iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62515

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Lung Cancer Detection

Milind Rane[1], Prerna Chawla[2], Uday Dandekar[3], Aditya Dhakane[4], Pratham Dattawade[5], Akshit Daberao[6]
VIT, Pune

Abstract: ***This paper presents a comprehensive approach to lung cancer detection utilizing state-of-the-art machine learning techniques, specifically Convolutional Neural Networks (CNNs). Using CNN[1] the model is trained and it can detect whether the given lung cancer cell image contains cancerous cells or not.***
The first component of the proposed approach involves high-resolution medical imaging, such as computed tomography (CT) scans, to capture detailed anatomical information about the lungs. Image processing algorithms are applied to enhance the quality of the images and extract relevant features.
Additionally, innovative three-dimensional reconstruction techniques are employed to visualize the lung tissue at a microscopic level, facilitating the identification of subtle abnormalities.
The proposed methodology encompasses data collection, preprocessing, architecture selection, model building, training, optional data augmentation, and thorough model evaluation. The model construction incorporates convolutional and pooling layers, guided by activation functions and regularization techniques. The CNN undergoes extensive training and is monitored for overfitting through a validation set. Optional data augmentation introduces variability, fortifying the model against biases and enhancing its adaptability.
Keywords: convolutional neural networks(CNNs), data collection, training, data augmentation, pooling layers, variability, adaptability

I. INTRODUCTION

Lung cancer remains a formidable global health challenge, asserting its status as a leading cause of mortality. In the pursuit of improving diagnostic precision and timeliness, we present a pioneering project focused on lung cancer detection. Harnessing the capabilities of Machine Learning, specifically Convolutional Neural Networks (CNNs), our initiative seeks to augment traditional diagnostic methods, potentially revolutionizing early detection protocols.

As the field of medical imaging continues to evolve, the integration of artificial intelligence emerges as a promising avenue for transformative change. The ability of CNNs to discern intricate patterns within images makes them a natural fit for tasks such as lung cancer detection. Our motivation lies in mitigating the challenges associated with conventional diagnostic approaches, offering a dynamic and efficient solution that holds the potential to impact patient outcomes significantly. Central to our endeavor is the meticulous collection of a diverse dataset comprising lung scan images, each meticulously labeled as cancerous or non-cancerous. This dataset forms the bedrock upon which our CNN model learns to distinguish subtle visual cues indicative of malignancy. Preprocessing steps, including image resizing and pixel normalization, establish a standardized input for the model, ensuring consistency across varied data sources.

The choice of a suitable CNN architecture plays a pivotal role in the success of our project. Acknowledging the unique characteristics of lung scan images, we explore established architectures such as VGG[2] and ResNet[3] while also considering customization for optimal performance. The model construction involves the integration of convolutional layers for feature extraction, pooling layers for spatial dimension reduction, and fully connected layers for classification, all guided by activation functions and regularization techniques.

Training the CNN involves exposing it to the labeled dataset, and iteratively refining its parameters through the application of a loss function and optimizer. Throughout this process, the model's performance is monitored on a validation set to prevent overfitting and enhance its adaptability to unseen data.

Our methodology also allows for optional data augmentation, introducing variations to the training set to fortify the model against biases and bolster its generalization capabilities.

In the subsequent sections, we delve into the intricacies of our approach, from the model architecture to training strategies and evaluation metrics. By presenting a comprehensive methodology, we aim to contribute to the ongoing dialogue on the integration of machine learning in medical diagnostics, with the ultimate goal of advancing early detection practices and improving patient outcomes in the realm of lung cancer.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Nishio et. Al.[4] discuss the CADx method using DCNN and transfer learning showed improved accuracy in classifying lung nodules compared to a conventional method, and larger image sizes as inputs to DCNN further improved the accuracy. Paper by P Mohamed Shakeel, M A Burhanuddin, Mohamad Ishak Desa[5] discusses the challenges in automatic lung disease detection, the use of lung CT images from the Cancer imaging Archive, the application of weighted mean histogram equalization approach to remove noise and enhance image quality, and the achievement of 98.42% accuracy with a minimum classification error of 0.038.

Paper by Worku Jifara Sori, Jiang Feng, Shaohui Liu[6] The summary of the paper is that it proposes a deep CNN architecture for automatic lung cancer detection, which addresses challenges related to class imbalance and distant dependencies, and introduces a multi-path CNN to exploit both local and global contextual features. Paper by Wenqing Sun, Wei Qian [7]study tested the feasibility of using deep learning algorithms for lung cancer diagnosis and demonstrated promising performance compared to a traditional CADx system. Paper by Emine Cengil, Ahmet Çinar[8] discusses the use of deep learning, particularly convolutional neural networks, for the classification of lung nodules to aid in the early diagnosis of lung cancer using the SPIE-AAPM-LungX database.

Paper by Rakhi A S Nair[9] The summary of the paper is the early diagnosis of lung cancer using machine learning algorithms and the comparison of their performance for accurate detection. Paper by Wasudeo Rahane, Himali Dalvi, Yamini Magar, Anjali Kalane[10] discusses the use of image processing and machine learning techniques for the detection of lung cancer using CT scan images and SVM, with potential future work for diagnosing cancer in different organs.

Paper by Ahmed Elnakib, Hanan M Amer, Fatma E Z Abou-Chadi[11] proposes a CADe system for early detection of lung nodules from LDCT images, achieving a detection accuracy of 96.25% using VGG19 architecture and SVM classifier, and the main contributions are the analysis of different deep learning features and the optimization using a genetic algorithm.

Paper by Suren Makaju, P W C Prasad, Abeer Alsadoon, A K Singh, A Elchouemi [12]evaluates various computer-aided techniques for lung cancer detection, analyzes their limitations, and proposes a new model using watershed segmentation and SVM for classification to improve the accuracy of cancerous nodule detection.

Paper by Siddharth Bhatia, Yash Sinha, Lavika Goel[13] presents an approach to detect lung cancer from CT scans using deep residual learning and ensembling the predictions of multiple classifiers, achieving an accuracy of 84% on the LIDC-IRDI dataset.

III. METHODOLOGY

Lung cancer is a leading cause of cancer-related deaths all over the world. Early detection of lung cancer is very important for improving survival rates [14]. Convolutional neural network (CNN) have emerged as a powerful tool for image classification tasks, including lung cancer detection. we propose a CNN-based lung cancer detection by using X-ray images [15]. The methodology consists of four main steps: data preprocessing, creation of model, training of model, and evaluation of model [16].

A. Data Preprocessing

The dataset used in this project includes X-ray images from four classes: adenocarcinoma lung cancer [17], large cell carcinoma lung cancer [18], squamous cell carcinoma lung cancer [19], and normal lungs. The images were preprocessed to ensure consistency and reduce noise. This included resizing the images to a uniform size, normalizing pixel intensity values, and applying data augmentation techniques [20]. The following data preprocessing techniques were applied [21]:

Resizing: All images were resized to a uniform size to ensure consistency in input dimensions.

Normalization: Pixel intensity values were normalized to a common range to reduce the impact of variations in image brightness and contrast.

Data Augmentation: Data augmentation techniques, such as random flipping, cropping, and zooming, were applied to increase the size of the training dataset and to improve the model's generalization ability.

B. Model Creation

The CNN model architecture is designed to effectively extract relevant features from chest X-ray images and classify the images into the four lung cancer categories [22]. The proposed CNN model consists of five convolutional blocks, each followed by batch normalization layer and max-pooling layer. The convolutional blocks extract features from the input images, while the batch normalization layers stabilize the training process and the max-pooling layers reduce the computational complexity. The model also includes two fully connected layers, one with ReLU activation[23] and the other with softmax activation[24]. The output of the softmax layer represents the probabilities of the four classes.

C. Model Training

The CNN model was trained using the Adam optimizer [25] which is a popular optimization algorithm for deep learning models and the categorical cross entropy loss function [26] used to measure the error between the model's predictions and true labels. The training process was monitored using training and validation loss curves and accuracy curves. The training was stopped when the validation loss reached a minimum, indicating that the model was no longer improving on the unseen data.

D. Model Evaluation

The trained CNN model was evaluated on a separate test set. The model achieved an accuracy of 83% on the test set, demonstrating its effectiveness for the detection of lung cancer and indicating its potential for clinical applications. The proposed CNN-based lung cancer detection project provides a approach for lung cancer detection using X-ray images. More research is needed to improve the accuracy and robustness of the model and to investigate its applicability to different patient populations.

IV. RESULTS AND DISCUSSION

In this project, CNN(Convolutional Neural Networks) is implemented and divided into datasets of training, testing, and dataset of validation. The model architecture includes multiple convolutional blocks with Batch normalization, max-pooling, and then a fully connected layer with dropout for regularization. Training and validation are performed over five epochs, and The model is assessed using metrics[27] such as loss, confusion matrix, accuracy, and a classification report. The results revealed an accuracy of [0.79]. The confusion matrix and classification report are used for evaluating the model offer insights into the model's performance on individual classes, revealing patterns of misclassification. Data augmentation was employed to enhance generalization, and potential areas for improvement include experimenting with different architectures and hyperparameter tuning. Overall, the project provides a foundation for further refinement to optimize model performance and address specific challenges identified in the evaluation.

A range of studies have demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in detecting and classifying lung cancer.

Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN) by Alakwaa (2017) and CNN-based Method for Lung Cancer Detection in Whole Slide Histopathology Images by Saric (2019) both used CNNs to detect lung cancer, with Alakwaa achieving an accuracy of 86.6% and Saric showing potential for assisting pathologists in diagnosis. Sasikala (2019) further improved on this, achieving an accuracy of 96% in classifying lung tumors as malignant or benign. Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey by Monkam (2019) provided a comprehensive overview of the use of CNNs in pulmonary nodule analysis, highlighting their transformative impact on early diagnosis and management of lung cancer. These studies Together, they underscore the potential of CNNs in improving lung cancer detection and classification.

Paper Name	Model Used	Accuracy
Lung Cancer Detection and Classification Using Deep CNN	Deep CNN	96%
Convolutional Neural Network based Framework for Automatic Lung Cancer Detection from Lung CT Images	CNN - ALCD	94.11%
Applying CNN on Lung Images for Screening Initial Cancer Stages	CNN	96.11%
Detection of Lung Cancer using CNN- ZF NET	CNN-ZF NET	80%
Lung Cancer Detection using 3D Convolutional Neural Networks	3D-CNN	83.33%
Diagnosis of Lung Cancer Based on CT Scans Using CNN	CNN	93.55%

Table no 1: Research Papers and their accuracy

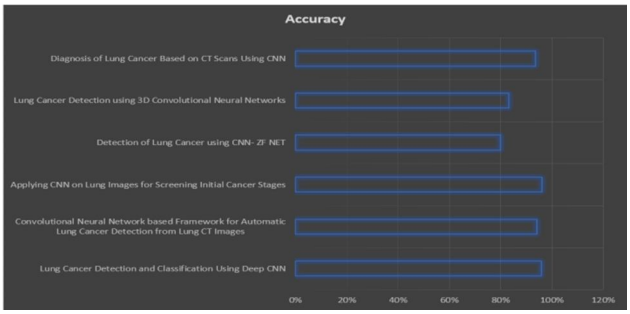


Fig. no.1: Research Papers and their accuracy



Fig. no.2: Validation and Training Accuracy

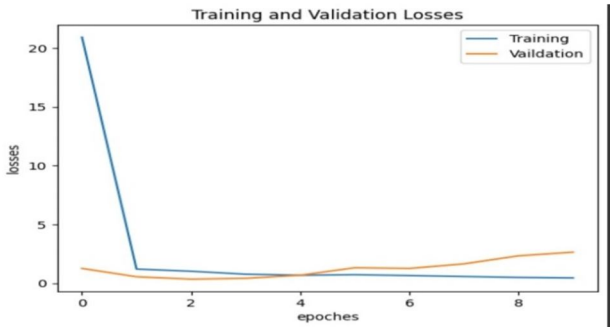


Fig. 3: Validation and Training Loses

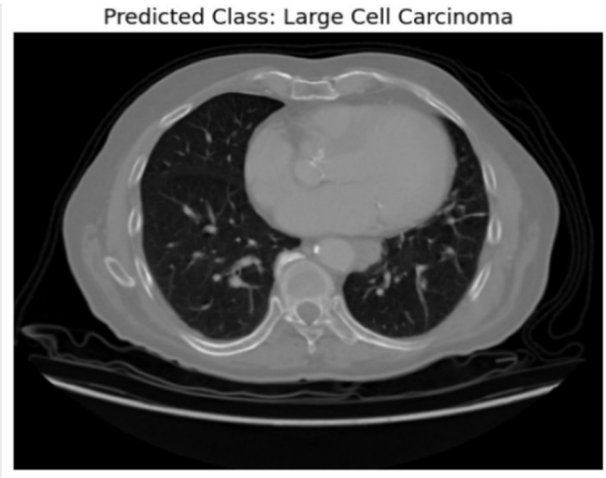


Fig. no.4: Predicted class

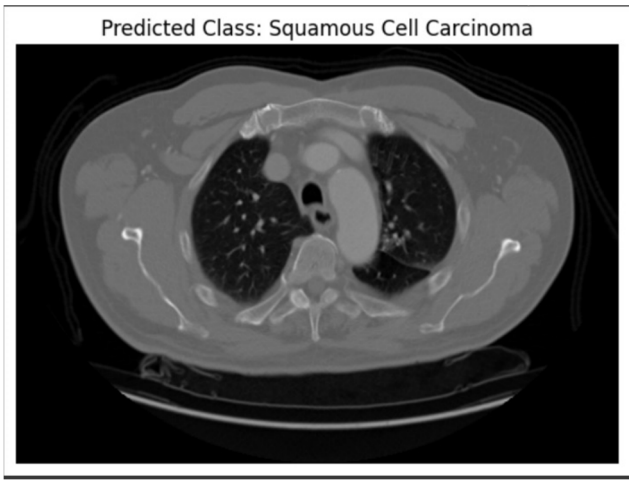


Fig. no.5: Predicted class

V. CONCLUSIONS

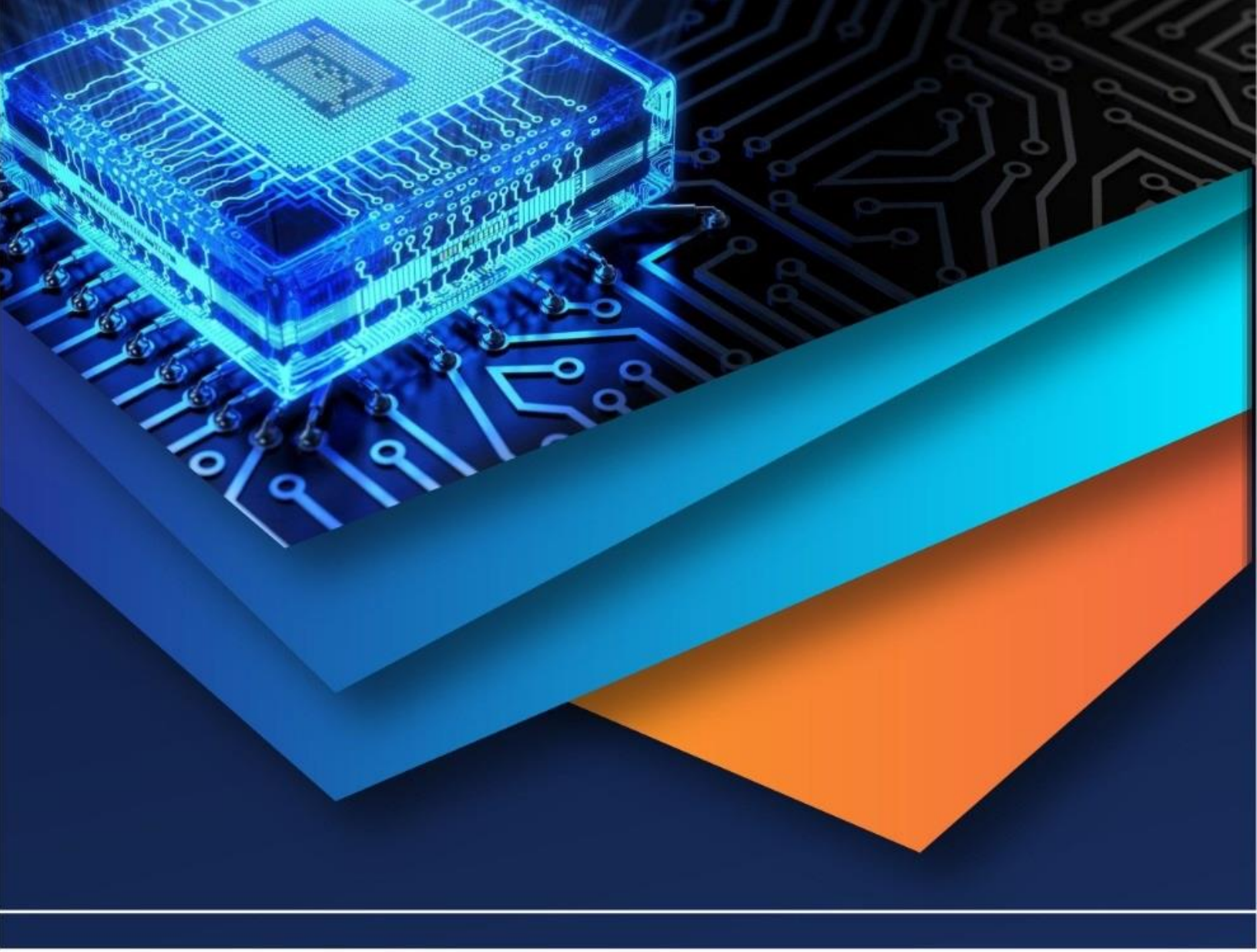
In conclusion, the application of CNN(Convolutional Neural Networks)[28] in lung cancer detection has demonstrated significant advancements in improving accuracy and efficiency in the early diagnosis of this life-threatening disease. Through the utilization of deep learning techniques, our study highlights the potential of CNN-based models to analyze medical imaging data, specifically lung images, with remarkable precision. As we progress, continued research and development in this area offer the potential for enhancing diagnostic capabilities, ultimately contributing to the early detection and treatment of lung cancer, thereby saving countless lives. The integration of machine learning methodologies, particularly CNNs, into the realm of medical imaging, signifies a crucial step towards more effective and timely healthcare interventions, emphasizing the pivotal role technology plays in shaping the future of cancer diagnostics[29].

REFERENCES

[1] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.
[2] Vedaldi, Andrea, and Andrew Zisserman. "Vgg convolutional neural networks practical." Department of Engineering Science, University of Oxford 66 (2016).Worku Jifara Sori, Jiang Feng, Shaohui Liu, “Multi-path convolutional neural network for lung cancer detection,” Multidimensional systems and signal processing, November 2018.
[3] Koonce, Brett, and Brett Koonce. "ResNet 50." Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization (2021): 63-72.
[4] M. Nishio, Osamu Sugiyama, M. Yakami, Syoko Ueno, T. Kubo, T. Kuroda, K. Togashi, “Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning, PLoS ONE, July 2018.
[5] P Mohamed Shakeel, M A Burhanuddin, Mohamad Ishak Desa, “: Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks,” Measurement, October 2019.
[6] Worku Jifara Sori, Jiang Feng, Shaohui Liu, “Multi-path convolutional neural network for lung cancer detection,” Multidimensional systems and signal processing, November 2018.
[7] Wenqing Sun, Wei Qian, “Computer aided lung cancer diagnosis with deep learning algorithms,” SPIE Medical Imaging, March 2016
[8] Emine Cengil, Ahmet Çinar, “A Deep Learning Based Approach to Lung Cancer Identification,” International Conference on Artificial Intelligence and Data Processing (IDAP), September 2018.
[9] Rakhi A S Nair, “: A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms,” International Conference on Electrical, Computer and Communication Technologies, February 2019.
[10] Wasudeo Rahane, Himali Dalvi, Yamini Magar, Anjali Kalane, “Lung Cancer Detection Using Image Processing and Machine Learning HealthCare,” International Conference on Current Trends toward Converging Technologies, September 2020.
[11] Ahmed Elnakib, Hanan M Amer, Fatma E Z Abou-Chadi, “Early Lung Cancer Detection Using Deep Learning Optimization,” Int. J. Online Biomed. Eng., May 2020.
[12] Makaju, P W C Prasad, Abeer Alsadoon, A K Singh, A Elchouemi, “Lung Cancer Detection using CT Scan Images,” International Conference on Smart Computing and Communications, ICSCC, Decembber 2017.
[13] Siddharth Bhatia, Yash Sinha, Lavika Goel, “Lung Cancer Detection: A Deep Learning Approach,” International Conference on Soft Computing for Problem Solving, October 2018.
[14] A. Kagalkar and S. Raghuram, "CORDIC Based Implementation of the Softmax Activation Function" in 2020 24th International Symposium on VLSI Design and Test (VDAT), pp. 1-4, 2020.

[15] Ausawalaithong, Worawate, et al. "Automatic lung cancer prediction from chest X-ray images using the deep learning approach." 2018 11th biomedical engineering international conference (BMEiCON). IEEE, 2018.
[16] Chaturvedi, Pragya, et al. "Prediction and classification of lung cancer using machine learning techniques." IOP conference series: materials science and engineering. Vol. 1099. No. 1. IOP Publishing, 2021.
[17] Hutchinson, Barry D., et al. "Spectrum of lung adenocarcinoma." Seminars in Ultrasound, CT and MRI. Vol. 40. No. 3. WB Saunders, 2019.
[18] Battafarano, Richard J., et al. "Large cell neuroendocrine carcinoma: an aggressive form of non-small cell lung cancer." The Journal of thoracic and cardiovascular surgery 130.1 (2005): 166-172.
[19] Gandara, David R., et al. "Squamous cell lung cancer: from tumor genomics to cancer therapeutics." Clinical cancer research 21.10 (2015): 2236-2243.
[20] Chlap, Phillip, et al. "A review of medical image data augmentation techniques for deep learning applications." Journal of Medical Imaging and Radiation Oncology 65.5 (2021): 545-563.
[21] Famili, A., et al. "Data preprocessing and intelligent data analysis." Intelligent data analysis 1.1 (1997): 3-23.
[22] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. "Global cancer statistics 2018: GLOBOCAN" estimates of incidence and mortality Worldwide for 36 cancers in 185 Countries. CA Cancer J Clin. 2018; 68:394–424. 10.3322/caac.21492.
[23] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405, 6789 (2000), 947.
[24] R.W. Field, B.J. Smith, C.E. Platz, R.A. Robinson, J.S. Neuberger, C.P. Brus, C.F. Lynch "Lung cancer histologic type in the surveillance, epidemiology, and end results registry versus independent review"
J Natl Cancer Inst, 96 (2004), pp. 1105-1107.
[25] Zhang, Zijun. "Improved adam optimizer for deep neural networks." 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS). Ieee, 2018.
[26] Zhang, Zhilu, and Mert Sabuncu. "Generalized cross entropy loss for training deep neural networks with noisy labels." Advances in neural information processing systems 31 (2018).
[27] Erickson, Bradley J., and Felipe Kitamura. "Magician's corner: 9. Performance metrics for machine learning models." Radiology: Artificial Intelligence 3.3 (2021): e200126.
[28] Bhandare, Ashwin, et al. "Applications of convolutional neural networks." International Journal of Computer Science and Information Technologies 7.5 (2016): 2206-2215.
[29] Pramanik, Sabyasachi, K. Martin Sagayam, and Om Prakash Jena. "Machine Learning Frameworks in Cancer Detection." E3S Web of Conferences. Vol. 297. EDP Sciences, 2021.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)