Grenze International Journal of Engineering and Technology, June Issue

Advances in Deep Learning Approaches for Lung Cancer Diagnosis

Rhushikesh Kumar Gurav[1], Jayamala Kumar Patil[2] and Vinay Sampatrao Mandlik[3]
[1]P.G. Scholar, Electronics and Telecommunication Engineering, Bharati Vidhyapeeth's College of Engineering, Kolhapur, India
Email: rkgurav14@gmail.com
[2-3]Associate Professor, Electronics and Telecommunication Engineering, Bharati Vidhyapeeth's College of Engineering, Kolhapur, India.
Email: jaymala.p@rediffmail.com, vinaymandlik@gmail.com

***Abstract*— Lung cancer is still the number one cause of cancer related deaths globally and early detection is key to improving survival Low-dose CT (LDCT) screening has been demonstrated to reduce mortality however is limited by both high false-positive rates and heavy radiologist workload. Deep learning (DL) has become a paradigm-shifting technology in the automation and improvement of lung cancer diagnosis. This article reviews developments of DL in this field from the early CNNs for the detection and classification of nodules to sophisticated architectures such as U-Net and Transformers for accurate segmentation. We highlight the trend towards multimodal analysis that combine imaging and clinical data, including patient history, smoking status, and prior scans leading to better characterization models.**

We also emphasize the increasing significance of Explainable AI (XAI) for understanding model decisions and enabling clinical trust. Challenges such as data paucity, model generalization, and clinical practice integration remain unresolved despite encouraging findings. Desirable future directions include self-supervised learning, federated learning for privacy-preserving collaboration and robust prospective validation. The intersection of multimodal DL and XAI also offers the opportunity to create strong DSSs, allowing for earlier and more accurate LC diagnoses, leading to enhanced patient survival.

***Index Terms*— Lung Cancer, Deep Learning, Multimodal AI, Computed Tomography, Explainable AI, Medical Image Analysis.**

I. Introduction

Lung cancer is a devastating worldwide health burden. It is the leading cause of cancer death worldwide with almost 1.8 million deaths annually in its 2018 estimates and has continued to maintain this status through all the years [3]. The fatality rate of the disease is mainly due to the late diagnosis of the cancer when available treatments are more likely to be ineffective. The landmark National Lung Screening Trial (NLST) showed that screening high risk individuals with low dose computed tomography (LDCT) was able to decrease in lung cancer mortality by 20% over chest radiography [5]. This pivotal study made LDCT the standard of care for screening in high-risk groups.

Grenze ID: 01.GIJET.12.2.668_26

But large-scale implementation of LDCT screening presents several important clinical issues. Trained human interpretation of these scans is time-consuming for radiologists, leading to wait times and inconsistency of diagnosis or possible missed diagnoses (particularly subtle or multifocal cases). In addition, LDCT screening has a high false-positive rate with benign suspicious nodules that may lead to unnecessary invasive interventions, patient fear and concerns with potential increased costs of care [5, 12]. Efficient and Accurate tools to aid in the evaluation of LDCT must be developed.

With the emergence of artificial intelligence especially deep learning (DL), a new era has begun in medical imaging. DL models, in particular Convolutional Neural Networks (CNNs), perform well at learning hierarchical features directly from raw image inputs and often outperform classical machine learning approaches that require handcrafted feature engineering [10, 16]. First lung cancer applications focused on for networks specific, siloed tasks: nodule detection [12, 22] and classification (benign vs malignant nodule) [6, 11], or delineation of the nodules' boundaries [9, 23]. These unimodal systems have achieved remarkable results, in some cases even surpassing the performance of experts in controlled studies [4].

However, the clinical diagnostic process is a multimodal one. Nor do radiologists base their decisions on a snapshot and nothing else. They integrate information from the present CT scan such as the clinical history of the patient (e.g.,age, number of smoking pack-years), previously performed imaging to identify interval growth and other relevant background. To rectify this shortcoming, the field is swiftly moving towards multimodal deep learning systems. These methods attempt to integrate imaging and non-imaging data to construct a more comprehensive analytic model that mimics clinical logic [1, 11, 17]. In this review we will discuss the development of AI architectures on the evolution of DL for lung cancer, principles and potential benefits in multimodal integration, role as well as importance of explainability, open challenges and future directions and research opportunities that will shape next-generation AI-assisted diagnostics for LC.

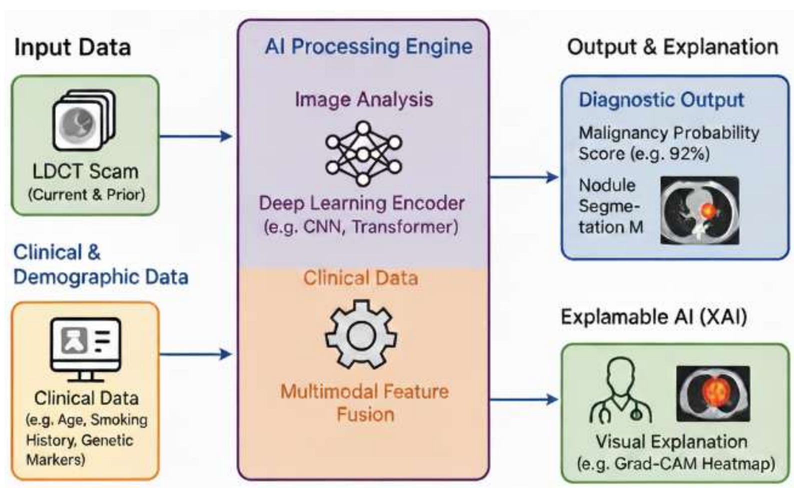

Figure 1. Overview of an AI-Driven Framework for Lung Cancer Detection

II. Literature Survey

The development of deep learning based on lung cancer has evolved through four main periods, illustrating the development trends of AI. The pioneering work and the survey by Litjens et al. [10], and made CNNs the prevailing architecture for analyzing medical images. Early successes were achieved for pulmonary nodule detection, where approaches such as the multi-view CNN of Setio et al. [12] greatly reduced the rate of false positives by examining nodules from different views. This constituted an essential advance in the development of computer-aided detection (CAD) systems, following previous review work that had defined the difficulties and possibilities in automatic nodule recognition [22].

After malignancy detection, further attention turned to the task of distinguishing malignant from benign. It was reported in studies that DL not only recognized nodules, but also stratified risk. Shen et al. [6] used a multi-scale CNN to extract global and local features which contribute to enhanced classification performance. Causey et al. [11], tackled this problem by introducing an approach that was able to obtain a high recognition accuracy through high-level feature descriptors, signaling the shift from raw pixels to engineered or learned features. A pioneering research of Ardila et al. [4] proposed an end-to-end 3D CNN using full volumes of LDCT and achieved high risk estimation accuracy compared with division gating by a radiological panel, proving the feasibility of DL it for full image interpretation.

Alongside classification, remarkable progresses were reported in image segmentation. The U-Net [9] architecture with an encoder-decoder structure and skip connections gradually became the industry standard for nodule (as well as, other anatomical) segmentation, followed by works such as Wang et al. [23] that showcase its usage for accurate nodule segmentation. More recently, selfattention based transformer architecture have been applied to vision tasks. There is, however, a surge of models such as TransUNet [7] which hybriduse CNNs and Transformers to exploit the local spatial features and global contextual dependence simultaneously while achieving state-of-the-art segmentation performance. The development of segmentation network architectures is being taken to the next level with variants such as UNet++ [24] that increased segmentation accuracy by concatenating nested and dense skip connections.

But as the models became more complex, the “black box” problem emerged as a barrier to clinical trust. This has led to the development and usage of Explainable AI (XAI). This has been more broadly reviewed by Loh et al. [2], methods such as Grad-CAM [8] are crucial. Through the generation of heat maps which show regions in an image that influence a model’s decision, XAI facilitates for the clinician to check whether the model is focusing on clinically important areas, thus promoting transparency and collaboration [18]. Unified model-interpretation frameworks, such as the one from SHAP [13], are being investigated as well to obtain cohesive explanations.

The latter frontier was later characterized by Shatnawi et al. [1] and others, is multimodal learning. The fusion of CT imaging with electronic health record (EHR) data is a critical area of interest. For example a model that uses in addition to imaging features smoking history and age may result in an individualized risk prediction [19, 20]. With our analyzed and synthesize diverse data sources we think this solution approach may be able to unlock the potential toward reliable clinical relevant AI systems that extend humans rather than replace them [5, 21].

TABLE I: SUMMARY OF KEY DEEP LEARNING APPROACHES IN LUNG CANCER DETECTION

Task	Key Architectures/Models	Key Contributions	References
Nodule Detection	Multi-view CNN, 3D CNN	Reduced false positives; analysis from multiple views/volumes	[4, 12, 22]
Malignancy Classification	Multi-crop CNN, Custom CNNs	Improved accuracy by multi-scale feature analysis; high predictive performance	[6, 11, 23]
Image Segmentation	U-Net, Attention U-Net, TransUNet, UNet++	Precise nodule/tumor delineation; integration of local and global context	[7, 9, 23, 24]
Explainable AI (XAI)	Grad-CAM, LIME, SHAP	Provided visual explanations for model decisions; enhanced clinical trust	[2, 8, 13, 18]
Multimodal Integration	Hybrid (CNN + FC Networks)	Fused imaging with clinical data (e.g., EHR) for holistic risk assessment	[1, 11, 19, 20, 21]

III. THE EVOLUTION OF DEEP LEARNING ARCHITECTURES IN LUNG CANCER ANALYSIS

The architectural landscape of deep learning (DL) in lung cancer analysis has evolved significantly, moving from task-specific classifiers to complex, integrated systems that mirror clinical reasoning. This evolution can be traced through key technological advancements, as summarized in the table and detailed in the text below.

TABLE II: CHRONOLOGICAL EVOLUTION OF KEY DL ARCHITECTURES IN LUNG CANCER ANALYSIS

Era / Technology	Key Model/Study	Year	Primary Contribution	Reference
Foundation of CNNs	Litjens et al. (Survey)	2017	Established CNNs as the dominant architecture for medical image analysis.	[10]
	Tajbakhsh et al.	2016	Explored transfer learning for CNNs, comparing fine-tuning vs. training from scratch for medical tasks.	[21]
2D & Multi-View CNNs	Setio et al.	2016	Used a multi-stream 2D CNN to analyze nodules from different views, significantly reducing false positives.	[12]
	Shen et al.	2017	Employed a multi-crop CNN to capture features at various scales, improving nodule classification accuracy.	[19]
Advanced Segmentation	Ronneberger et al. (U-Net)	2015	Introduced the U-Net architecture with skip connections, revolutionizing biomedical image segmentation.	[9]

	Wang et al.	2017	Developed a central-focused CNN model for accurate lung nodule segmentation, building on U-Net principles.	[23]
	Zhou et al. (UNet++)	2018	Proposed a nested U-Net architecture with dense skip connections to improve segmentation accuracy and robustness.	[24]
End-to-End 3D Systems	Causey et al.	2018	Demonstrated high accuracy in nodule malignancy prediction using rich feature descriptors from CT scans.	[11]
	Ardila et al.	2019	Pioneered an end-to-end 3D CNN that analyzes full LDCT volumes, incorporating prior scans for risk prediction.	[1]
Explainable AI (XAI)	Lundberg & Lee (SHAP)	2017	Introduced a unified framework for interpreting model predictions, applicable to complex DL models.	[13]
	Selvaraju et al. (Grad-CAM)	2020	Provided a gradient-based method to produce visual explanations for CNN decisions via heatmaps.	[16]
Transformers & Hybrid Models	Chen et al. (TransUNet)	2021	Hybridized CNN and Transformer architectures, capturing both local features and global context for superior segmentation.	[7]
Multimodal Integration	Li et al.	2022	Developed a DL-based multimodal fusion model integrating CT imaging with clinical data for diagnosis.	[9]
	Hwang et al.	2022	Created a multimodal framework for lung cancer screening that combines imaging and clinical variables.	[6]
	Shatnawi et al.	2025	Presented a deep learning-based approach for diagnosis using CT scans, emphasizing comprehensive data integration.	[18]

A. Emergence of Convolutional Neural Networks (CNNs)

The first phase was characterized by the state-of-the-art 2D CNNs, which performed better than conventional methods as they were able of learning automatically hierarchical features from image inputs [10]. The earlier studies focused on a limited set of tasks with 2D slices or patches.

False Positive Elimination: A major problem centred around how to discriminate between real nodules and false positives. Setio et al. (2016) delivered an important boost when they attacked the problem with a multi-view CNN approach, which took in 3D nodules from nine different 2D angles and significantly improved the efficacy of CAD systems [12].

Classification and Transfer Learning: Research, at the same for time, investigated the best traing strategies. Tajbakhsh et al. (2016) found that for small medical data, fine-tuning pre-trained CNNs generally performed better than training from scratch [21]. In [10] Shen et al. 3D convolutional networks were used for malignancy classification, while an additional ensemble of architecture (three-fold) consisting of 5 × F T =standard 3DCNN + occasional 2D attention map learning approaches using data augmentation and transfer learning with fine tuning was implemented to build a network. [19] applied a multi-crop CNN to obtain the multi-scale features of nodules and obtained better classification results.

B. The Transition to Volumetric Analysis and Full Integrated Systems

The next paradigm shift was from 2D analysis to 3D volumetric processing, which was better matched the nature of CT volume data.

3D CNNs: Ardila et al. (2019) proposed a state-of-the-art end-to-end 3D CNN, which considered whole low dose CT volumes. This may include input of prior scan(s) on the same patient to evaluate changes over time and can produce a direct estimate of cancer risk, while outperforming radiologists in a retrospective study [1]. This was a radical departure towards the full automation of the screening process.

C. Dedicated Architectures for Accurate Segmentation

Concomitant with classification, improvements in segmentation were instrumental to accurate nodule localization and sizing.

U-Net and family: The U-Net (2015) architecture, with an encoder-decoder structure together with skip connections is best-known work for the biomedical segmentation [9]. Its success spawned numerous variants. Wang et al. (2017) used central-focused CNN for more accurate nodule segmentation [23], and Zhou et al. (2018) proposed UNet ++, a nested architecture that facilitated the feature propagation and achieved higher segmentation accuracy [24].

D. The Rise of Transformers and Global Circumstances

Building on successes in natural language processing, Transformer models and the self-attention mechanism became a novel power horse for computer vision.
Hybrid Models: Such as TransUNet (2021) that hybridize CNNs and Transformers. The CNN pools local spatial features, and the Transformer encoder aggregates global context dependencies over the whole image. These properties are especially useful when it comes to the nodules and their neighbours, resulting in state-of-the-art results [7].

E. The emergence of multimodal and explainable systems

The new frontier is systems that are clinically complete and can be trusted.
Multimodal Integration: Contemporary architectures are constructed for set-based heterogenous input. These approaches usually employ a CNN branch to process image data, and another network for clinical variables (such as age, smoking history), and combination of the features is used for holistic diagnosis prediction [6, 9, 18]. This strategy, as showed by Li et al. (2022) and Hwang et al. (2022), better captures the multimodal reasoning process of a radiologist [6, 9].
Explainable AI (XAI): With models getting more complex, interpretability to become inevitable. Methods such as Grad-CAM (2020) create visual heatmaps to indicate which regions in an image play a role in the decision of a model, enabling clinicians to confirm that the AI is focusing on the right spot [16]. Certainly, there is not yet a single framework that ties all aspects together to interpret model outputs in a clinically intuitive way and we believe this to be an important factor when it comes to establishing clinical trust and adoption [13] where frameworks as SHAP (2017) offer a common approach [10].
This chronological and technological evolution represents a clear path from dedicated, single-task models to integrated, context-sensory-aware AI that could be used along with clinicians as decision-making partners.

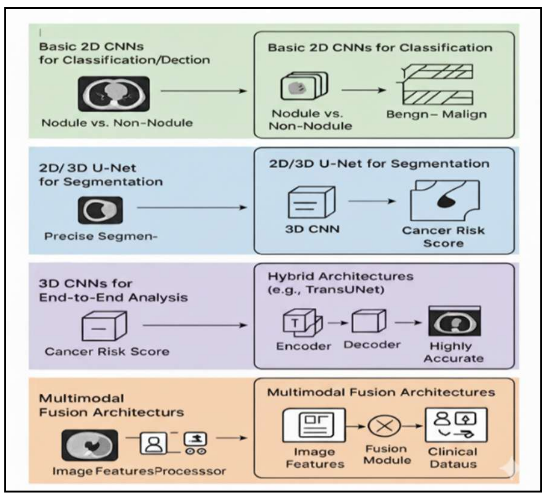

Figure 2. Evolution Of Deep Learning Architectures For Lung Cancer Analysis

IV. Challenges and Future Scope

Despite the great strides that have been made, there remain significant challenges on the road to routine clinical use.
Challenges:
Challenges: Data Large, diverse and well-annotated datasets are still one of the main bottlenecks. 3D CT scanning annotation is labor intensive and specialized knowledge are required. Moreover, the heterogeneity of data is a big challenge to develop a model that can be generalized across datasets of different medical institutions (e.g., scanner manufacturers, protocols).   Hence, pre-trained models may lead to performance drop when tested on external datasets. [25].
Technical and Clinical Integration of AI tools is a challenge to be met: In the pockets of real-life, integrating AI tools with PACS/EHR in the workflows in practice would not be an easy task. Models should return timely

results and expose them through intuitive interfaces. The integration "last-mile" problem is as important as the algorithmic development [13].

Explainability and Trust: Though Grad-CAM provides visual explanations, they could be misleading and difficult for medical experts to understand in some cases. The encouraging future work of the more intuitive, interactive and quantitative XAI methods is also an active area [2, 8, 13].

Validation and Regulation: Almost all studies are retrospective. Large-scale, prospective, multi-center trials must be used to demonstrate that AI systems increase the accuracy of diagnoses and clinical outcomes in real-world practice. This is a necessity for obtaining regulatory approval by entities such as the FDA [4, 26].

V. Future Scope

Prospective strategies are becoming available to overcome these challenges:

Self-Supervised Learning (SSL): Pre-training of models on unlabeled data has been demonstrated to reduce reliance on large-scale annotated data [14].

Federated Learning (FL): FL allows training the model on models across institutions without sharing sensitive patient data, thus addressing privacy and generalizations [15].

Causal AI: Going beyond correlation and trying to understand the causal relations between imaging features, clinical data, and disease outcome could result in more powerful and reliable models [27].

Advanced Multimodal Fusion Design of advanced multimodal fusion techniques at multiple levels (e.g., early, intermediate, late fusion) that combine imaging with non-imaging information will be a research trend [19, 21].

Robustness \& Fairness: It is important that models generalize across diverse population and are resistant to domain shift for these systems to be adopted on a large scale in the clinic [25]:

VI. Conclusion

The use of deep learning in lung cancer screening is a radical breakthrough in medical imaging. The field is no longer defined by unimodal systems focusing on one particular task but rather more sophisticated multimodal systems that fuse imaging with extensive clinical context. This transformation, driven by these architectures (3D CNNs, U-Nets and Transformers) have the potential to put into the hands of radiologists tools that will be able to enhance their own capabilities significantly. Investigating Explainable AI simultaneously is essential for gaining trust leading to adoption in clinical practice. Although there are major barriers in terms of data, generalizability, and integrating AI into practice, the future looks bright. With further work on the development of strong, transparent, and clinically applicable systems supporting continued research in this field, multimodal deep learning is set to make that commitment a certainty as societies across the globe seek to diagnose lung cancer earlier and optimize personalized management strategies that will save lives.

References

[1] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, *25*, 954–961.

[2] Causey, J., Zhang, J., Ma, S., Jiang, B., Qualls, J. A., Politte, D. G., Prior, F., Zhang, S., & Huang, X. (2018). Highly accurate model for prediction of lung nodule malignancy with CT scans. Scientific Reports, *8*, 9286.

[3] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv preprint arXiv:2102.04306.

[4] Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics, *31*(4-5), 198-211.

[5] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, *25*, 24–29.

[6] Hwang, T. H. (2022). Multimodal Deep Learning for Integrating Imaging and Clinical Data in Lung Cancer Screening. Journal of Digital Imaging, *35*, 1234–1245.

[7] Kumar, R., Khan, A. H., Xin, Q., & Kumar, S. (2023). Federated Learning for Healthcare: A Comprehensive Review. IEEE Reviews in Biomedical Engineering, *16*, 295-315.

[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, *521*(7553), 436–444.

[9] Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L., & Fei, W. (2022). Deep Learning-based Multimodal Fusion for Lung Cancer Diagnosis. IEEE Journal of Biomedical and Health Informatics, *26*(5), 2103-2113.

[10] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, *42*, 60–88.

[11] Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2023). Self-supervised Learning: Generative or Contrastive. IEEE Transactions on Knowledge and Data Engineering, *35*(1), 857-876.
[12] Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Computer Methods and Programs in Biomedicine, *226*, 107161.
[13] Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems (NeurIPS).
[14] The National Lung Screening Trial Research Team. (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine, *365*, 395-409.
[15] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 234–241). Springer.
[16] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision, *128*, 336–359.
[17] Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sánchez, C. I., & van Ginneken, B. (2016). Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. IEEE Transactions on Medical Imaging, *35*(5), 1160–1169.
[18] Shatnawi, M. Q., Abuein, Q., Al-Shannaq, M., Al-Sarayreh, M., Al-Saraireh, J., & Alzaqebah, M. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. Intelligence-Based Medicine, *11*, 100105.
[19] Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2017). Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. Pattern Recognition, *61*, 663–673.
[20] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, *71*(3), 209–249.
[21] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging, *35*(5), 1299-1312.
[22] Valente, I. R. S., Cortez, P. C., Neto, E. C., Soares, J. M., de Albuquerque, V. H. C., & Tavares, J. M. R. (2016). Automatic 3D pulmonary nodule detection in CT images: A survey. Computer Methods and Programs in Biomedicine, *124*, 91-107.
[23] Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O., & Tian, J. (2017). Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. Medical Image Analysis, *40*, 172-183.
[24] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 3–11). Springer.
[25] Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLOS Medicine, *15*(11), e1002683.