International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2024): 6.623

Deep Learning-Based Classification of Lung Cancer in CT Images Using Transfer Learning and Random Forest

Timothy Olukunle Olaniyi

School of Computing, Robert Gordon University, Aberdeen
Email: *timolsoft[at]gmail.com*

Abstract: ***Lung cancer is the leading cause of cancer-related deaths globally, with late diagnoses contributing to poor survival rates. Early diagnosis and accurate detection of the lung cancer stage can save the patients' lives. However, as early detection is vital, it is often hindered by non-specific symptoms and limitations due to traditional diagnostic methods. Several image processing, biomarker based, and machine automation approaches are used to identify lung cancer, but accuracy and early diagnosis are challenging for medical practitioners. This study presents a compelling approach to early lung cancer detection using CT scan images, leveraging deep learning for feature extraction and machine learning for classification. By implementing pretrained models like VGG16 and DenseNet121 alongside a Random Forest classifier, the research aims to improve diagnostic accuracy while minimizing computational cost. It is evident that the VGG16 + RF combination outperformed DenseNet121 in both 70/30 and 80/20 training-test splits, achieving up to 95.9% accuracy. This suggests that hybrid architecture can play a vital role in resource-constrained settings, such as rural healthcare. However, limitations due to dataset homogeneity must be acknowledged, with future work proposed around diverse datasets and explainable AI to enhance generalizability and clinical trust.***

Keywords: Lung Cancer Detection, CT Scan Classification, Transfer Learning, Random Forest Classifier, Deep Learning

1. Introduction

Lung Cancer is one of the most common and serious types of cancer with over 43,000 people diagnosed with the condition annually in the United Kingdom [1]. Globally, it is the leading cause of cancer related death worldwide. In 2020 alone, over 2.2 million persons were estimated to have been diagnosed of lung cancers globally and estimated 1.8 million deaths recorded in the same year, this makes lung cancer a serious public health concern [2]. The prevalence of lung cancer has significant impact on the whole global economy. The impact is not limited to the healthcare sector alone, but other factors such as premature death, cost of medication, time off work, unpaid care by friends and family of a death patient and so on. In research published by Oxford University, opined that the cost of cancer to the United Kingdom economy is as high as 15.8 billion annually and lung cancer has a higher cost than other which is almost 3 billion pounds as at 2009 [3]. This economic burden underscores the urgency of improved detection and treatment strategies.

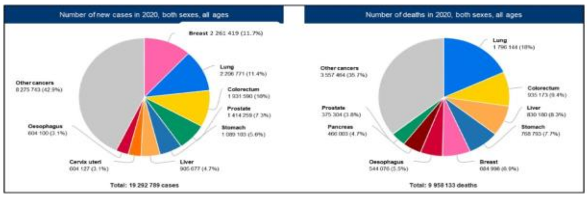


Figure 1: Chart showing Case notification and Death by Lung Cancer

Moreover, lung cancer is a leading prevalent type of cancer globally with high death rate [2]. Therefore, early and accurate diagnosis becomes critical to improving patient outcome. The survival rate of lung cancer is higher if it's diagnosed as early as possible, however, the early detection is not an easy task [4]. There are usually no signs or symptoms of lung cancer in its early stage, however, people with lung cancers develop symptoms such as persistent cough, coughing up blood, persistent breathlessness, unexplained tiredness and weight loss, an ache or pains when breathing [1]. However, these symptoms are not exclusive to lung cancer nor all of them are apparent in every patient, hence the need for thorough diagnostic procedure to determine if the patient is having a lung cancer [5]. Notably, lung cancers are mainly two kinds: No-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). Other types of cancer along the lungs include lymphomas, sarcomas and pleural mesothelioma [7]. Non-Small Cell Lung Cancer is most common and account for 80% of reported cases; the common type of NSCLC include; adenocar cinoma and squamous cell carcinoma, while adenosuamous carcinoma and sarcomoid are not too common types [7]. Besides, Small Cell Lung Cancer accounts for about 20% of the reported cases but grows more quickly and very difficult to treat. It is often found as a relatively small lung tumor that is already spread to the body. Its variant include oat cell carcinoma and combined cell carcinoma [7] Lung cancer is often diagnosed with a combination of imaging procedures, including tissue biopsies to analyse the cancer cells under a microscope and computed tomography (CT), positron emission tomography (PET), or chest X-rays [7] [1]. Beyond imaging, molecular diagnostics also play an important role. To locate specific variations in genetics or biomarkers that can inform treatment choices, molecular testing may also be carried out. The type and stage of the disease, as well as the patient's general state of health are factors that determine the types of treatment. Surgery to remove the tumor and any nearby lymph nodes, radiation therapy to target and kill cancer cells, chemotherapy to kill cancer cells throughout the body, targeted therapy to stop particular molecules in volved in cancer growth, and immunotherapy to boost the immune system are common forms of treatment system to fight against cancer cells [7].

The prognosis for lung cancer varies depending on factors such as the stage of the disease at diagnosis, the type of lung cancer, and the overall wellbeing of the patient [7]. Lung cancer is often diagnosed at advanced stages when it has already spread to other organs, leading to lower survival rates, however, early detection and treatment offer the best chances of favorable outcomes. Lung cancer are diagnosed using different modalities, some of the diagnostic procedure include X-ray, Computed Tomography (CT) scan (Low Dose CT, Positron Emission Tomography (PET) CT), Magnetic Resonance Imaging (MRI), Liquid Biopsy, Bron choscopy or Volatile Organic Compounds [8],[9]. Chest X-ray screening is not advised 13 for early screening and diagnosis of lung cancer, because of its difficulty to show intra pulmonary and small lesion. Low Dose CT which offers low radiation, fast scanning speed and high sensitivity, is the most promising imaging technique for early screening of lung cancer [9]. The Computed Tomography (CT) is one of the lowest cost and effective modalities considered in clinics to examine the condition of lungs [10]. Lung cancer is evaluated using radiographs from imaging devices; while reviewing the image to determines grounds to suspect cancer, the Physician may use additional method for further investigation such as Biopsy. Once a cancer is diagnosed, it is staged, which defines how large the tumor is and where it has progressed [11]. Classification of lung cancer is defined by the location of the cancer, the spread of tumor to other areas of the body such as lymph node, other lung, the brain and others [12]. There are two major method of staging lung cancer; the number staging system which classifies the disease into Stages 1– 4 and the TNM staging system. However, the most commonly used staging system for lung cancer is the TNM system, developed by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) [12]. The TNM is an acronym of Tumor, Node and Metastasis. The Tumor (T) describes the size of the tumor and categories lung cancer into four stages T1, T2, T3 and T4. In T1 the tumor is ¡= 3cm, T2 the tumor is ¿= 3cm and ¡= 5cm, at T3 the tumor is ¿=5cm and ¡= 7cm while at T4 the tumor is ¿=7cm. Node (N) describes whether the cancer has spread to the lymph nodes and are categorized into 3, N1, N2 and N3. Metastasis (M) describes whether the cancer has spread to a different part of the body and is categorized into two, M0 and M1 [12]. While traditional methods provide the framework for diagnosis and staging, recent advances in artificial intelligence offer new tools to improve early detection and classification.

This study explores the application of Deep Learning (DL) model for the early diagnosis of lung cancer using CT scan images. The system aims to reduce false positives in medical imaging and augment clinical decision-making through automation, accuracy, and speed. This research is significant because it proposes a cost-effective, accurate, and automated method for early lung cancer diagnosis, which could improve patient survival and reduce diagnostic burden in under-resourced settings

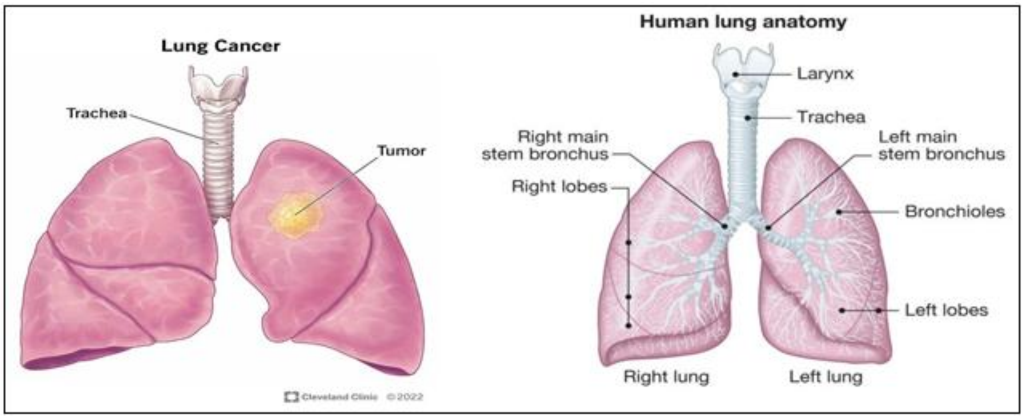


Figure 2: Images of tumor lung and a normal lung.

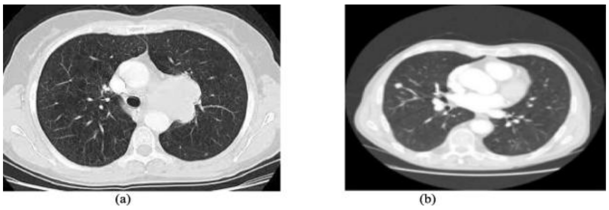


Figure3: Computed Tomography images (a) Normal (b) Diseased Lungs

2. Related Works

Several prior studies have explored various techniques for lung cancer diagnosis and classification, particularly using radiological imaging and machine learning. Chest X-rays have traditionally been the primary and most accessible method for the initial diagnosis of lung cancer Panunzio2020 [22]. On conventional radiographs, lung cancer manifests with various pathological characteristics, including nodules, central or peripheral masses, partial or total lung collapse, and infiltrates in the lung tissue [23]. The presence of pulmonary nodules with focal densities measuring between 3mm and 3cm in diameter suggests the presence of lung cancer [24]. Evaluation of more evasive features, such as invasion of the phrenic nerve or the superior vena cava, is limited by chest X-ray images. It may not be able to categorize the tumour as benign or malignant. However, the combination of a chest X-ray and a Computed Tomography (CT) image aids in determining the complete disease staging. CT plays an important role in the definition of local invasiveness and the determination of the lesion's resectability. CT is useful for both benign and malignant tumour diagnosis. For lung nodules, CT can accurately distinguish between benign and malignant nodules based on various signs; irregular shape, ground glass shadow, marginal spinous sign, and burr sign are important bases for qualitative determination of pulmonary nodules by chest [25]. CT Radiological imaging aids experienced radiologists in the identification, classification, management, and treatment planning of lung cancer patients.

The availability of radiographic tumor characteristics has made it very possible to assess tumors' features quantitatively rather than quantitatively [28] and noninvasive description of tumor has better predictive accuracy than standard clinical assessment [28]. The use of deep learning approaches to aid diagnosis has recently sparked a great deal of interest in the diagnosis of lung cancer, the shift to Deep Learning in the recent times is attributed to its capability to learn representative feature automatically from data [29].

CNN is the most used Deep Learning algorithm [30] particularly in lung cancer diagnosis. CNNs offer the advantage of end-to-end learning, it learns to extract relevant features and make predictions directly from the raw image data without the need for handcrafted feature engineering. Their ability to learn intricate patterns and adapt to various image variations has proven to be beneficial for lung cancer diagnosis, especially in tasks like nodule detection, cancer classification, and segmentation of lung tumors. The CNN approach to lung cancer diagnosis can be through an end-to-end learning or transfer learning model [30]. In a CNN, the convolutional layer applies multiple small filters, also known as kernels, to the input image. Each filter slides (convolves) over the image, performing element-wise multiplication and summation operations, which effectively extracts specific features or patterns from the input. The convolutional layer works by detecting simple patterns in the early layers and progressively more complex patterns in deeper layers. For example, in the early layers, the filters might detect basic features like edges, corners, and textures. As the information is passed through deeper layers, the filters combine these basic features to represent more abstract features, like shapes and object parts. By stacking multiple convolutional layers, CNNs can learn a hierarchy of features, with each layer building upon the previous ones. The final layer(s) of the CNN often combines these learned features to make predictions for the specific task, such as classifying the input image or segmenting specific regions of interest. Most CNN architecture consist of series of convolutional layers along with max-pooling layers as well as FC layer and final layer are SoftMax layer or output layer []. AlexNet,LeNet and VGGNet are common examples of CNN models, ResNet, GoogleNet and DenseNet are more efficient advance architectures; they all have fundamental componets such as convolutional layer, max-pooling and others in common.

To overcome limitations of insufficient data to train deep learning networks from scratch, researchers increasingly rely on transfer learning. Transfer learning is gaining deep interest in recent times among the researchers, this is because, it enables transfer of knowledge from previous activity to improve learning in the new ones. Its prominence in the fact that it provides the large properties essential to train a deep learning model and, large and complex dataset on deep learning [31]. Also, reduction in the training time, improvement in performance of neural networks and the absence of a large amount of data are some primary attractions to transfer learning techniques of deep learning [32].

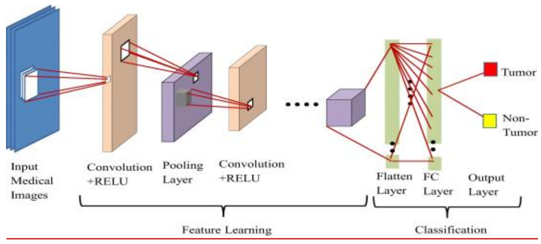


Figure 4: A sample CNN Architecture

In [33], the authors in an attempt to improve early detection and classification of lunger cancer, developed two models using two pretrained CNN models; ResNet50 and EfficientNet-B7 models. The model was trained using 5000 images categorized into three classes. The labels are adenocarcinoma, squamous cell carcinoma, carcinoma which are types of small cell lung cancer (SCLC). ResNet50 performed very better with an accuracy of 95% and 91% for EfficientNet-B7. However, the focus of the research was on speed and accuracy instead of classification of the disease into different types, such as benign and malignant and the their stagings.

Also, the authors in [34] This study uses the Accelerated Wrapper-based Binary Artificial Bee Colony algorithm (AWBABCA) for feature selection and VGG19+CNN for cancer phase classification. Morphological features are extracted from pre-processed images, and the chosen features are used for cancer classification, resulting in high strength and precision. The proposed classifier achieved the best accuracy 98.34%, precision 97.96%, recall 98.72%, and f1-score 98.32%.

Furthermore, research by [35] utilized transfer learning approach to classify lung CT images based on cancer type in patients. A pre-trained hybrid model extracts features from images and sends them to a multi-class SVM model for classification. The model is trained using the IQ-OTH/NCCD dataset, which includes 1190 cases classified into normal, benign, or malignant categories. During training and validation phases, the data is segmented using various ratios to evaluate the model's effectiveness. The 23 proposed model achieves 97% accuracy and better results in precision, sensitivity, and specificity, as well as a f1-score.

Moreover, [32] developed an efficient training model for accurately classifying and detecting lung cancer types using CT-scanned chest images. Pre-trained models like DCNN, VGG16, Inceptionv3, and RestNet50 were used. The ResNet50 model achieved 94% accuracy during training, while the DSTL model based on Deep CNNs surpassed state-of-the-art models in classification tasks. This approach improves the precision and effectiveness of lung Cancer screening and detection

3. Methodology

To develop and validate the proposed model, a well-curated public dataset was utilized. The dataset used in this study is the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset was collected in specialist hospitals in fall 2019. It's a publicly available dataset from Mendeley that was first published in

October, 2020 and the most recent version, 4th version released in May 2023. It includes 1190 images of 110 cases, categorized into normal, benign, and malignant cases. The dataset includes 1190 DICOM scans, using a Siemens SOMATOM scanner. The 110 cases range in gender, age, educational attainment, and living sta[tus, including employees, farmers, and gainers. The choice of the dataset is based on its public status and its wide use in literatures. The dataset was resized into 512 by 512 pixels each, this is to ensure standardization of the dataset and remove noises from the images. Also, the dataset was divided into two for each of the experiments. Experiment A has 70% of the dataset for training and 30% for testing, while experiment B has 80% and 20% of the dataset for training and testing respectively, the splitting stratified sampling method to maintain class distribution and ensured unbiased evaluation.

Min–Max normalization was employed to rescale pixel intensity values to the [0, 1] interval. This preprocessing technique promotes numerical stability, expedites model convergence, and enhances generalisation by ensuring uniform input distributions across diverse imaging conditions.

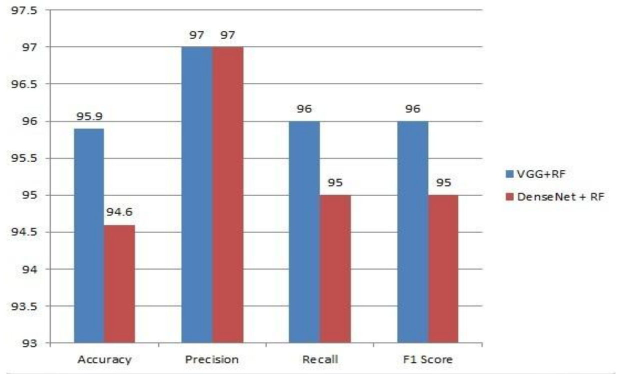
In this research work, I employed VGG16 and DenseNet121 solely as feature extractors. This means we remove the fully connected (FC) layers or dense layers from these networks. Instead, we utilize the features extracted from the convolutional layers. These extracted features are subsequently fed into a random forest classifier for the purpose of classification, specifically categorizing data into disease type, namely Normal, Benign and Malignant. This approach enables us to make informed assessments and predictions based on the deep features obtained from the pretrained VGG16 and DenseNet121 models. Standard performance metrics was employed for this research, including accuracy, recall, precision, and F-measure, for the evaluation of lung cancer classification models. Accuracy serves as a fundamental measure, representing the ratio of correctly classified cases to the total number of cases in the dataset.

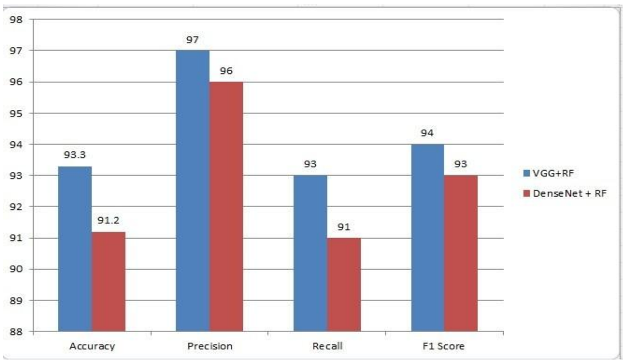
4. Results and Discussion

The evaluation of the proposed models, combining VGG16 and DenseNet121 with Random Forest (RF) classifiers, was conducted using two dataset splits: 80/20 (training/testing) and 70/30. The VGG16 + RF model achieved the highest performance, with an accuracy of 95.9%, precision of 97%, recall of 96%, and F1-score of 96% in the 80/20 split, outperforming the DenseNet121 + RF model, which recorded an accuracy of 94.6%, precision of 97%, recall of 94%, and F1-score of 95%. In the 70/30 split, VGG16 + RF maintained its superior performance with an accuracy of 93.3%, precision of 97%, recall of 93%, and F1-score of 94%, compared to DenseNet121 + RF, which achieved an accuracy of 91.2%, precision of 96%, recall of 90%, and F1-score of 93%. Both models demonstrated high precision, indicating robust classification capabilities, though VGG16 + RF consistently exhibited better recall and overall balance, as reflected in the F1-scores. To further analyze performance, confusion matrices and bar charts of performance metrics were generated (Figures 4.1–4.6). The confusion matrices detail true positive, true negative, false positive, and false negative predictions for each model and split, while the bar charts visually compare accuracy, precision, recall, and F1-scores, highlighting the consistent advantage of VGG16 + RF over DenseNet121 + RF. The results indicate that the VGG16 + RF model consistently outperformed the DenseNet121 + RF model across all metrics and dataset splits, suggesting that VGG16's feature extraction, combined with the RF classifier, is more effective for the task. However, a limitation lies in the models' generalization, as the dataset was sourced from a single origin, which may restrict applicability to diverse datasets and necessitates further validation with multi-source data to ensure robustness and generalizability. A summary of the comparative performance metrics for both models across the two experiments is shown below.

Table 4.3: The accuracy of the models based on the testing parameters

	80% - 20%	70% - 30%
VGGG16 + RF	95.90%	93.30%
DenseNet211 + RF	94.60%	91.20%





This study demonstrates the effectiveness of a deep learning approach for medical image classification, achieving high accuracy in distinguishing between [specify classes, e.g., benign and malignant tumors]. By leveraging transfer learning and an ensemble classifier, we were able to create a simple yet powerful architecture that performs well even in resource-constrained settings. This is particularly significant for applications such as telemedicine or rural healthcare, where computational power and data availability are limited. The high classification accuracy underscores the model's potential to deliver reliable results without the need for extensive computational resources, making it a viable solution for environments with restricted access to advanced technology. Despite these promising results, some limitations must be acknowledged. The dataset used was homogeneous, coming from a single scanner and source, which may limit the model's generalizability to other settings. This homogeneity raises concerns about the model's ability to perform consistently across different imaging conditions, such as variations in scanner types or patient demographics. Additionally, the small dataset size and lack of external validation mean that further testing is needed to confirm the model's performance in diverse, real-world scenarios. Future research could address these limitations by expanding the dataset to include images from multiple scanners and sources, as well as conducting external validation on independent datasets to ensure the model's robustness and generalizability. Despite these limitations, our findings have important implications. They highlight the potential of deep learning models for use in resource-constrained settings, where traditional deep learning approaches may be impractical due to high computational demands. This could have significant implications for improving healthcare access and outcomes in underserved areas, such as rural or low-income regions, where access to advanced medical imaging technology is often limited. Moreover, the success of our approach suggests that similar strategies could be adapted for other tasks requiring efficient yet accurate models, such as [specify related applications, e.g., remote diagnostics or mobile health applications]. Overall, our study demonstrates the promise of deep learning and suggests avenues for future research to further validate and expand upon our approach, ultimately contributing to more accessible and scalable solutions in medical imaging and beyond.

5. Conclusion and Future Works

This study confirms that combining pretrained deep learning models like VGG16 with a Random Forest classifier can deliver high accuracy in classifying lung cancer from CT scans. While the models performed well on the available dataset, broader generalization remains a limitation due to data homogeneity. Future research should include more diverse datasets, explore additional imaging modalities like MRI, and incorporate explainable AI tools to build clinical trust. These steps could significantly enhance the model's applicability in real-world, resource-constrained healthcare settings.

References

[1] NHS. WHO; 2022. Available from: https://www.nhs.uk/conditions/lung-cancer/.

[2] WHO. WHO; 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/ cancer.

[3] Oxford University U. Cancer Cost to UK the UK Economy; 2012. Available from: https://www. ox.ac.uk/news/2012-11-07-cancer-costs-uk-economy-%C2%A3158bn-year.

[4] K LS, Mohanty SN, K S, N A, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 2019;92:374-82. Available from: https://www.sciencedirect.com/science/article/pii/S0167739X18317011.

[5] Masud M, Sikder N, Nahid AA, Bairagi AK, AlZain MA. A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. Sensors. 2021;21(3). Available from: https://www.mdpi.com/1424-8220/21/3/748.

[6] Support MC. Cancer information and support; 2023. Available from: https://www.macmillan.org.uk/cancer-information-and-support/lung-cancer/ causes-and-risk-factors-of-lung-cancer.

[7] Cleveland C. Lung Cancer; 2023. Available from: https://my.clevelandclinic.org/health/ diseases/4375-lung-cancer.

[8] Masud M, Sikder N, Nahid AA, Bairagi AK, AlZain MA. A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. Sensors. 2021 Jan;21(3):748. Available from: https://doi.org/10.3390/s21030748.

[9] Ning J, Ge T, Jiang M, Jia K, Wang L, Li W, et al. Early diagnosis of lung cancer: which is the optimal choice? Aging. 2021 Feb;13(4):6214-27. Available from: https://doi.org/10.18632/ aging.202504.

[10] Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, et al. Deeplearning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters. 2020 Jan; 129:271-8. Available from: https://doi.org/10. 1016/j.patrec.2019.11.013.

[11] MEDICINES JH. Lung Cancer Diagnosis; 2023. Available from: https://www.hopkinsmedicine.org/health/conditions-and-diseases/lung-cancer/lung-cancerdiagnosis#:~:text=Lung%20cancer%20is%20diagnosed%20through,emission%20tomography%20(PET)%20scans.

[12] UK CR. TNM Staging for Lung Cancer; 2023. Available from: https://www.cancerresearchuk. org/about-cancer/lung-cancer/stages-types-grades/tnm-staging.

[13] Liam CK, Liam YS, Poh ME, Wong CK. Accuracy of lung cancer staging in the multidisciplinary team setting. Translational Lung Cancer Research. 2020 Aug;9(4):1654-66. Available from: https: //doi.org/10.21037/tlcr.2019.11.28.

[14] Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications. 2020 Jan;79(11-12):7731-62. Available from: https://doi.org/10.1007/ s11042-019-08394-3.

[15] Nanglia P, Kumar S, Mahajan AN, Singh P, Rathee D. A hybrid algorithm for lung cancer classification using SVM and Neural Networks. ICT Express. 2021 Sep;7(3):335-41. Available from: https://doi.org/10.1016/j.icte.2020.06.007.

[16] Elyan E, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-Garc´ıa CF, et al. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery. 2022. Available from: https://doi.org/10. 20517/ais.2021.15.

[17] Pourasad Y, Zarouri E, Parizi MS, Mohammed AS. Presentation of Novel Architecture for Diagnosis and Identifying Breast Cancer Location Based on Ultrasound Images Using Machine Learning. Diagnostics. 2021 Oct;11(10):1870. Available from: https://doi.org/10.3390/ diagnostics11101870.

[18] Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, et al. Deeplearning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters. 2020;129:271-8. Available from: https://www.sciencedirect.com/science/article/pii/S0167865519303277.

[19] Lanjewar MG, Panchbhai KG, Charanarur P. Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers. Expert Systems with Applications. 2023;224:119961. Available

from: https://www.sciencedirect.com/science/article/pii/S0957417423004633.

[20] Sleeman KE, Gomes B, de Brito M, Shamieh O, Harding R. The burden of serious healthrelated suffering among cancer decedents: Global projections study to 2060. Palliative Medicine. 2021;35(1):231-5. PMID: 32945226. Available from: https://doi.org/10.1177/0269216320957561.

[21] Zhang C, Sun X, Dang K, Li K, wei Guo X, Chang J, et al. Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network. The Oncologist. 2019 Apr;24(9):1159-65. Available from: https://doi.org/10.1634/theoncologist.2018-0908.

[22] Panunzio A, Sartori P. Lung Cancer and Radiological Imaging. Current Radiopharmaceuticals. 2020 Nov;13(3):238-42. Available from: https://doi.org/10.2174/1874471013666200523161849.

[23] Gharraf HS, Mehana SM, ElNagar MA. Role of CT in differentiation between subtypes of lung cancer is it possible? The Egyptian Journal of Bronchology. 2020 Sep;14(1). Available from: https://doi.org/10.1186/s43168-020-00027-w.

[24] Rampinelli C, Calloni SF, Minotti M, Bellomi M. Spectrum of early lung cancer presentation in low-dose screening CT: a pictorial review. Insights into Imaging. 2016 May;7(3):449-59. Available from: https://doi.org/10.1007/s13244-016-0487-4.

[25] Robinson H, Roberts MJ, Gardiner RA, Hill MM. Extracellular vesicles for precision medicine in prostate cancer – Is it ready for clinical translation? Seminars in Cancer Biology. 2023 Feb;89:1829. Available from: https://doi.org/10.1016/j.semcancer.2023.01.003.

[26] Hu Q, de F Souza LF, Holanda GB, Alves SSA, dos S Silva FH, Han T, et al. An effective approach for CT lung segmentation using mask region-based convolutional neural networks. Artificial Intelligence in Medicine. 2020 Mar;103:101792. Available from: https://doi.org/10.1016/j.artmed.2020.101792.

[27] Rehman A, Kashif M, Abunadi I, Ayesha N. Lung Cancer Detection and Classification from Chest CT Scans Using Machine Learning Techniques. In: 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA); 2021. p. 101-4.

[28] Rajasekar V, Vaishnnave MP, Premkumar S, Sarveshwaran V, Rangaraaj V. Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques. Results in Engineering. 2023 Jun;18:101111. Available from: https://doi.org/10.1016/j.rineng.2023.101111.

[29] Chaunzwa TL, Hosny A, Xu Y, Shafer A, Diao N, Lanuti M, et al. Deep learning classification of lung cancer histology using CT images. Scientific Reports. 2021 Mar;11(1). Available from: https://doi.org/10.1038/s41598-021-84630-x.

[30] Sahu M, Dash R. A Survey on Deep Learning: Convolution Neural Network (CNN). In: Smart Innovation, Systems and Technologies. Springer Singapore; 2020. p. 317-25. Available from: https://doi.org/10.1007/978-981-15-6202-0_32.

[31] Promy TF, Joya NI, Turna TH, Sukhi ZN, Ashraf FB, Uddin J. Cancer Diseases Diagnosis Using Deep Transfer Learning Architectures. In: Satu MS, Moni MA, Kaiser MS, Arefin MS, editors. Machine Intelligence and Emerging Technologies. Cham: Springer Nature Switzerland; 2023. p. 226-37.

[32] Atiya SU, Ramesh NVK, Reddy BNK. Classification of non-small cell lung cancers using deep convolutional neural networks. Multimedia Tools and Applications. 2023 Jul. Available from: https://doi.org/10.1007/s11042-023-16119-w.

[33] Venkatesh P, Kiran KR, Bharath SY, Kalyani G. Deep Learning-Based Lung Cancer Detection and Classification. In: Joshi A, Mahmud M, Ragel RG, editors. Information and Communication Technology for Competitive Strategies (ICTCS 2022). Singapore: Springer Nature Singapore; 2023. p. 33-40.

[34] Nandipati BL, Devarakonda N. VGG19CNN: Deep Learning-Based Lung Cancer Classification with Meta-Heuristic Feature Selection Methodology. Indonesian Journal of Electrical Engineering and Informatics (IJEEI). 2023 Mar;11(1). Available from: https://doi.org/10.52549/ijeei. v11i1.4394.

[35] Nigudgi S, Bhyri C. Lung cancer CT image classification using hybrid-SVM transfer learning approach. Soft Computing. 2023 May;27(14):9845-59. Available from: https://doi.org/10.1007/s00500-023-08498-x.

[36] Mader KS. The Lung Image Database Consortium image collection (LIDC-IDRI). IEEE DataPort; 2021. Available from: https://ieee-dataport.org/documents/ lung-image-database-consortium-image-collection-lidc-idri.

[37] Alyasriy H. The IQ-OTHNCCD lung cancer dataset. Mendeley; 2020. Available from: https: //data.mendeley.com/datasets/bhmdr45bh2/1.

[38] Shafi I, Din S, Khan A, D´ıez IDLT, del Jesu´s Pal´ı Casanova R, Pifarre KT, et al. An Effective Method for Lung Cancer Diagnosis from CT Scan Using Deep Learning-Based Support Vector Network. Cancers. 2022 Nov;14(21):5457. Available from: https://doi.org/10.3390/cancers14215457.

[39] Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Medical Informatics and Decision Making. 2011 Jul;11(1). Available from: https://doi.org/10.1186/1472-6947-11-51.

Author Profile

Timothy Olaniyi received the BSc. Degree in Computer Science from Kwara State University, Malete, Nigera, and M.Sc. degree in Data Science, from Robert Gordon University, Aberdeen, United Kingdom in 2020 and 2023, respectively. His research interest areas are Artificial neural network, Data Analytics, Image processing, Machine Learning, Deep Learning and Robotics.
+447917909821, timolsoft@gmail.com