Emerging Trends and Key Themes in Deep Learning for Lung Cancer: A Bibliometric Analysis from 2014 to 2025

Rapeepat Klangbunrueang[1*]

[1*] Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand
Corresponding author: rapeepat.klangbunrueang@kkumail.com

Abstract

This study conducts a bibliometric analysis exploring deep learning (DL) applications in lung cancer diagnostics from 2014 to 2025. Through analysis of 2,529 publications from 868 sources, the research identifies key trends, emerging themes, and collaborative networks within this rapidly evolving field. Central research domains include "deep learning," "lung cancer," and "medical imaging," with notable advancements in "transfer learning" and "bioinformatics" contributing to enhanced diagnostic precision. Despite the demonstrated potential of convolutional neural networks (CNNs) and explainable AI frameworks, significant challenges persist, including dataset heterogeneity, model interpretability limitations, and inequitable access to AI technologies. Examination of global collaboration patterns reveals pronounced disparities in research activity and resource distribution. This study emphasizes the critical importance of integrating computational and clinical perspectives to advance DL-driven lung cancer diagnostics, advocating for inclusive and cost-effective solutions to improve early detection capabilities, personalized treatment approaches, and equitable healthcare delivery worldwide.

Keywords: Emerging Trends, Deep Learning, Lung Cancer, Bibliometric Analysis, Machine Learning.

Received: 20 November 2024 **Revised:** 5 December 2024 **Accepted:** 30 December 2024

---

1. Introduction

Lung cancer remains a leading cause of cancer-related mortality worldwide, accounting for 1.8 million deaths annually (Siegel et al., 2021). Despite significant advancements in medical technologies and treatment methods, disparities in lung cancer outcomes persist across different populations, largely influenced by socio-economic inequalities, access to healthcare, and systemic issues within healthcare delivery systems (Islami et al., 2021). Addressing these disparities requires not only innovations in diagnostic and treatment tools but also equitable implementation of policies that ensure access to quality care for all individuals. Interdisciplinary efforts and collaboration between healthcare providers, policymakers, and researchers are critical to overcoming these challenges (Mulshine, 2005; Islami et al., 2021).

Advancements in artificial intelligence (AI), particularly deep learning (DL), have transformed the landscape of lung cancer diagnosis and management. Convolutional neural networks (CNNs) have demonstrated remarkable capabilities in automating complex tasks such as tumor detection, segmentation, and classification from medical imaging data (Ardila et al., 2019). These technologies offer the potential to improve early detection rates, reduce diagnostic errors, and provide personalized treatment strategies (Sarkar et al., 2024). Moreover, radiomics, which quantifies imaging features like texture and shape, enhances prognostic accuracy by uncovering patterns that may not be visible to the human eye (Aerts et al., 2014). AI methods have been applied to large datasets such as LUNA16 and have achieved high precision and recall rates, enabling more accurate nodule detection (Bharti et al., 2022; Setio et al., 2016).

The integration of DL techniques, such as transfer learning and hybrid CNN models, has further refined diagnostic systems, allowing for higher accuracy in distinguishing between benign, malignant, and normal cases (Sarkar et al., 2024; Lalitha & Murugan, 2023). Furthermore, explainable AI (XAI) frameworks emerge as essential tools for improving the transparency and interpretability of these models, fostering trust among clinicians and enhancing their usability in clinical workflows (Apostolidis & Papakostas, 2021; Wani et al., 2023). Explainability is particularly critical when adopting advanced methodologies like adversarial learning, as even minor perturbations in input data can degrade model performance (Apostolidis & Papakostas, 2021).

Despite these promising developments, several barriers hinder the adoption of DL technologies in real-world clinical workflows. Variability in imaging datasets, differences in acquisition protocols, and a lack of standardized data remain major challenges that limit the generalizability of AI models (Setio et al., 2016). Additionally, the "black-box" nature of many DL models raises concerns among clinicians, who require transparent and interpretable systems to ensure trust and reliability in AI-driven decision-making processes (Davri et al., 2023). In resource-limited settings, the lack of infrastructure and trained personnel exacerbates these issues, creating disparities in access to cutting-edge diagnostic tools (Jung et al., 2023; Rao et al., 2023). Collaborative efforts to develop cost-effective tools and provide education on AI implementation are necessary to ensure that these innovations benefit all populations (Islami et al., 2021).

Another critical aspect of improving lung cancer diagnostics lies in addressing systemic and societal barriers. Policies that support funding for AI research and its integration into healthcare systems are essential for achieving sustainable progress (Siegel et al., 2022; Makubhai et al., 2024). Additionally, global collaboration plays a pivotal role in advancing research and technology development, with interdisciplinary approaches driving the integration of computational and medical sciences (Tran et al., 2023; Jacobs et al., 2021). For example, the development of frameworks like DeepXplainer and Marine Predators Algorithm has demonstrated the potential of integrating biology, imaging, and computational techniques for comprehensive cancer diagnostics (Mengash et al., 2023; Wani et al., 2023).

This study seeks to explore the intersection of AI technologies and healthcare policies in improving lung cancer diagnosis. Specifically, it aims to address three critical questions: (1) What are the emerging trends and advancements in the application of deep learning techniques for lung cancer detection and classification? (2) What are the challenges and limitations in integrating deep learning algorithms into clinical workflows for lung cancer diagnosis? (3) What are the key factors influencing the effectiveness of deep learning-based approaches in predicting lung cancer risk and outcomes? By examining these questions, this research aims to provide insights into the role of DL in addressing disparities in lung cancer care while proposing strategies for the equitable and effective adoption of AI technologies.

The findings of this study will contribute to the growing body of knowledge on AI-driven healthcare solutions, offering practical recommendations for integrating DL technologies into clinical workflows. Additionally, the study highlights the importance of global collaboration and policy reform in ensuring that advancements in lung cancer diagnostics benefit all patients, regardless of socio-economic status or geographic location.

2. Methodology

2.1 Research Design

To ensure methodological transparency and reproducibility, this study adheres to the PRISMA framework for systematic literature selection. The methodology follows a rigorous four-phase process: identification, screening, eligibility assessment, and inclusion. Bibliometric data were initially retrieved from Scopus and Web of Science databases using standardized search terms including "lung cancer" and "deep learning." Following the removal of duplicates and non-English publications, the final corpus comprised 2,529 documents from 868 sources. The analytical approach incorporated both performance analysis and science mapping techniques to examine publication trends, co-authorship networks, and keyword co-occurrences, thereby providing a comprehensive assessment of research progression and collaboration patterns in the field.

2.2 Data Collection
The bibliometric data for this study was retrieved from two leading academic databases, Scopus known for their comprehensive coverage of high-quality, peer-reviewed literature. The dataset spans publications from 2014 to 2025, ensuring an up-to-date exploration of trends in the application of deep learning for lung cancer diagnostics. A systematic search strategy was employed, utilizing keywords such as "lung cancer," "deep learning," Only English-language journal articles, reviews, and conference proceedings were included. After removing duplicates and irrelevant records, the final dataset comprised 2,529 documents sourced from 868 journals and conferences. These documents reflect contributions from 8,102 authors, with an international co-authorship rate of 21.59%. Key metadata, such as publication year, author affiliations, keywords, and citation counts, were extracted for analysis, forming the foundation for the subsequent bibliometric mapping and performance evaluations.

2.3 Data Analysis
The data analysis in this study utilized bibliometric techniques to evaluate the development and dynamics of research on deep learning applications in lung cancer diagnostics. A combination of performance analysis and science mapping was employed to identify publication trends, keyword patterns, and collaborative networks within the field. Performance analysis focused on examining descriptive metrics such as the total number of publications, co-authorship rates, and the distribution of keywords across studies. These metrics provided a quantitative understanding of the research productivity and the impact of studies in this domain.

Science mapping techniques were implemented to explore thematic structures and relationships between research areas. Co-occurrence network analysis was applied to keywords to identify central themes and emerging topics, highlighting the interdisciplinary nature of the research. Furthermore, collaboration networks were analyzed to map partnerships among institutions and countries, illustrating the extent of global cooperation in advancing this field.

Advanced bibliometric tools, including Bibliometric in R were used to process and visualize the data. These tools enabled the identification of significant research clusters, trends in topic evolution, and key contributors to the field. By integrating these approaches, the study provided a comprehensive view of the progress and potential future directions of research in deep learning for lung cancer diagnostics.

3. Results

The bibliometric analysis covered publications from 2014 to 2025, capturing data from 2,529 documents authored by 8,102 contributors across 868 sources. Single-authored documents accounted for 50 publications, while international co-authorship represented 21.59% of all collaborations. The average number of citations per document was 12.48, indicating robust academic engagement with the field. The average age of the documents analyzed was 1.47 years, reflecting the recent and dynamic nature of the research. Figure 1 summarizes these key bibliometric characteristics, illustrating the growth and collaboration trends in the field of deep learning applications for lung cancer diagnostics.

Figure 1. Main information.

.

The analysis of annual scientific production reveals a substantial increase in research activity on deep learning applications in lung cancer diagnostics from 2014 to 2024. The steady growth observed until

2020 corresponds with the maturation of deep learning technologies and increased availability of annotated medical imaging datasets. A notable surge in publications occurs between 2020 and 2024, likely driven by advancements in computational resources and growing interdisciplinary collaborations. However, a sharp decline is observed in 2025, which may reflect incomplete indexing of recent publications rather than a true reduction in research activity. Figure 2 illustrates the annual distribution of publications, emphasizing the dynamic and rapidly evolving nature of the field.

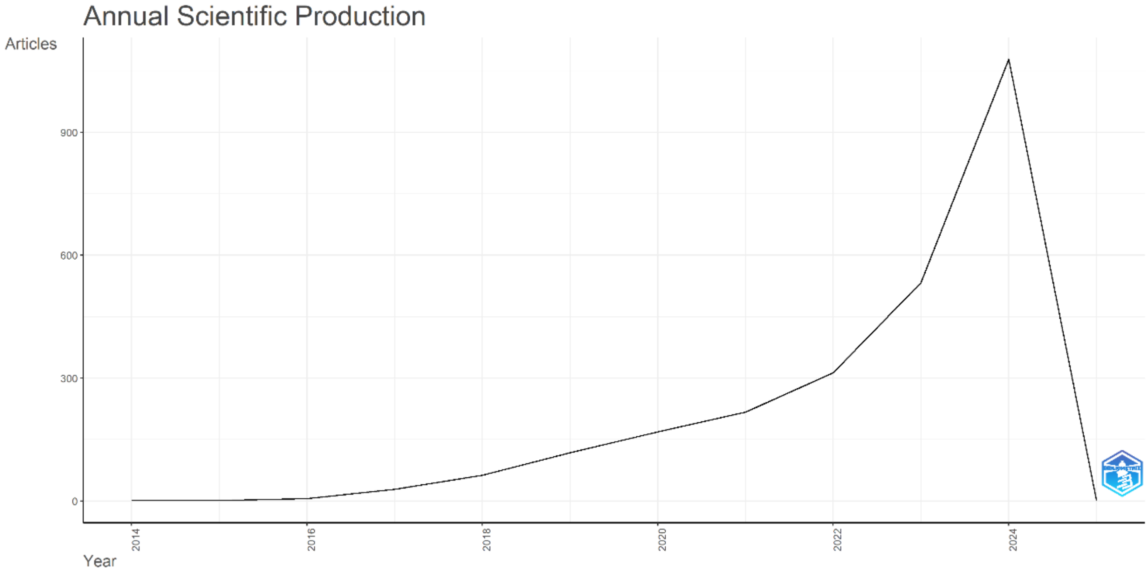

Figure 2. Annual scientific production in deep learning for lung cancer diagnostics (2014–2025).

3.1 Three field plot

To understand the intellectual structure of the research field, a three-field plot was generated to analyze the connections between keywords, article titles, and contributing institutions. As shown in Figure 3, "deep learning," "lung cancer," and "medical imaging" emerge as central themes, connecting to various advanced methodologies, including "transfer learning," "convolutional neural networks," and "radiomics." These keywords indicate a focus on diagnostic accuracy and novel analytical frameworks, demonstrating the field's growing sophistication.

The plot further highlights the strong contributions from institutions like Shanghai Jiao Tong University, Northeastern University, and other key players driving advancements in lung cancer diagnostics. These connections reflect the interdisciplinary and collaborative nature of the field, underscoring the integration of computational sciences with medical research.

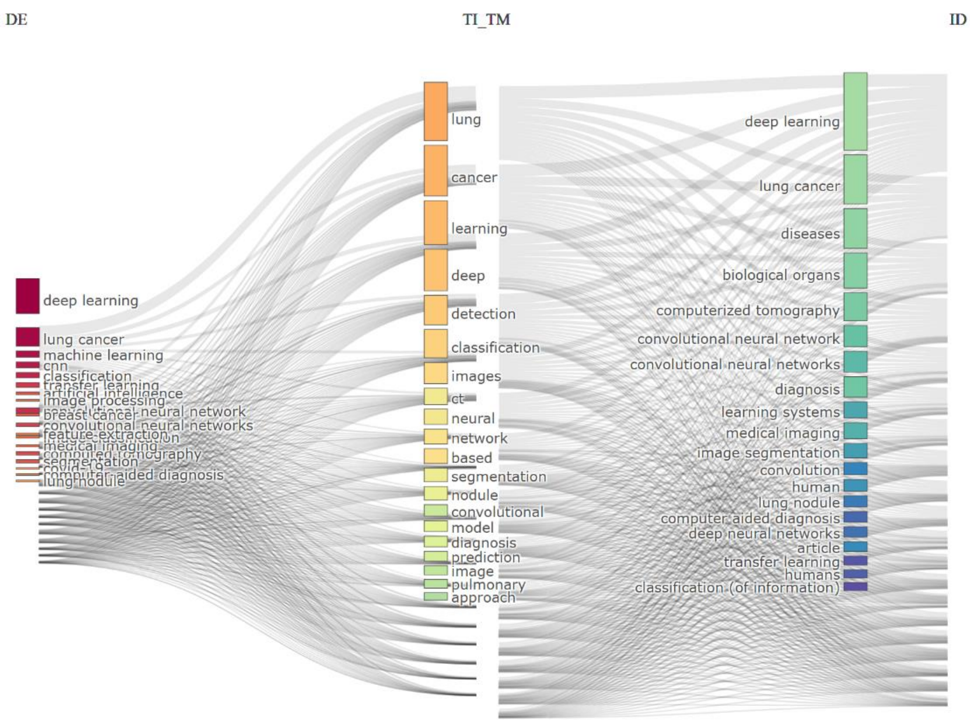

Figure 3. Three-field plot.

3.2 Co-occurrence Network

The co-occurrence network presented in Figure 4 illustrates the relationships among the most frequently occurring keywords in deep learning applications for lung cancer research. This visualization highlights the dominance of "deep learning" as a central node, closely linked to other key terms such as "lung cancer," "diagnosis," "medical imaging," and "convolutional neural networks."

The network also reveals emerging themes such as "contrastive learning," "image segmentation," and "neural networks," suggesting a growing emphasis on advanced methodologies in medical image analysis. The clustering of terms like "biological organs," "medical imaging," and "classification" reflects a focus on leveraging deep learning to improve diagnostic precision and outcomes in clinical settings. Additionally, the dense interconnections emphasize the interdisciplinary nature of this research area, bridging computational, medical, and biological sciences.

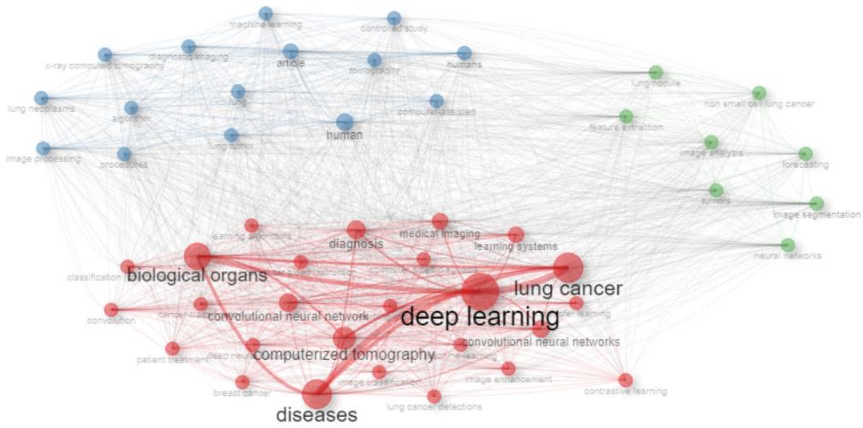

Figure 4. Co-occurrence network.

3.3 Trend topics

The analysis of trending topics over the study period, presented in Figure 5, highlights the evolution of research focus areas in the application of deep learning for lung cancer diagnostics. Terms such as "lung cancer," "deep learning," and "medical imaging" remain consistently prominent, reflecting their foundational importance to the field.

Emerging topics like "transfer learning," "contrastive learning," and "bioinformatics" have shown significant growth in recent years, indicating a shift towards more sophisticated methodologies and their integration with biological data for enhanced diagnostic precision. The prominence of terms like "neural networks" and "lung nodule detection" underscores ongoing efforts to improve the accuracy and efficiency of automated systems for early cancer detection.

Additionally, the timeline reveals that topic such as "10-fold cross-validation" and "convolutional neural networks (CNN)" gained attention in earlier years, serving as a foundation for the subsequent exploration of more advanced techniques. The consistent rise of terms like "collaborative learning" and "medical computing" reflects the interdisciplinary nature of the research, emphasizing the importance of integrating computational and clinical perspectives.

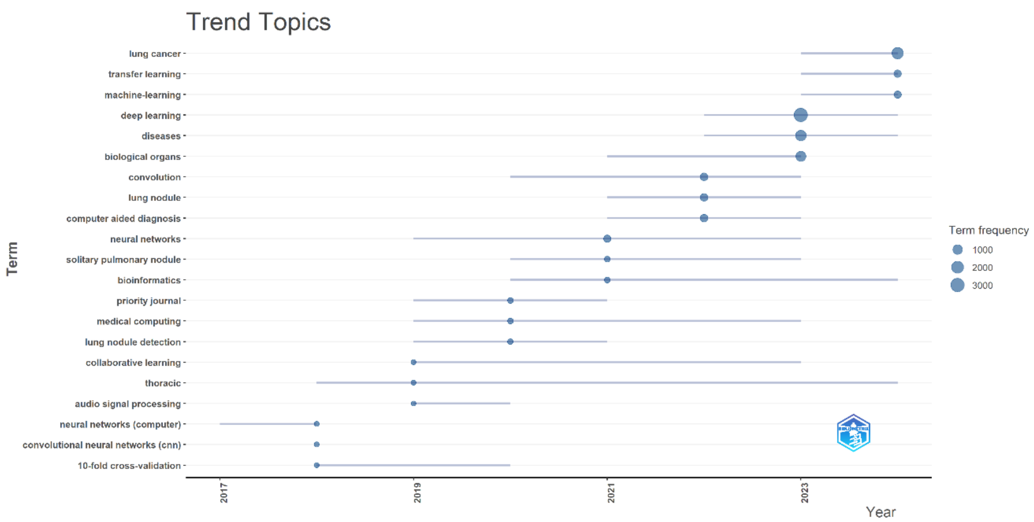

Figure 5. Trend topics.

3.4 Collaborative Networks
The collaborative network presented in Figure 6 highlights the interconnected relationships between leading institutions contributing to deep learning research in lung cancer diagnostics. Key institutions, such as Shanghai Jiao Tong University and Northeastern University, appear as central hubs, demonstrating their significant influence and extensive collaborative efforts within the research community.

The clustering of institutions into distinct groups reflects regional and thematic concentrations of expertise. For example, institutions in China and India, including Chitkara University Institute of Engineering and Technology, show strong internal collaborations, while also engaging with global partners. These networks underscore the critical role of international collaboration in advancing the field.

Despite the global nature of these networks, disparities remain evident. Leading institutions predominantly represent resource-rich regions, highlighting a potential gap in participation from underrepresented and resource-limited areas. Addressing this gap is crucial for ensuring inclusive and equitable advancements in AI-driven lung cancer research.

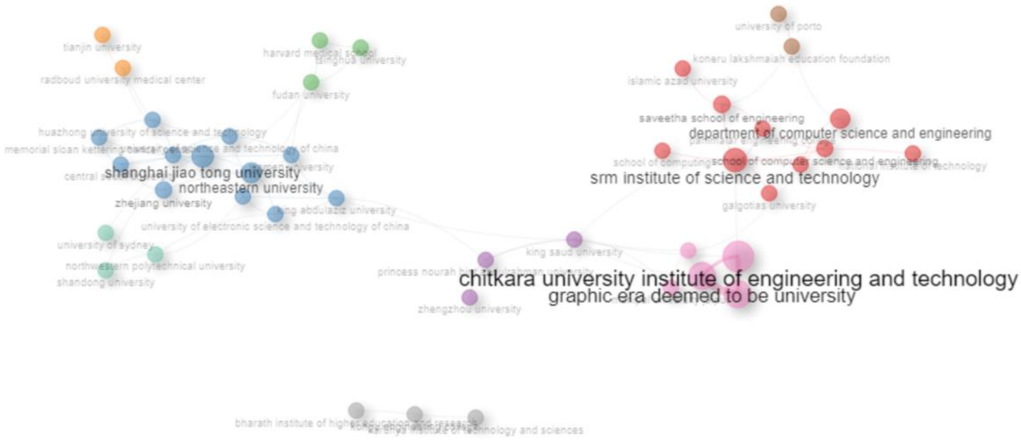

Figure 6. Collaboration network.

4. Discussion

This bibliometric analysis offers a comprehensive examination of the development and dynamics of deep learning applications in lung cancer diagnostics, uncovering key trends, collaborative networks, and thematic advancements. The steady growth in publications over the past decade highlights the increasing recognition of artificial intelligence as a transformative tool in healthcare. A surge in research activity since 2020 correlates with advancements in computational power, the widespread adoption of machine learning techniques, and the availability of annotated medical imaging datasets. These factors have facilitated the development of complex models like convolutional neural networks (CNNs) and transfer learning techniques, which have shown significant potential in improving diagnostic precision.

Thematic analysis revealed "deep learning," "lung cancer," and "medical imaging" as central topics in the field. Emerging themes such as "transfer learning," "contrastive learning," and "bioinformatics" reflect a progression toward more sophisticated methodologies, including the integration of multimodal data for personalized medicine. However, several barriers to clinical adoption persist. Dataset variability, differences in imaging acquisition protocols, and a lack of standardized frameworks hinder the generalizability of AI models. Furthermore, concerns surrounding the "black-box" nature of deep learning algorithms continue to impede trust and adoption among healthcare practitioners. Ethical considerations, particularly data privacy and potential biases, present additional challenges that require immediate attention.

Collaboration networks demonstrate the critical role of international and institutional partnerships in driving advancements in this field. Prominent contributors, such as Shanghai Jiao Tong University and Northeastern University, underscore the global character of the research. However, disparities in resource distribution and the concentration of research activity in specific regions point to the need for more inclusive efforts to bridge gaps in access to AI tools and research opportunities. Additionally, this study highlights the importance of developing explainable AI (XAI) frameworks to increase transparency in decision-making, allowing clinicians to better interpret AI-generated diagnoses and predictions. Increased interdisciplinary collaboration between AI researchers and medical professionals is essential to refining AI models to ensure their real-world applicability and reliability in clinical settings.

This study provides valuable insights into the development and application of deep learning (DL) techniques in lung cancer diagnostics; however, several limitations should be acknowledged. First, the analysis relied exclusively on data retrieved from Scopus and Web of Science databases. While these sources are widely recognized for their comprehensive coverage of peer-reviewed research, they may exclude relevant studies from other databases or non-English publications, potentially limiting the diversity and comprehensiveness of the dataset. Additionally, the temporal scope of the analysis, which focused on publications from 2014 to 2025, may overlook foundational research conducted prior to this period that shaped the field's trajectory.

Another limitation arises from the reliance on predefined keywords such as "deep learning" and "lung cancer." This approach, while effective in targeting the field's core topics, may have excluded studies employing alternative terminologies or interdisciplinary perspectives, thereby introducing potential biases in dataset composition. Moreover, the study emphasized quantitative metrics, such as publication counts, keyword co-occurrence, and collaboration networks, without assessing the methodological rigor or practical applicability of individual studies. This lack of qualitative analysis may result in an incomplete understanding of the clinical relevance of the findings.

The observed stabilization in annual growth rates for 2025 may also reflect incomplete indexing of recent publications rather than a true plateau in research activity. Furthermore, the concentration of collaborations among a few leading institutions highlights potential disparities in resource accessibility and research opportunities across different regions, which could limit the generalizability of the findings. Lastly, the study did not directly address the real-world implementation of DL technologies in clinical workflows, leaving a gap in understanding their practical impact and effectiveness in diverse healthcare settings. Future research should explore the practical integration of DL models into clinical practice, including physician acceptance, interpretability concerns, and regulatory challenges.

By acknowledging these limitations, this study provides a foundation for refining future bibliometric analyses and advancing the field of DL applications in lung cancer diagnostics, ensuring that future research addresses these gaps to achieve a more comprehensive and inclusive understanding of the domain.

5. Conclusion

This study highlights the rapid growth and evolving nature of deep learning applications in lung cancer diagnostics, emphasizing its interdisciplinary and global character. By identifying key themes and emerging trends, such as "transfer learning," "contrastive learning," and "bioinformatics," this analysis addresses Research Question 1 and demonstrates the progression toward sophisticated and personalized diagnostic tools.

Despite significant advancements, challenges related to data standardization, model interpretability, and ethical implementation remain, directly addressing Research Question 2. These findings highlight the importance of explainable AI frameworks and the integration of multimodal data to improve diagnostic accuracy and clinical acceptance. Additionally, this study emphasizes the importance of enhancing AI fairness and mitigating biases in lung cancer diagnostics to ensure equitable healthcare delivery.

The analysis also underscores the critical role of collaboration in advancing the field. However, disparities in resource distribution and the concentration of research activity in specific regions suggest a need for more inclusive and equitable efforts, aligning with Research Question 3. Future research should focus on fostering collaborations with underrepresented regions, developing scalable and cost-effective solutions, and addressing ethical concerns to ensure that AI-driven tools benefit diverse healthcare systems. Additionally, regulatory frameworks and standardization protocols should be developed to support the clinical adoption of AI-based diagnostic tools.

By bridging the gap between academic advancements and clinical practice, the potential of deep learning to revolutionize lung cancer diagnostics and reduce global health disparities can be fully realized. This study provides a foundation for future research, offering valuable insights into thematic evolution, collaboration patterns, and emerging directions in the field. Researchers, clinicians, and policymakers can leverage these findings to advance the role of artificial intelligence in transforming lung cancer care for all populations.

References

Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M. M., Leemans, C. R., Dekker, A., Quackenbush, J., Gillies, R. J., & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications, 5*, Article 4006. https://doi.org/10.1038/ncomms5006

Apostolidis, K. D., & Papakostas, G. A. (2021). A survey on adversarial deep learning robustness in medical image analysis. *Electronics*, *10*(17), Article 2132. https://doi.org/10.3390/electronics10172132

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L. H., Tse, D., Etemadi, M., Ye, W., Corrado, G. S., 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*(7), 954–961. https://doi.org/10.1038/s41591-019-0447-x

Bharti, M., Choudhary, J., & Singh, D. P. (2022). Detection and classification of pulmonary lung nodules in CT images using 3D convolutional neural networks. *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1319–1324. https://doi.org/10.1109/ICACCS54159.2022.9785134

Davri, A. S., Birbas, E., Kanavos, T., Ntritsos, G., Giannakeas, N., Tzallas, A. T., & Batistatou, A. (2023). Deep learning for lung cancer diagnosis, prognosis and prediction using histological and cytological images: A systematic review. *Cancers*, *15*(15), Article 3981. https://doi.org/10.3390/cancers15153981

Islami, F., Guerra, C. E., Minihan, A. K., Yabroff, K. R., Fedewa, S. A., Sloan, K., Wiedt, T. L., Thomson, B., Siegel, R. L., Nargis, N., Winn, R., Lacasse, L. A., Makaroff, L., Daniels, E., Patel, A. V., Cance, W., & Jemal, A. (2021). American Cancer Society's report on the status of cancer disparities in the United States, 2021. *CA: A Cancer Journal for Clinicians, 72*(1). https://doi.org/10.3322/caac.21703

Jacobs, C., Setio, A. A., Scholten, E. T., Gerke, P. K., Bhattacharya, H., Hoesein, F. A. M., Brink, M., Ranschaert, E. R., de Jong, P. A., Silva, M., Geurts, B., Chung, K., Schalekamp, S., Meersschaert, J., Devaraj, A., Pinsky, P. F., Lam, S., van Ginneken, B., & Farahani, K. (2021). Deep learning for lung cancer detection on screening CT scans: Results of a large-scale public competition and an observer study with 11 radiologists. *Radiology: Artificial Intelligence*, *3*(6), Article e210027. https://doi.org/10.1148/ryai.2021210027

Jung, K., Kang, M. J., Bang, S. H., Choi, S. H., Park, E., Yun, H., Kim, H., Kong, H. J., Im, J. S., Seo, H. G., Mee, O. I., & Kang, J. (2023). Cancer statistics in Korea: Incidence, mortality, survival, and prevalence in 2020. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *55*(2), 385–399. https://doi.org/10.4143/crt.2023.447

Lalitha, S., & Murugan, D. (2023). Segmentation and classification of 3D lung tumor diagnoses using convolutional neural networks. *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 230–238. https://doi.org/10.1109/ICAISS58487.2023.10250625

Makubhai, S. S., Pathak, G. R., & Chandre, P. R. (2024). Predicting lung cancer risk using explainable artificial intelligence. *Bulletin of Electrical Engineering and Informatics, 13*(2), 628–640. https://doi.org/10.11591/eei.v13i2.6280

Mengash, H. A., Alamgeer, M., Maashi, M. S., Othman, M., Hamza, M. A., Ibrahim, S. S., Zamani, A. S., & Yaseen, I. (2023). Leveraging marine predators algorithm with deep learning for lung and colon cancer diagnosis. *Cancers, 15*(5), Article 1591. https://doi.org/10.3390/cancers15051591

Mulshine, J. L. (2005). Issues in the management of early lung cancer. *Chest, 128*(1), 83–85. https://doi.org/10.1378/chest.128.1.83

Rao, K. B., Bhardwaj, Y., Rao, G. V., Gurrala, J., Jain, A., & Gupta, K. (2023). Early lung cancer prediction by AI-inspired algorithm. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics, and Computer Engineering (UPCON)*, 1466–1469. https://doi.org/10.1109/UPCON59197.2023.10434702

Sarkar, S., Singh, H., & Chawla, J. (2024). Precision tumor identification in lung CT scans: Leveraging convolutional neural networks for accurate segmentation. *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 1–6. https://doi.org/10.1109/ACROSET62108.2024.10743910

Setio, A. A. A., Ciompi, F., Litjens, G. J., Gerke, P. K., Jacobs, C., Riel, S. J. V., 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. https://doi.org/10.1109/TMI.2016.2536809

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, *71*(1), 33–54. https://doi.org/10.3322/caac.21654

Tran, T., Vo, T. H., & Le, N. Q. (2023). Omics-based deep learning approaches for lung cancer decision-making and therapeutics development. *Briefings in Functional Genomics, 22*(3), 239–250. https://doi.org/10.1093/bfgp/elad031

Wani, N. A., Kumar, R., & Bedi, J. (2023). DeepXplainer: An interpretable deep learning-based approach for lung cancer detection using explainable artificial intelligence. *Computer Methods and Programs in Biomedicine, 243*, Article 107879. https://doi.org/10.1016/j.cmpb.2023.107879