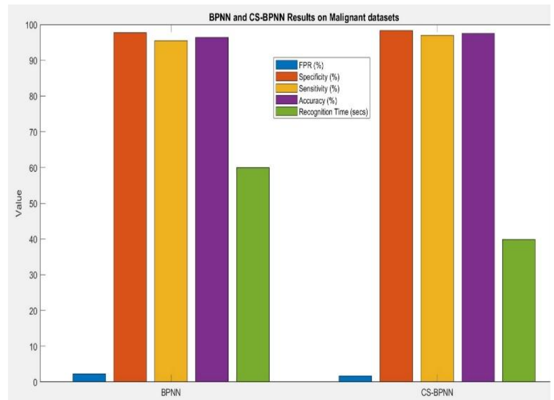

Fig. 8. BPNN and CS-BPNN Representation for Malignant Dataset Metrics

The above metrics evaluation indicates that CS-BPNN technique outperformed the BPNN technique in the detection and categorization of cancer of the lung illnesses. Optimal threshold for the CS-BPNN is responsible for the technique's enhanced performance over BPNN regarding FPR, Specificity, Sensitivity, Accuracy and response Time. Given the aforementioned outcome, all of the study's datasets had increased accuracy, specificity, sensitivity, and reduced FPR as a result of the combination of the BPNN and CS techniques.

V. CONCLUSION

This research has added to the body of knowledge by developing an optimized Backpropagation Neural Network (BPNN) model using chameleon swarm optimization for parameter selection. By comparative analysis and statistical evaluation of integrating CSO with BPNN through Matlab2022a using the adopted performance metrics, has established the capability of CSO in improving the precision, sensitivity and specificity of the classification model; as well as reduction in both false positive rate and recognition time. The hereby provides insight into the effectiveness of CSO-based BPNN optimization algorithm. This developed technique led to the development of an improved algorithms for the purpose of recognising and categorising lung cancer; ultimately enhancing identification and management of the disease. This research work will assist medical practitioners in comprehending the model's rationale and facilitating informed decisions about the cancer patient's treatment.

ACKNOWLEDGEMENT

The authors wish to thank the Federal Ministry of Education, Nigeria, through the TETFUND Institution Based Research (IBR) Intervention funding for sponsorship.

REFERENCES

[1] A. Asuntha, A. Brindha, S. Idirani, & A. Srinivasan. “Lung Cancer Detection using SVM Algorithm and Optimization Techniques”. Journal of Chemical and Pharmaceutical Sciences, 9(4), 2016, 3198-3203.

[2] J. Ji, W. Zhang, Y. Dong, R. Lin, Y. Geng, L. Hong. “Automated Lung Cancer Detection using Histopathological Images”. Researchsquare 2023, 1–13.

[3] S. K. Thakur, D. P. Singh & J. Choudhary. “Lung cancer identification: a review on detection and classification”. Cancer and Metastasis Reviews - Springer, 39: 2020, 989–998.

[4] A. Bhandary, Prabhu, G. A., M. Basthikodi & K.M. Chaitra. “Early Diagnosis of Lung cancer Using Computer Aided Detection via Lung Segmentation Approach”. International Journal of Engineering Trends and Technology, 69(5), 2021, 85-93.

[5] A. Bhandary, & M. Basthikodi. “Early Diagnosis of Lung Cancer Using Computer Aided Detection via Lung Segmentation Approach”. arXiv preprint arXiv:2107.12205, 2021.

[6] R.M. Mohan, R. D. H. Devi, & A. Bai. “Lung cancer detection using Nearest Neigbour classifier”. International Journal of Recent Technology and Engineering (IJRTE). 8(2S11), 2019, pp. 3641–3645.

[7] R. R. Asaad, & R.I. Ali. “Back Propagation Neural Network (BPNN) and sigmoid activation function in multi-layer networks”. Academic Journal of Nawroz University, 8(4), 2019, 216-221.

[8] A. Srinivasulu, K. Ramanjaneyulu, R. Neelavenin, S.R. Karanam, S. Majji, M. Jothilingam, T.M. Patnala. “Advanced lung cancer prediction based on blockchain material using extended CNN". Applied Nanoscience, 13 (1), 2021, 1-13.

[9] D. Yang, Y. Liu, C. Bai, X. Wang, & C. A. Powell. “Epidemiology of lung cancer and lung cancer screening program in China and the United States”. Cancer Letters, vol. 468, 2020, pp. 82-87.

[10] N. Howlader, G. Forjaz, M. J. Mooradian, R. Meza, C.Y Kong, K. A. Cronin, A. B. Mariotto, D. R. Lowy, & E. J. Feuer. "The effect of advances in lung-cancer treatment on population mortality." New England Journal of Medicine, 383(7), 2020, 640-649.

[11] F. Calabrese, F. Lunardi, F. Pezzuto, F. Fortarezza, S.E.Vuljan, C. Marquette, & P. Hofman. “Are There New Biomarkers in Tissue and Liquid Biopsies for the Early Detection of Non-Small Cell Lung Cancer?”. Journal of clinical medicine, 8(3), 2019, pp. 414.

[12] L. H. Araujo, L. Horn, R. E. Merritt, K. Shilo, M. XuWelliver, & D. P. Carbone. “Cancer of the Lung: Non-Small Cell Lung Cancer and Small Cell Lung Cancer”. Abeloff’s Clinical Oncology, 2020, 1108-1158.

[13] Y. Q. Song, N. Wang, Y. Qiao, L. He, X. Li, X. F. Zhang, Q. K. Yang, R. Z. Wang, R. He, C. Y. Wang, Y. W. Ren, G. Li, & T. L. Wang “Treatment patterns and survival after18F-fluorodeoxyglucose positron emission tomography/computed tomography-guided local consolidation therapy for oligometastatic non-small cell lung cancer: a two-center propensity score-matched analysis”. Journal of Cancer Research and Clinical Oncology, 146(4), 2020, 1021–1031.

[14] Z. Megyesfalvi, C. M. Gay, H. Popper, R. Pirker, G. Ostoros, S. Heeke, C. Lang, K. Hoetzenecker, A. Schwendenwein, K. Boettiger,

P. A. Bunn, F. R. Vamos, K. Schelch, H. Prosch. "Clinical insights into small cell lung cancer:Tumor heterogeneity, diagnosis, therapy, and future directions". CA: a cancer Journal for clinicians, 73(6), 2023, 620-652.

[15] J. Kim, H. Lee, & B. W. Huang. "Lung cancer: diagnosis, treatment principles, and screening". American family physician, 105(5), 2022, 487-494.

[16] S. K. Vinod, & E. Hau. "Radiotherapy treatment for lung cancer: Current status and future directions". Respirology, 25(S2), 2020 61–71.

[17] G. Sharma, & C. Prabha. "Applications of machine learning in cancer predictionand prognosis". Cancer Prediction for Industrial IoT 4.0, 2021, 119-135.

[18] A. Zafar, M. Aamir, N. Mohd Nawi, A. Arshad, S. Riaz, A Alruban, A. K. Dutta, & S. Almotairi. "A Comparison of Pooling Methods for Convolutional Neural Networks". Applied Sciences (Switzerland), 12(17), 2022, 1–21.

[19] M. Krichen. "Convolutional Neural Networks: A Survey". Computers, 12(8), 2023, pp.151.

[20] M. M. Taye. "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions". Computation, 11(3), 2023, pp.52.

[21] S. Minaee, & A. Abdolrashidi. "DeepIris: iris recognition using a deep learning approach". arXiv preprint arXiv:1907.09380. 2019.